



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Υπηρεσίες Διαλειτουργικότητας Δεδομένων με εφαρμογή σε
Ετερογενείς Υποδομές και Εφαρμογές Υγείας
(Services of Data Interoperability applied to Heterogeneous Healthcare
Infrastructures and Applications)

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Αθανάσιος Δ. Κιούρτης

ΠΕΙΡΑΙΑΣ
Ιούνιος 2019



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

**Υπηρεσίες Διαλειτουργικότητας Δεδομένων με εφαρμογή σε
Ετερογενείς Υποδομές και Εφαρμογές Υγείας
(Services of Data Interoperability applied to Heterogeneous Healthcare
Infrastructures and Applications)**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Αθανάσιος Δ. Κιούρτης

Συμβουλευτική επιτροπή:

Επ. Καθ. Δημοσθένης Κυριαζής
Καθ. Γεώργιος Βασιλακόπουλος
Αν. Καθ. Ηλίας Μαγκλογιάννης

Εγκρίθηκε από την επιταμελή εξεταστική επιτροπή την

.....
Δ. Κυριαζής
Επ. Καθ. Πανεπ. Πειραιά

.....
Γ. Βασιλακόπουλος
Καθ. Πανεπ. Πειραιά

.....
Η. Μαγκλογιάννης
Αν. Καθ. Πανεπ. Πειραιά

.....
Μ. Φιλιππάκης
Αν. Καθ. Πανεπ. Πειραιά

.....
Χ. Δουλκερίδης
Επ. Καθ. Πανεπ. Πειραιά

.....
Χ. Ξενάκης
Αν. Καθ. Πανεπ. Πειραιά

.....
Π. Τσανάκας
Καθ. Εθνικού Μετσόβιου Πολυτεχνείου

ΠΕΙΡΑΙΑΣ
Ιούνιος 2019

.....

Αθανάσιος Δ. Κιούρτης

Copyright © Αθανάσιος Δ. Κιούρτης, 2019

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Πειραιά.

Περίληψη

Την τελευταία δεκαετία, υπήρξε μια μετάβαση από έναν περιορισμένο σε δεδομένα κόσμο σε έναν πλούσιο σε δεδομένα κόσμο με στόχο τη βελτίωση της ποιότητας του τομέα των μεταφορών, της διακυβέρνησης, του περιβάλλοντος, της επικοινωνίας και της υγείας. Μεγάλο μέρος αυτής της άνευ προηγουμένου αύξησης της παραγωγής δεδομένων μπορεί να αποδοθεί στην παρουσία πολλών κινητών συσκευών, φορητών συσκευών και αισθητήρων. Αυτό έχει ως αποτέλεσμα ένα πλήθος συσκευών που έχουν τη δυνατότητα σύνδεσης με το διαδίκτυο, παράγοντας δεδομένα διαφορετικών τύπων που μπορεί να έχουν συλλεχθεί σε διαφορετικές χρονικές στιγμές και τοποθεσίες, από ετερογενείς συσκευές. Οι περισσότερες από αυτές τις συσκευές τυπικά χαρακτηρίζονται από υψηλό βαθμό ετερογένειας, όσον αφορά την ύπαρξη διαφορετικών χαρακτηριστικών, δυνατοτήτων ή προδιαγραφών δικτύου, τα οποία πρέπει να είναι αποδοτικά διαχειρίσιμα αναφορικά κυρίως με τη διαχείριση των ετερογενών δεδομένων που παράγουν. Ειδικά για τις ιατρικές συσκευές - Internet of Medical Things (IoMT), οι οποίες υιοθετούνται ευρέως και χρησιμοποιούνται στην 4^η Βιομηχανική Επανάσταση του χώρου της Υγείας (Healthcare 4.0), η ανάγκη χρήσης, κατανόησης και επεξεργασίας των δεδομένων τους χρήζει ως ένα ζήτημα ζωτικής σημασίας. Αναφορικά ωστόσο με τη βέλτιστη αξιοποίηση των δεδομένων αυτών, η ετερογένεια, η πολυπλοκότητα, ο θόρυβος και η ατέλειά τους, συγκαταλέγονται στις πιο κοινές τους προκλήσεις. Είναι εμφανές πως μεγάλα ποσά ετερογενών ιατρικών δεδομένων καθίστανται διαθέσιμα σε διάφορους οργανισμούς και συσκευές υγειονομικής περίθαλψης, αναμορφώνοντας έτσι πλήρως το χώρο του Healthcare 4.0. Προκειμένου αυτά τα ετερογενή ιατρικά δεδομένα να ανταλλάσσονται με όσο το δυνατόν περισσότερους ενδιαφερόμενους και να αποτελούν βασικό γνώμονα παροχής εξατομικευμένης και αποδοτικής ιατρικής φροντίδας προς τους ασθενείς και τους πολίτες, η διαλειτουργικότητα είναι ο μόνος τρόπος. Σαν πρόσθετα αποτελέσματα, μπορεί να αυξηθεί η ποιότητα των υπηρεσιών υγείας, να μειωθεί το κόστος της και τα ιατρικά σφάλματα, να αυξηθεί η ιδιωτικότητα των ασθενών, και να ανανεωθεί ολοκληρωτικά η δημόσια υγεία.

Επί του παρόντος, η ταχέως αυξανόμενη διαθεσιμότητα των ιατρικών αρχείων ωθεί στην υιοθέτηση προσεγγίσεων που βασίζονται σε δεδομένα, όπως η ακριβέστερη πρόγνωση ασθενειών και συμπτωμάτων, η αυτοματοποίηση των καθηκόντων που σχετίζονται με την υγειονομική περίθαλψη, η ευκολότερη

ανίχνευση ασθενειών, και η αποδοτικότερη κλινική έρευνα. Παρόλα αυτά, οι οργανισμοί υγειονομικής περίθαλψης εξακολουθούν να αντιμετωπίζουν πολλές δυσκολίες στην εφαρμογή, στη συντήρηση και στην αναβάθμιση των συστημάτων υγειονομικής περίθαλψης, συμπεριλαμβανομένων πολλών προκλήσεων στους τομείς της τεχνολογίας, της ασφάλειας και της ανθρώπινης επικοινωνίας. Μεταξύ άλλων, μια πρόσθετη πρόκληση αναφέρεται στην έλλειψη μιας ολοκληρωμένης προσέγγισης ανταλλαγής δεδομένων, τα οποία προκειμένου να ανταλλάσσονται με όσο το δυνατόν περισσότερους ενδιαφερόμενους ώστε να βελτιώσουν τη δημόσια υγεία, απαιτούν την ύπαρξη της διαλειτουργικότητας, η οποία αποτελεί το βασικό μέσο για να δοθεί η δυνατότητα στα συστήματα να αλληλοεπιδρούν μεταξύ τους. Έχοντας κατά νου ότι η αγορά φορητών ιατρικών συσκευών αναμένεται να τετραπλασιαστεί, και πως τα νοσοκομεία και τα ιατρεία σε εθνικό επίπεδο θα μπορούσαν να βελτιώσουν δραματικά την ποιότητα ζωής των ασθενών, αυτό έχει ως αποτέλεσμα να καταστεί ακόμη πιο δύσκολο και επιτακτικό το έργο της διαλειτουργικότητας στον τομέα της υγείας. Για το σκοπό αυτό, πολλές τεχνικές τυποποίησης αλλά και έρευνες επενδύονται ετησίως στη διαλειτουργικότητα των δεδομένων για την υγεία. Μεταξύ των διαφόρων υφιστάμενων ερευνών και προτύπων, ο οργανισμός (Health Level 7) HL7 παρέχει την ανάπτυξη και το πλαίσιο των προτύπων που χρησιμοποιούνται ευρέως στην ιατρική αγορά και την έρευνα, με το HL7 Fast Healthcare Interoperability Resources (FHIR) να είναι το πιο πρόσφατο πρότυπο που δημιουργήθηκε από τον οργανισμό για την ανταλλαγή κλινικών πληροφοριών.

Βάσει όλων των προαναφερθέντων υφιστάμενων προκλήσεων, η παρούσα Διδακτορική Διατριβή στοχεύει στη δημιουργία μιας προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων σε μορφή ενός αυτοματοποιημένου μετασχηματισμού ανομοιογενών δεδομένων υγείας από ετερογενείς πηγές δεδομένων. Η προσέγγιση αυτή είναι ικανή να μετασχηματίζει τόσο σε σημασιολογικό όσο και σε συντακτικό επίπεδο δεδομένα διαφορετικής αναπαράστασης και μορφής, προκειμένου αυτά να μπορούν να ταυτοποιούνται, να αντιστοιχίζονται σε μία ενιαία γλώσσα, και στο τέλος να συγχωνεύονται σε ένα κοινό επίπεδο. Για να επιτευχθεί αυτό, αρχικά έγινε εμβάθυνση στους μηχανισμούς που έχουν υλοποιηθεί ώστε τα δεδομένα υγείας και οι πληροφορίες να μπορούν να ανταλλάσσονται διαλειτουργικά μεταξύ συστημάτων που ανήκουν σε διαφορετικές ομάδες ή υπακούν σε διαφορετικά πρότυπα υγείας. Κατόπιν σχετικής έρευνας, προτάθηκε μία προσέγγιση η οποία είναι ικανή να διαχωρίζει τα δεδομένα υγείας σε κατηγορίες ανάλογα με τον τρόπο αναπαράστασής τους, και στη συνέχεια εξαγοντας γνώσεις, πληροφορίες και μεταδεδομένα, τα δεδομένα μετασχηματίζονται σε μία ενιαία

γλώσσα, η οποία διατηρεί συνδυαστικά στοιχεία από άλλες ήδη προτυποποιημένες γλώσσες αναπαράστασης δεδομένων υγείας. Εν συνεχεία, η συγκεκριμένη προσέγγιση μετασχηματισμού προσαρμόστηκε ώστε να παράγει οντολογίες μέσω συγκεκριμένων επιπέδων μεταμοντέλου και οντολογιών, με απώτερο στόχο την κοινή αναπαράσταση των δεδομένων σε μία ενιαία μορφή. Χρησιμοποιώντας ως γνώμονα τις παραχθείσες οντολογίες, η προσέγγιση που ακολουθήθηκε εστίασε αφενός στη δημιουργία των οντολογιών, και αφετέρου στη σύγκρισή τους με τις οντολογίες των επιμέρους πόρων (resources) του προτύπου υγείας HL7 FHIR, λόγω της παγκόσμιας διάδοσης και υιοθέτησής του. Η σύγκριση αυτή αφορούσε τόσο τη συντακτική όσο και τη σημασιολογική απεικόνιση των εκάστοτε οντολογιών, ώστε στο τέλος να καθίσταται εφικτή η αντιστοίχιση των οντολογιών των δεδομένων υγείας με τις οντολογίες των πόρων HL7 FHIR. Η τελική προσέγγιση βασίστηκε στην προαναφερθείσα λύση, όπου τόσο οι μηχανισμοί υλοποίησης των οντολογιών όσο και της συντακτικής και της σημασιολογικής σύγκρισης και αντιστοίχισης βελτιώθηκαν, βάσει σχετικών πειραματικών δοκιμών.

Εστιάζοντας στο χώρο της υγείας, οι περισσότερες από τις ήδη υπάρχουσες πρακτικές και τεχνικές διαλειτουργικότητας δεδομένων στηρίζονται σε πρότυπα υγείας όπως και σε μηχανισμούς οντολογικών μετασχηματισμών και αντιστοιχίσεων. Ωστόσο είναι σχεδιασμένες για συγκεκριμένες περιπτώσεις χρήσης με ήδη προκαθορισμένη τη μορφή με την οποία πρέπει να εισέρχονται τα δεδομένα, χωρίς να έχουν τη δυνατότητα προσαρμογής, μη μπορώντας έτσι να θεωρηθούν ως ολιστικές λύσεις. Η συγκεκριμένη πρόκληση απαντάται από την παρούσα Διδακτορική Διατριβή, αποδίδοντας μια ενιαία και γενικευμένη προσέγγιση η οποία μπορεί να εφαρμοστεί αυτοματοποιημένα σε οποιοδήποτε σύνολο δεδομένων υγείας, χωρίς την ανάγκη προ-επεξεργασίας τους, αποδίδοντας και αντιστοιχίζοντας τα σε ένα από τα πιο ισχυρά πρότυπα υγείας, το HL7 FHIR. Τα αποτελέσματα των πειραματικών δοκιμών που πραγματοποιήθηκαν αξιολογούν και αποδεικνύουν τη λειτουργικότητα και αποδοτικότητα της προτεινόμενης προσέγγισης, καθιστώντας την εφικτή για χρήση και ενσωμάτωση σε πολλαπλούς τομείς και περιβάλλοντα, και ειδικότερα στο χώρο της υγειονομικής περίθαλψης, καθώς και στους τομείς των τηλεπικοινωνιών και των δικτύων, της αναγνώρισης συσκευών, της ηλεκτρονικής διακυβέρνησης και των ηλεκτρονικών προμηθειών.

Abstract

In the last decade, there has been a transition from a data-poor to a data-rich world, with the aim of improving the quality of transport, governance, environment, communication and health. Much of this unprecedented increase in data generation can be attributed to the abundance of thousands of mobile devices, wearables, and sensors. Such thing results into a myriad of heterogeneous devices that will be connected to the world of Internet of Things (IoT), producing data of different types that may have been collected at different locations and time scales, by different devices. Most of these devices are typically characterized by a high degree of heterogeneity, in terms of having different characteristics, capabilities, or network specifications, needing to be easily manageable in particular with regard to the management of the heterogeneous data they generate. Especially for the Internet of Medical Things (IoMT) devices that are widely adopted and used in Healthcare 4.0, the need for using, understanding, and processing of these devices' data is an issue of vital importance. However, with regard to the best use of this data, heterogeneity, complexity, noise and imperfection are among their most common challenges. It is undeniable that vast amounts of heterogeneous medical data are becoming available in various healthcare organizations and devices, thus completely reshaping the Healthcare 4.0. In order for these heterogeneous medical data to be exchanged with as many stakeholders as possible, and to be a key driver of providing personalized and efficient medical care to patients and citizens, interoperability is the only way. Additionally, the quality of the healthcare services can be improved, health costs and medical errors can be reduced, and patients' privacy can be increased, while public health can be completely renewed.

Currently, the rapidly increasing availability of health records is pushing towards the adoption of data-driven approaches, bringing the opportunities for more accurate disease and symptom prognosis, to automate healthcare related tasks, providing better disease detection, and more efficient clinical research. Nevertheless, the healthcare organizations are still facing many difficulties in implementing, maintaining and upgrading their healthcare systems, including many challenges in the technical, security and human interaction fields. Among others, what is missing is an integrated data exchange system for which, in order to exchange data with as many stakeholders as possible to improve public health, interoperability is the only way for letting systems interact with each other. Having in mind that the wearable medical devices market is expected to quadruple, and that hospitals and doctors' offices

nationwide could dramatically improve their patients' quality of life, such thing results into making the health interoperability task even more daunting. For this purpose, many standardization techniques are annually invested in health data interoperability. Among the different existing researches and standards, the Health Level Seven (HL7) organization provides the development and the framework of standards that are widely used in the medical market and research, with HL7 Fast Healthcare Interoperability Resources (FHIR) being the latest standard created by this organization, for the exchange of clinical information.

The current PhD dissertation aims to create an Interoperability Services' approach in the form of an automated transformation of heterogeneous medical data deriving from heterogeneous devices. This approach will transform both semantically and syntactically health data of different representation and morphology, in order to identify, match into a single language, and eventually merge it into a common level. For this reason, the mechanisms that were studied more are those that aim in the interoperable exchange of health data and information, between systems belonging to different groups or under different health standards. Following this research, an approach was developed which was able to separate health data into categories according to how they were modelled, and then exporting knowledge, information and metadata, the data was translated into a single language, which retained combined data from other already-standardized health data representation languages. Subsequently, this translation approach was adapted to generate ontologies through specific metamodel layers and ontologies, with the ultimate goal of jointly rendering the data in a single format. Using the created ontologies, the next approach was to create ontologies and to compare them with individual HL7 FHIR resources' ontologies due to its worldwide dissemination and adoption. This comparison concerned both the syntactic and the semantic mapping of the ontologies so that the ontologies of the health data can be mapped to the ontologies of the HL7 FHIR resources. The final approach was therefore based on the above-mentioned solution, where both the ontology implementation mechanisms, but also the syntactic and semantic comparison and mapping, were improved based on relevant experimental tests.

Focusing on the healthcare domain, most of the existing health data interoperability practices and techniques are based on health standards as well as mechanisms for ontological transformations and mappings. However, these are designed for specific cases of use with the pre-defined format in which the data must be entered without having adaptability, without the ability to be considered as holistic solutions. This gap is covered by the current PhD dissertation, delivering a unified and generalized approach that can be applied

automatically to any set of health data without the need for pre-processing, assigning and matching this data to one of the most powerful and promising health standards, the HL7 FHIR. The results of the experimental tests listed below, evaluate and demonstrate the operation and efficiency of the proposed approach, making it possible to use it for integration into multiple sectors and environments, particularly in the field of health care, as well as in the fields of telecommunications and networks, devices' identification, e-Government and e-Procurement.

Ευχαριστίες

Η Διδακτορική Διατριβή που παρουσιάζεται στις επόμενες σελίδες εκπονήθηκε από τον Δεκέμβριο του 2015 μέχρι τον Ιούνιο του 2019, στο τμήμα Ψηφιακών Συστημάτων της σχολής Τεχνολογιών Πληροφορικής και Τηλεπικοινωνιών, του Πανεπιστημίου Πειραιά.

Κατά τη διάρκεια της εκπόνησης της Διατριβής, είχα την ευκαιρία να συνεργαστώ με αρκετούς ερευνητές στον κλάδο της Πληροφορικής, να παρουσιάσω το έργο μου και την πορεία του σε διεθνή συνέδρια και περιοδικά, καθώς και να ασχοληθώ με αρκετά ενδιαφέροντα θέματα που αφορούν τον ερευνητικό χώρο της διαλειτουργικότητας των δεδομένων.

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου Επ. Καθηγητή κ. Δημοσθένη Κυριαζή για το ενδιαφέρον που έδειξε, για τις πολύτιμες συμβουλές του, για την υποστήριξη που μου παρείχε κατά τη διάρκεια της πορείας μου, και για τις ευκαιρίες που μου έδωσε ώστε να εξελιχθώ και να διαδώσω το έργο μου. Θα ήθελα επίσης να ευχαριστήσω ιδιαίτερα τους καθηγητές της Τριμελούς Συμβουλευτικής Επιτροπής, Καθηγητή κ. Γεώργιο Βασιλακόπουλο και Αν. Καθηγητή κ. Ηλία Μαγκλογιάννη, καθώς και τον Καθηγητή κ. Μαρίνο Θεμιστοκλέους, για τις συμβουλές τους και τις προτάσεις βελτίωσης του έργου μου.

Θα ήθελα επίσης να ευχαριστήσω το Ίδρυμα Παιδείας και Ευρωπαϊκού Πολιτισμού, καθώς και το Ελληνικό Ίδρυμα Έρευνας και Καινοτομίας (ΕΛΙΔΕΚ) για την οικονομική ενίσχυση που μου παρείχαν τα χρονικά διαστήματα 2016-2017 και 2017-2019, αντίστοιχα.

Επιπλέον, θα ήθελα να ευχαριστήσω τα μέλη της ερευνητικής ομάδας, με τους οποίους συνεργάστηκα και συνδυάσαμε τις διαφορετικές ερευνητικές μας προσεγγίσεις, σε πολλαπλούς επιστημονικούς τομείς. Θα ήθελα επίσης να ευχαριστήσω τη συνάδελφο, συνοδοιπόρο, συνεργάτη, και σύντροφό μου, η οποία με στήριξε ιδιαίτερα στις δύσκολες στιγμές της πορείας μου, ώστε τελικά να τα καταφέρω.

Ένα μεγάλο ευχαριστώ οφείλω στην οικογένειά μου, στη συμπαράσταση και τη δύναμη της μητέρας μου, στη δυναμικότητα και στον εγωισμό της αδερφής μου, και στις θυσίες του πατέρα μου. Στη μνήμη του αφιερώνω κάθε λεπτό του έργου μου, και κάθε σελίδα της Διατριβής μου.

Σε αυτούς που με έφτασαν εδώ...

Πίνακας Περιεχομένων

Κατάλογος Συντμήσεων.....	5
Κατάλογος Σχημάτων.....	9
Κατάλογος Πινάκων.....	13

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή.....	15
1.1 Διαλειτουργικότητα Δεδομένων.....	16
1.2 Τομείς, Προκλήσεις και Όραμα Διαλειτουργικότητας Δεδομένων	21
1.3 Αντικείμενο και Καινοτομία Διατριβής.....	28
1.4 Δομή Διατριβής	30

ΚΕΦΑΛΑΙΟ 2

Διαλειτουργικότητα Δεδομένων.....	33
2.1 Δεδομένα και Διαλειτουργικότητα Δεδομένων	34
2.1.1 Ανάγκη Διαλειτουργικότητας και Δεδομένων	34
2.1.2 Κατηγορίες Διαλειτουργικότητας.....	44
2.1.2.1 Νομική Διαλειτουργικότητα.....	47
2.1.2.2 Οργανωτική Διαλειτουργικότητα	48
2.1.2.3 Σημσιολογική Διαλειτουργικότητα.....	50
2.1.2.4 Τεχνική Διαλειτουργικότητα.....	54
2.1.3 Πρότυπα Διαλειτουργικότητας.....	56
2.1.3.1 ISO (International Organization for Standardization)	56
2.1.3.2 CEN (European Committee for Standardization).....	59
2.1.3.3 UN/CEFACT (United Nations/Centre for Trade Facilitation and Electronic Business)	61
2.1.3.4 W3C (World Wide Web Consortium)	62
2.1.3.5 EIF (European Interoperability Framework).....	64
2.1.3.6 OASIS (Organization for the Advancement of Structured Information Standards)	66
2.1.3.7 IDABC (Interoperable Delivery of European eGovernment Services to Public Administrations, Businesses and Citizens).....	68
2.1.4 Τεχνικές Διαλειτουργικότητας.....	70

2.2	Διαλειτουργικότητα στον τομέα της Υγείας	74
2.2.1	Ανάγκη Διαλειτουργικότητας στον τομέα της Υγείας.....	74
2.2.2	Κατηγορίες Διαλειτουργικότητας στον τομέα της Υγείας.....	82
2.2.2.1	Νομική Διαλειτουργικότητα.....	83
2.2.2.2	Οργανωτική Διαλειτουργικότητα	83
2.2.2.3	Σημσιολογική Διαλειτουργικότητα.....	84
2.2.2.4	Τεχνική (Συντακτική) Διαλειτουργικότητα.....	84
2.2.3	Πρότυπα Διαλειτουργικότητας στον τομέα της Υγείας.....	85
2.2.3.1	DICOM (Digital Imaging and Communications in Medicine)	90
2.2.3.2	SNOMED (Systematized Nomenclature of Medicine - Clinical Terms) 93	
2.2.3.3	ICD-10 (International Classification of Diseases).....	94
2.2.3.4	LOINC (Logical Observation Identifiers Names and Codes)	96
2.2.3.5	HL7 v2/HL7 v3 (Health Level Seven Version 2/ Version 3).....	97
2.2.3.6	HL7 FHIR (Health Level Seven Fast Healthcare Interoperability) 100	
2.2.4	Τεχνικές Διαλειτουργικότητας στον τομέα της Υγείας	104
2.2.4.1	Κομβικές Τεχνικές Διαλειτουργικότητας στον τομέα της Υγείας	107

ΚΕΦΑΛΑΙΟ 3

Βασικές Τεχνολογίες.....	115
3.1 Οντολογίες.....	116
3.1.1 Ανάγκη Χρήσης Οντολογιών	116
3.1.2 Δημιουργία Οντολογιών.....	120
3.1.2.1 OWL (Web Ontology Language)	120
3.1.2.2 Κομβικές Τεχνικές Δημιουργίας	124
3.1.3 Χρήση Οντολογιών στον τομέα της Υγείας.....	128
3.2 Επεξεργασία Φυσικής Γλώσσας.....	131
3.2.1 Ανάγκη και Χρήση Επεξεργασίας Φυσικής Γλώσσας	131
3.2.2 Τεχνικές Επεξεργασίας Φυσικής Γλώσσας στον τομέα της Υγείας	136
3.2.2.1 Τεχνικές Συντακτικής Ομοιότητας	136
3.2.2.1.1 Απόσταση <i>Levenshtein</i>	137
3.2.2.1.2 Ομοιότητα <i>Cosine</i>	138
3.2.2.1.3 Ομοιότητα <i>Jaccard</i>	138
3.2.2.1.4 Ομοιότητα <i>Jaro-Winkler</i>	139
3.2.2.2 Τεχνικές Σημσιολογικής Ομοιότητας.....	140

3.2.2.2.1	<i>Word2Vec</i>	141
3.2.2.2.2	<i>WordNet</i>	142
3.2.2.2.3	<i>Cortical.io</i>	144

ΚΕΦΑΛΑΙΟ 4

Προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων	147
4.1 Αρχικά Στάδια Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων	148
4.1.1 Στάδιο Μεταφοράς Γνώσεων συγκεκριμένου τομέα σε έναν άγνωστο τομέα, κατά μήκος του Κύκλου Ζωής των Δεδομένων	148
4.1.2 Στάδιο Συσσωμάτωσης Ετερογενών δεδομένων υγείας μέσω μιας Οντολογικής Κοινής Γλώσσας Υγείας	153
4.1.3 Στάδιο απόκτησης της Οντολογικής Αναπαράστασης των Δεδομένων της Υγείας μέσω των Τεχνικών Μεταμοντέλων	160
4.1.4 Στάδιο Οντολογικού Αντιστοιχιστή FHIR (FHIR Ontology Mapper - FOM): Συγκέντρωση Συντακτικών και Σημασιολογικών Ομοιοτήτων Οντολογιών προς την αντιστοίχισή τους σε HL7 FHIR.....	168
4.2 Προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων	177
4.2.1 Επίπεδο Οντολογιών και Συσχετίσεων.....	180
4.2.2 Γνωσιακή Βάση Δεδομένων.....	182
4.2.3 Αλγόριθμος Συσχέτισης Δομής.....	185
4.2.3.1 Τεχνικές Συντακτικής και Σημασιολογικής Ομοιότητας.....	185
4.2.3.1.1 Τεχνική Συντακτικής Ομοιότητας	185
4.2.3.1.2 Τεχνική Σημασιολογικής Ομοιότητας	193
4.2.3.1.3 Τεχνική Ολικής Οντολογικής Αντιστοίχισης.....	199
4.2.3.2 Τεχνική Οντολογικής Ευθυγράμμισης.....	201
4.2.4 Τεχνική Μετάφρασης Δομής FHIR.....	203

ΚΕΦΑΛΑΙΟ 5

Αξιολόγηση.....	205
5.1 Περιβάλλον Δοκιμών	206
5.2 Αξιολόγηση Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων	207
5.2.1 1 ^ο Πείραμα: Εφαρμογή Τεχνικών Συντακτικής και Σημασιολογικής Ομοιότητας συνδυαστικά με την Τεχνική Οντολογικής Ευθυγράμμισης.....	207
5.2.1.1 Περιγραφή Συνόλου Δεδομένων Πειράματος.....	207
5.2.1.2 Αποτελέσματα Πειράματος	208

5.2.1.3	Συζήτηση Αποτελεσμάτων Πειράματος.....	211
5.2.2	2 ^ο Πείραμα: Συγκριτική Αξιολόγηση Απόδοσης Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων	217
5.2.2.1	Περιγραφή Μηχανισμού Σύγκρισης	217
5.2.2.2	Περιγραφή Συνόλου Δεδομένων Πειράματος.....	218
5.2.2.3	Αποτελέσματα Πειράματος	220
5.2.2.4	Συζήτηση Αποτελεσμάτων Πειράματος.....	224

ΚΕΦΑΛΑΙΟ 6

Σύνοψη Διατριβής - Ανοιχτά Ερευνητικά Θέματα	231	
6.1	Σύνοψη Διατριβής και Κυριότερων Συνεισφορών.....	232
6.1.1	Εφαρμογή και Χρήση της Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων	235
6.2	Ανοιχτά Ερευνητικά Θέματα και Μελλοντικά Σχέδια.....	238
Βιβλιογραφία	243	

Κατάλογος Συντμήσεων

Σύντμηση	Επεξήγηση
5G	5 th Generation
ACR	American College of Radiology
ADL	Archetype Description Language
AEs	Application Entities
API	Application Programming Interface
BII	Business Interoperability Interfaces
C3	Command and Control
CAP	College of American Pathologists
CBOW	Continuous Bag of Words
CCTS	Core Components Technical Specification
CEN	European Committee for Standardization
CHIME	College of Healthcare Information Management Executives
CHISTAR	Cloud Health Information Systems Technology Architecture
CHL	Common Health Language
COM	Component Object Model
CORBA	Common Object Request Broker Architecture
CSS	Cascading Style Sheets
CSV	Comma Separated Values
CTv3	Clinical Terms version 3
DAL	Domain Agnostic Language
DCM	Detailed Clinical Models
DCOM	Distributed Component Object Model
DI	Data Integrator
DICOM	Digital Imaging and Communications in Medicine
DSL	Domain Specific Language
ebMS	ebXML Messaging Service Specification
ebXML	Electronic business eXtensible Markup Language
eHealth	Electronic Health
EHRs	Electronic Health Records
EIF	European Interoperability Framework
ESB	Enterprise Service Bus
FCAT	Federative Committee on Anatomical Terminology
FHIR	Fast Healthcare Interoperability Resources
FOM	FHIR Ontology Mapper

HIMSS	Healthcare Information and Management Systems Society
HIPAA	Health Insurance Portability and Accountability Act
HITECH	Health Information Technology for Economic and Clinical Health Act
HL7	Health Level 7
HL7 CDA	HL7 Clinical Document Architecture
HL7 v2	HL7 version 2
HL7 v3	HL7 version 3
HSB	Health Service Bus
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
ICD-10	International Classification of Diseases 10
IDABC	Interoperable Delivery of European eGovernment Services to Public Administrations, Businesses and Citizens
IDC	International Data Corporation
IDL	Interface Description Language
IHE	Integrating the Healthcare Enterprise
IODs	Information Object Definitions
IoE	Internet of Everything
IoMT	Internet of Medical Things
IoT	Internet of Things
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
JUNG	Java Universal Network/Graph
JVM	Java Virtual Machine
KB	Kilobytes
LISI	Levels of Information System Interoperability
LOINC	Logical Observation Identifiers Names and Codes
MDE	Model Driven Engineering
MOF	Meta-Object Facility
NC3TA	NATO Consultation, Command and Control Technical Architecture
NCOIC	Network Centric Operations Industry Council Interoperability
NEMA	National Electrical Manufacturers Association
NHS	National Health Service
NLP	Natural Language Processing
NLU	Natural Language Understanding
NoSQL	Non Structured Query Language

OAEI	Ontology Alignment Evaluation Initiative
OASIS	Organization for the Advancement of Structured Information Standards
ODBC	Open Database Connectivity
OIM	Organisational Interoperability Maturity
ONC	Office of the National Coordinator
OTS	Off-The-Shelf
OWL	Web Ontology Language
OWL-DL	OWL-Description Logic
PGS	Patient Guidance System
PHRs	Personal Health Records
PPACA	Patient Protection and Affordable Care Act
RDF	Resource Description Framework
RDFS	RDF Schema
RIM	Reference Information Model
RMI	Remote Method Invocation
ROM	Reverse Ontology Mapping
SCPs	Service Class Providers
SCUs	Service Class Users
SDRs	Sparse Distributed Representations
SGML	Structured General Markup Language
SMTP	Simple Mail Transfer Protocol
SNOMED	Systematized Nomenclature of Human and Veterinary Medicine
SOAP	Simple Object Access Protocol
SOPs	Service Object Pairs
SQL	Structured Query Language
SVM	Support Vector Machines
TF-IDF	Term Frequency - Inverse Document Frequency
UBL	Universal Business Language
UML	Unified Modelling Language
UMLS	Unified Medical Language System
UMM	UN/CEFACT Modelling Methodology
UN/CEFACT	United Nations Centre for Trade Facilitation and Electronic Business
URI	Uniform Resource Identifiers
URL	Uniform Resource Locator
VMR	Virtual Medical Record
W3C	World Wide Web Consortium
WHO	World Health Organization

Wi-Fi	Wireless-Fidelity
WONCA	World Organization of Family Doctors
WSDL	Web Services Description Language
www	World Wide Web
XML	eXtensible Markup Language
XPath	XML Path
XSD	XML Schema Definition
XSG	XML Schema Graph
XSL	EXtensible Stylesheet Language
XSLT	eXtensible Stylesheet Language Transformations
XSOM	XML Schema Object Model
XTR	XML-to-Reational
ΤΠΕ	Τεχνολογίας Πληροφοριών και Επικοινωνιών

Κατάλογος Σχημάτων

ΚΕΦΑΛΑΙΟ 2

Εικόνα 2-1 - Προκλήσεις επίτευξης διαλειτουργικότητας	34
Εικόνα 2-2 - Αύξηση όγκου δεδομένων.....	35
Εικόνα 2-3 - Τα 3Vs των Big Data	37
Εικόνα 2-4 - Βασικές κατηγορίες διαλειτουργικότητας	46
Εικόνα 2-5 - Ποιότητα δεδομένων για σημασιολογικά διαλειτουργικά EHRs.....	107
Εικόνα 2-6 - Ενίσχυση της σημασιολογικής διαλειτουργικότητας στην υγειονομική περίθαλψη με τη χρήση εξόρυξης σημασιολογικών διαδικασιών.....	108
Εικόνα 2-7 - Ένα τεχνικό σχέδιο για την επίτευξη της σημασιολογικής διαλειτουργικότητας στα ηλεκτρονικά συστήματα καταγραφής της υγείας.....	109
Εικόνα 2-8 - SAPHIRE - έξυπνη παρακολούθηση της υγειονομικής περίθαλψης με βάση μία πλατφόρμα σημασιολογικής διαλειτουργικότητας.....	110
Εικόνα 2-9 - Ένα πλαίσιο για τη σημασιολογική διαλειτουργικότητα στην υγειονομική περίθαλψη: Αρχιτεκτονική με βάση τα πρότυπα υγείας	111
Εικόνα 2-10 - Οι οντολογίες ενισχύουν τη σημασιολογική διαλειτουργικότητα	112
Εικόνα 2-11 - Επίλυση της πρόκλησης διαλειτουργικότητας ενός κατακεκομμένου πολύπλοκου συστήματος καθοδήγησης ασθενών.....	113
Εικόνα 2-12 - Μια προσέγγιση για τη σημασιολογική διαλειτουργικότητα των αρχέτυπων ISO EN 13606 και OpenEHR.....	114

ΚΕΦΑΛΑΙΟ 3

Εικόνα 3-1 - Αντιστοίχιση XML σε OWL Οντολογίες	125
Εικόνα 3-2 - Δημιουργία οντολογιών από πηγές δεδομένων XML.....	126
Εικόνα 3-3 - Μετατρέποντας το XML Schema σε OWL μέσω μοτίβων	127

ΚΕΦΑΛΑΙΟ 4

Εικόνα 4-1 - 1ο Βήμα μεταφοράς δεδομένων σε DSL.....	151
Εικόνα 4-2 - 2ο Βήμα μεταφοράς δεδομένων σε DAL	151
Εικόνα 4-3 - 3ο Βήμα μεταφοράς και αναγνώρισης δεδομένων σε DSL.....	152

Εικόνα 4-4 – Αρχιτεκτονική προσέγγισης Κοινής Γλώσσας Υγείας.....	155
Εικόνα 4-5 – Μηχανισμός φιλτραρίσματος OWL οντολογιών	157
Εικόνα 4-6 – Εύρεση κοινών σχέσεων, τιμών και εννοιών.....	158
Εικόνα 4-7 – Δημιουργία κανόνων	158
Εικόνα 4-8 – Μετασηματισμός κανόνων σε οντολογική μορφή	159
Εικόνα 4-9 – Μεταφορά δεδομένων σε OWL οντολογίες.....	162
Εικόνα 4-10 – Αρχιτεκτονική επιπέδων μεταμοντέλου και οντολογιών.....	164
Εικόνα 4-11 – Μετασηματισμός ενός ιατρικού προτύπου EHR στην αναπαράσταση του μεταμοντέλου.....	166
Εικόνα 4-12 – Αρχιτεκτονική Οντολογικού Αντιστοιχιστή FHIR.....	171
Εικόνα 4-13 – Προσέγγιση υπηρεσιών διαλειτουργικότητας δεδομένων	179
Εικόνα 4-14 - Δημιουργία οντολογιών από XML δεδομένα.....	180
Εικόνα 4-15 – Παράδειγμα ανάλυσης XML Schema μέσω του XSOM.....	181
Εικόνα 4-16 – Δομή γνωσιακής βάσης δεδομένων.....	182
Εικόνα 4-17 – Δεδομένα περίπτωσης χρήσης για εύρεση βέλτιστου αλγορίθμου συντακτικής ομοιότητας.....	187
Εικόνα 4-18 – Συμπεριφορά αλγορίθμων συντακτικής ομοιότητας συμβολοσειρών	190
Εικόνα 4-19 – Βήματα τεχνικής συντακτικής ομοιότητας.....	191
Εικόνα 4-20 - Δεδομένα περίπτωσης χρήσης για εύρεση βέλτιστου αλγορίθμου σημασιολογικής ομοιότητας	194
Εικόνα 4-21 - Συμπεριφορά αλγορίθμων συντακτικής ομοιότητας συμβολοσειρών.....	197
Εικόνα 4-22 - Βήματα τεχνικής σημασιολογικής ομοιότητας.....	198
Εικόνα 4-23 - Βήματα τεχνικής ολικής οντολογικής αντιστοιχισής.....	200
Εικόνα 4-24 – Τεχνική σαρωτή FHIR	201
Εικόνα 4-25 – Παράδειγμα οντολογικής ευθυγράμμισης.....	202

ΚΕΦΑΛΑΙΟ 5

Εικόνα 5-1 – Εννοιολογικό διάγραμμα οντοτήτων συνόλου δεδομένων πρώτου πειράματος.....	207
Εικόνα 5-2 – Δεδομένα περίπτωσης χρήσης για το πρώτο πείραμα	208
Εικόνα 5-3 – Οντολογική αναπαράσταση δεδομένων περίπτωσης χρήσης για το πρώτο πείραμα	208
Εικόνα 5-4 - Παράδειγμα αποτελεσμάτων της τεχνικής συντακτικής ομοιότητας... ..	209

Εικόνα 5-5 - Παράδειγμα αποτελεσμάτων της τεχνικής συντακτικής ομοιότητας...	210
Εικόνα 5-6 - Απεικόνιση διαφοράς μεταξύ της υπολογισμένης αντιστοίχισης και της περίπτωσης ακριβούς αντιστοίχισης (χειροκίνητα αποτελέσματα)	213
Εικόνα 5-7 - Αποτελέσματα οντολογικής ευθυγράμμισης.....	215
Εικόνα 5-8 - Περιοχές εφαρμογής Apache Dozer.....	218
Εικόνα 5-9 - Εννοιολογικό διάγραμμα οντοτήτων συνόλου δεδομένων πρώτου πειράματος.....	219
Εικόνα 5-10 - Δεδομένα περίπτωσης χρήσης για το δεύτερο πείραμα.....	219
Εικόνα 5-11 - Οντολογική αναπαράσταση δεδομένων περίπτωσης χρήσης για το δεύτερο πείραμα	220
Εικόνα 5-12 - Απεικόνιση κριτηρίων και αποτελεσμάτων σύγκρισης.....	227
Εικόνα 5-13 - Απεικόνιση προόδου μετατροπής μικρού και μεγάλου συνόλου δεδομένων.....	228

Κατάλογος Πινάκων

ΚΕΦΑΛΑΙΟ 2

Πίνακας 2-1 - Πρότυπα στον τομέα της υγείας.....	89
--	----

ΚΕΦΑΛΑΙΟ 4

Πίνακας 4-1 - Βέλτιστα αποτελέσματα συντακτικής αντιστοίχισης για την απόσταση Levenshtein	187
Πίνακας 4-2 - Βέλτιστα αποτελέσματα συντακτικής αντιστοίχισης για την ομοιότητα Cosine.....	188
Πίνακας 4-3 - Βέλτιστα αποτελέσματα συντακτικής αντιστοίχισης για την ομοιότητα Jaccard.....	188
Πίνακας 4-4 - Βέλτιστα αποτελέσματα συντακτικής αντιστοίχισης για την απόσταση	188
Πίνακας 4-5 - Αποτελέσματα χειροκίνητης συντακτικής αντιστοίχισης	189
Πίνακας 4-6 - Κριτήρια σύγκρισης αλγορίθμων συντακτικής ομοιότητας	190
Πίνακας 4-7 - Βέλτιστα αποτελέσματα σημασιολογικής αντιστοίχισης για το Cortical.io.....	195
Πίνακας 4-8 - Βέλτιστα αποτελέσματα σημασιολογικής αντιστοίχισης για το Word2Vec	195
Πίνακας 4-9 - Βέλτιστα αποτελέσματα σημασιολογικής αντιστοίχισης για το Wordnet	195
Πίνακας 4-10 - Αποτελέσματα χειροκίνητης σημασιολογικής αντιστοίχισης	196
Πίνακας 4-11 - Κριτήρια σύγκρισης αλγορίθμων σημασιολογικής ομοιότητας	197

ΚΕΦΑΛΑΙΟ 5

Πίνακας 5-1 - Πόροι HL7 FHIR με υψηλότερους βαθμούς συντακτικής ομοιότητας	209
Πίνακας 5-2 - Πόροι HL7 FHIR με υψηλότερους βαθμούς σημασιολογικής ομοιότητας	210

Πίνακας 5-3 - Μεγαλύτερα υπολογισμένα ζεύγη των αποτελεσμάτων του Syntactic Similarity και του Semantic Similarity Identifier.....	211
Πίνακας 5-4 - Αποτελέσματα του χειροκίνητου μετασχηματισμού	213
Πίνακας 5-5 - Πόροι HL7 FHIR με υψηλότερους βαθμούς συντακτικής ομοιότητας	221
Πίνακας 5-6 - Πόροι HL7 FHIR με υψηλότερους βαθμούς σημασιολογικής ομοιότητας	222
Πίνακας 5-7 - Μεγαλύτερα υπολογισμένα ζεύγη των αποτελεσμάτων του Syntactic Similarity και του Semantic Similarity Identifier.....	223
Πίνακας 5-8 - Αποτελέσματα του χειροκίνητου μετασχηματισμού.....	225
Πίνακας 5-9 - Κριτήρια και αποτελέσματα σύγκρισης προσεγγίσεων	227
Πίνακας 5-10 - Σύγκριση αποτελεσμάτων περίπτωσης μικρού συνόλου δεδομένων	229
Πίνακας 5-11 - Σύγκριση αποτελεσμάτων περίπτωσης μεγάλου συνόλου δεδομένων	230

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Περιεχόμενα

- 1.1. Διαλειτουργικότητα Δεδομένων
- 1.2. Τομείς, Προκλήσεις και Όραμα
- 1.3. Αντικείμενο και Καινοτομία Διατριβής
- 1.4. Δομή Διατριβής

Περίληψη

Στο **ΚΕΦΑΛΑΙΟ 1**, πραγματοποιείται μία εισαγωγή στην έννοια της διαλειτουργικότητας δεδομένων σε γενικό πλαίσιο, εκφράζοντας την ανάγκη ύπαρξης της στο χώρο της υγείας. Ακολούθως, αναφέρονται οι σχετικές προκλήσεις σε συνδυασμό με τους τομείς που συναντάται η έννοια της διαλειτουργικότητας των δεδομένων, επισημαίνοντας το όραμα του μέλλοντος για έναν κόσμο αποτελούμενο από διαλειτουργικές υπηρεσίες που βασίζονται και χρησιμοποιούν δεδομένα που παράγονται από διαφορετικές και ετερογενείς πηγές. Παράλληλα, πραγματοποιείται αναφορά στην καινοτομία, στη συνεισφορά και στον τελικό στόχο της Διδακτορικής Διατριβής, πραγματοποιώντας μία σύντομη ανασκόπηση σε παλαιότερες και κομβικές προσωπικές έρευνες, που μπορούν να θεωρηθούν ως ορόσημα για το τελικό αποτέλεσμα της έρευνας.

1.1 Διαλειτουργικότητα Δεδομένων

Ο κόσμος γίνεται όλο και πιο μικρός καθώς περισσότεροι άνθρωποι συνδέονται μεταξύ τους μέσω δικτυακών υπολογιστικών συστημάτων [1]. Οι ενισχυμένες συνδέσεις επιταχύνουν τους χρόνους αντίδρασης καθώς οι άνθρωποι μπορούν να επικοινωνούν ευκολότερα, σημειώνοντας επίσης κέρδη καθώς διάφορα λογισμικά και συστήματα υπολογιστών συνδέονται μεταξύ τους για αυτοματοποιημένη κοινή χρήση δεδομένων. Με τόσα πολύπλοκα συστήματα που διασυνδέονται, τα ζητήματα διαλειτουργικότητας απασχολούν όσους εργάζονται στους διάφορους τύπους βιομηχανιών, καθώς κάθε βιομηχανία έχει ένα μοναδικό λεξιλόγιο, με την υγειονομική περίθαλψη να μην αποτελεί κάποια εξαίρεση. Το πρόβλημα είναι πως τα λεξιλόγια αυτά τυγχάνουν να αποκτούν μία διαφορετική έννοια, των οποίων η σημασία παρεξηγείται εύκολα, με αποτέλεσμα να καταλήγουν σε εσφαλμένη χρησιμοποίησή τους. Καθώς η τεχνολογία έχει εισχωρήσει σε όλους τους τομείς της καθημερινότητάς μας, έτσι έχει γίνει και με το σύνολο των εφαρμογών και των εργαλείων που χρησιμοποιούν οι διαφορετικοί πάροχοι για την υποστήριξη των υπηρεσιών της υγείας. Υπάρχουν χιλιάδες συστήματα που παρακολουθούν τα δεδομένα ασθενών, παρέχουν ή αναλύουν πληροφορίες, περιεχόμενο και πόρους, ή επιτρέπουν πιθανές ψηφιακές αξιολογήσεις [2]. Ενώ οι δυνατότητες αυτών των εργαλείων δημιουργούν συναρπαστικές ευκαιρίες τόσο για τους παρόχους υγείας όσο και για τους ασθενείς, μπορούν επίσης να οδηγήσουν σε πολλαπλές δυσκολίες λόγω της πληθώρας των δεδομένων υγείας, της ιδιωτικότητας των ασθενών, και των απομονωμένων δεδομένων. Στο σημείο αυτό είναι όπου η διαλειτουργικότητα τίθεται σε λειτουργία, καθώς υπάρχει ένα αυξανόμενο ενδιαφέρον όπως και μία πιεστική ανάγκη να βρεθούν τρόποι για να διευκολυνθούν και να επιλυθούν αυτές οι δυσκολίες.

Είναι λοιπόν απαραίτητο όλα τα διαθέσιμα συστήματα να επικοινωνούν με ομοιόμορφο και ενιαίο τρόπο μεταξύ τους, χωρίς να αντιμετωπίζουν κανενός είδους δυσκολία. Παράλληλα, χρειάζεται να γίνεται εύκολα η αναδιοργάνωση των συστημάτων και να καθίσταται εύκολη η πρόσβαση στην τεχνολογία από ένα μεγάλο πλήθος φορέων και προμηθευτών, με ταυτόχρονη αξιοποίηση της τεχνολογίας. Όλα αυτά επιτυγχάνονται χάρη στην ύπαρξη της διαλειτουργικότητας. Στην πράξη, η διαλειτουργικότητα αποτελεί την ικανότητα των πληροφοριακών συστημάτων να παρέχουν ανοιχτά, σαφή και προτυποποιημένα σημεία επικοινωνίας, ώστε να παράγονται πραγματικά ολοκληρωμένες υπηρεσίες [3]. Η έννοια της διαλειτουργικότητας, στην πραγματικότητα, δεν αποτελεί κάτι νέο. Στις ημέρες μας, η ανάγκη για την επίτευξη της διαλειτουργικότητας συνεχώς αυξάνεται, αφού παράλληλα η

τεχνολογία εξελίσσεται με ραγδαίους ρυθμούς, δημιουργώντας έτσι την ανάγκη εφαρμογής νέων προτυποποιήσεων, ώστε να ακολουθείται αυτή η πορεία σε μια κοινωνία όπου οι απαιτήσεις και οι ανάγκες της συνεχώς μεταβάλλονται. Διαφαίνεται έτσι, πως το ζήτημα της διαλειτουργικότητας αποτελεί ένα πολύπλοκο και πολυδιάστατο ζήτημα, που περιτριγυρίζεται από διοικητικούς, οργανωτικούς, νομικούς, λειτουργικούς και τεχνικούς περιορισμούς. Γενικότερα, για την έννοια της διαλειτουργικότητας, έχουν δοθεί πολλοί προσδιορισμοί. Ορισμένοι από αυτούς καταγράφονται στη συνέχεια:

- **IEEE [4]:** «*Η ικανότητα δύο ή περισσότερων συστημάτων ή βασικών στοιχείων να ανταλλάσσουν πληροφορίες και να τις επαναχρησιμοποιούν προς όφελός τους*»
- **ISO/IEC 2382-01 [5]:** «*Η δυνατότητα να επικοινωνούν, να εκτελούνται προγράμματα ή να μεταφέρονται δεδομένα μεταξύ ποικίλων λειτουργικών μονάδων, με τρόπο που απαιτεί τον χρήστη να έχει ελάχιστες έως καθόλου γνώσεις των χαρακτηριστικών αυτών των μονάδων*»
- **European Commission/IDABC [6]:** «*Η ικανότητα διαφορετικών πληροφοριακών και επικοινωνιακών συστημάτων, όπως ακριβώς συμβαίνει και με τις επιχειρησιακές διαδικασίες που υποστηρίζουν, να ανταλλάσσουν δεδομένα, όντας σε θέση να διαμοιράζονται πληροφορίες και γνώση*»
- **Wikipedia [7]:** «*Η δυνατότητα ενός προϊόντος ή συστήματος – του οποίου οι διεπαφές είναι πλήρως δημόσια τεκμηριωμένες – να συνδέεται και να λειτουργεί με άλλα προϊόντα ή συστήματα, χωρίς περιορισμούς στην πρόσβασή τους ή φραγμούς στην υλοποίηση*»

Βάσει των παραπάνω ορισμών, η διαλειτουργικότητα συνοπτικά, αποτελεί την ικανότητα διαφορετικών συστημάτων, συσκευών ή εφαρμογών να συνδέονται με συντονισμένο τρόπο εντός και εκτός των διαφορετικών οργανωτικών ορίων, με σκοπό την πρόσβαση, την ανταλλαγή και τη συνεργατική χρήση δεδομένων, βελτιώνοντας έτσι τις καθημερινές ανάγκες των εκάστοτε ατόμων και πληθυσμών, χωρίς να απαιτείται κάποια προσπάθεια από την πλευρά του τελικού χρήστη. Η έννοια της διαλειτουργικότητας (interoperability), πολλές φορές συγχέεται με τις «συγγενείς» έννοιες της ενσωμάτωσης (integration), της συνεργασίας (collaboration), της φορητότητας (portability) και της ανταλλαξιμότητας (interchangeability). Ωστόσο, υπάρχει ένα σύνολο ουσιαστικών διαφορών μεταξύ αυτών των εννοιών και της έννοιας της διαλειτουργικότητας, όπως παρουσιάζονται στη συνέχεια:

- **Interoperability - Integration:** Σε γενικές γραμμές, η διαλειτουργικότητα εμπεριέχει την έννοια της συνύπαρξης και της αυτονομίας. Με τη σειρά της η ενσωμάτωση, έχει την έννοια της στενής διασύνδεσης των εμπλεκόμενων οντοτήτων σε τέτοιο βαθμό, όπου δεν καθίσταται εφικτός ο διαχωρισμός αυτών από το σύνολο το οποίο συνθέτουν. Αντιθέτως, η διαλειτουργικότητα σχετίζεται με την έννοια της χαλαρής διασύνδεσης, μιας κατάστασης στην οποία οι εμπλεκόμενοι οργανισμοί και οντότητες μπορούν να ανταλλάσσουν υπηρεσίες και να συνεργάζονται για την παροχή υπηρεσιών, διατηρώντας ωστόσο την αυτονομία τους.
- **Interoperability - Collaboration:** Η διαλειτουργικότητα συσχετίζεται άμεσα με την συμβατότητα και την ικανότητα της συνεργασίας, μιας και δύο διαλειτουργούντες οργανισμοί είναι απαραίτητο να συνεργάζονται μεταξύ τους. Από την άλλη μεριά, αν δύο οργανισμοί συνεργάζονται, αυτό δε συνεπάγεται και το γεγονός πως θα μπορούν και να διαλειτουργούν μεταξύ τους.
- **Interoperability - Portability:** Η φορητότητα αφορά αποκλειστικά την ικανότητα της μεταφοράς συστημάτων, λογισμικού ή χρηστών από το περιβάλλον ενός οργανισμού στο περιβάλλον ενός άλλου. Αντιθέτως, η διαλειτουργικότητα αναφέρεται στην ικανότητα ενός συστήματος να χρησιμοποιεί και να κατανοεί πλήρως την πληροφορία που προέρχεται από ένα άλλο σύστημα.
- **Interoperability - Interchangeability:** Η ανταλλαξιμότητα αναφέρεται στην ικανότητα της συνέχισης παροχής του ίδιου επιπέδου σταθερής συμπεριφοράς υπηρεσιών παρά την αλλαγή ενός συστήματος ή ενός τμήματος αυτού. Όσον αφορά τη διαλειτουργικότητα, όμως, αυτός δεν είναι κάποιος από τους βασικούς σκοπούς της.

Γενικότερα, οι λειτουργίες της διαλειτουργικότητας περιλαμβάνουν την πρόσβαση σε δεδομένα, τη μετάδοση δεδομένων και τη δια-οργανωτική συνεργασία ανεξάρτητα από τον προγραμματιστή ή την προέλευσή τους. Παρόμοια με τη συμβατότητα, η διαλειτουργικότητα βοηθά τους οργανισμούς να επιτύχουν υψηλότερη απόδοση και να αποκτήσουν μία πιο ολιστική άποψη των πληροφοριών. Μέσω της διαλειτουργικότητας δίνεται η δυνατότητα πρόσβασης και επεξεργασίας δεδομένων αναγνώσιμων από μηχανές από πολλαπλές πηγές, μερικές φορές αυτόματα, χωρίς τα δεδομένα αυτά να χάνουν το νόημα ή την ακεραιότητά τους. Τα πρότυπα τεχνολογίας αποτελούν βασικούς παράγοντες που επιτρέπουν την επίτευξη της διαλειτουργικότητας.

Συλλογικά, επιτρέπουν την ενσωμάτωση διαφορετικών στοιχείων ενός συστήματος και την απρόσκοπτη συνεργασία τους. Σε τεχνικό επίπεδο, η διαλειτουργικότητα των δεδομένων απαιτεί «συντακτική διαλειτουργικότητα», η οποία βασίζεται στην ευρεία υιοθέτηση τυποποιημένων μορφών δεδομένων και στην υλοποίηση διεπαφών προγραμματισμού εφαρμογών (Application Programming Interfaces - API) και συνδέσμων που επιτρέπουν την πρόσβαση και την ενσωμάτωση δεδομένων από πολλαπλές πηγές. Η διαλειτουργικότητα απαιτεί επίσης την ανταλλαγή δεδομένων και πληροφοριών μεταξύ των συστημάτων χωρίς να χάνεται το πλαίσιο και η σημασία τους. Με άλλα λόγια, υπάρχει ανάγκη για «σημσιολογική διαλειτουργικότητα» σε όλα τα συστήματα δεδομένων. Η σημσιολογική διαλειτουργικότητα αναφέρεται στον τρόπο με τον οποίο τα στοιχεία δεδομένων περιγράφονται και αντιστοιχίζονται σε διαφορετικά πληροφοριακά συστήματα, προκειμένου να ελαχιστοποιηθεί η ασάφεια και να εξασφαλιστεί η κατάλληλη ερμηνεία των μεμονωμένων στοιχείων δεδομένων. Στο πλαίσιο αυτό, η ανάπτυξη κοινών λεξιλογίων και ταξινομήσεων χρήζει ζωτικής σημασίας.

Ειδικά στον τομέα της υγείας, τα τελευταία χρόνια παράγονται, συγκεντρώνονται και αποθηκεύονται τεράστιες ποσότητες δεδομένων που σχετίζονται με την υγεία σε Ηλεκτρονικούς Φακέλους Υγείας (Electronic Health Records - EHRs) [8] ή σε Προσωπικούς Φακέλους Υγείας (Personal Health Records - PHRs) [9] πολλαπλών μορφών, τα οποία συγκεντρώνουν ένα ευρύ και συχνά ποικίλο σύνολο δεδομένων, όπως ιατρικές εξετάσεις, ή βιοσήματα. Το τελευταίο είναι αποτέλεσμα της έκρηξης των διαθέσιμων υπηρεσιών Τεχνολογίας Πληροφοριών και Επικοινωνιών (ΤΠΕ) που αποσκοπούν στη συλλογή, στην παραγωγή και στην επεξεργασία των δεδομένων για την επίτευξη καλύτερης λήψης αποφάσεων ή τη βελτίωση της ποιότητας ζωής των πολιτών. Σύμφωνα με μία πρόσφατη έρευνα [10], η παγκόσμια αγορά φορητών ιατρικών συσκευών προβλέπεται να φθάσει τα 12,14 δισ. δολάρια μέχρι το 2021 από 5,31 δισ. δολάρια το 2016. Σύμφωνα με μια μελέτη [11], θα μπορούσαν να επιτευχθούν ετήσιες εξοικονομήσεις περίπου 78 δισ. δολαρίων εάν χρησιμοποιούνταν πρότυπα ανταλλαγής δεδομένων σε ολόκληρο τον τομέα της υγειονομικής περίθαλψης. Αυτές οι συσκευές έχουν σχεδιαστεί για να παρακολουθούν τις καθημερινές δραστηριότητες των ασθενών, να ανιχνεύουν τα πρότυπα ύπνου και τα συναισθήματά τους, καθώς και να καταγράφουν ζωτικά σήματά τους όπως είναι η αρτηριακή πίεση ή ο καρδιακός παλμός. Πολλές εφαρμογές και συστήματα έχουν αναπτυχθεί βασισμένα στα ιατρικά δεδομένα ή δεδομένα φορητών συσκευών, προκειμένου να ανιχνευθούν και να προβλεφθούν οι ανωμαλίες της υγείας των ασθενών, να αντιμετωπιστούν οι

ασθένειες και να μελετηθεί γενικά η ανθρώπινη συμπεριφορά. Ωστόσο, ενώ οι φορείς υγειονομικής περίθαλψης συλλέγουν ποικίλα ιατρικά δεδομένα, τα δεδομένα αυτά δεν ρέουν μεταξύ αυτών των οντοτήτων με συνεκτικό ή τυποποιημένο τρόπο. Η εξέλιξη λοιπόν στον τομέα της υγειονομικής περίθαλψης, προς τη λεγόμενη Βιομηχανία της Υγείας 4.0 (Healthcare 4.0) [12], περιλαμβάνει ένα ευρύ φάσμα συστημάτων συλλογής δεδομένων, συμπεριλαμβανομένων των ερευνών για την υγεία, των διοικητικών εγγραφών και των λογαριασμών, και των ιατρικών αρχείων που χρησιμοποιούνται από φορείς όπως νοσοκομεία, υγειονομικές αρχές, φαρμακεία ή ακόμα και πολίτες.

Είναι αδιαμφισβήτητο ότι οι περισσότερες υπηρεσίες και οντότητες υγειονομικής περίθαλψης είναι ετερογενείς και λειτουργούν ανεξάρτητα, ενώ έγκειται στην ευθύνη του εκάστοτε ενδιαφερόμενου να εμπιστευθεί και να εκμεταλλευτεί τις διάφορες πηγές για να λάβει τις απαιτούμενες πληροφορίες. Ωστόσο, στην περίπτωση αυτή, συχνά χάνονται ευκαιρίες για την επαναχρησιμοποίηση αυτών των δεδομένων για την παροχή προηγμένων υπηρεσιών υγειονομικής περίθαλψης καθώς και για ερευνητικούς σκοπούς, αφενός λόγω της έλλειψης διαλειτουργικότητας των δεδομένων και αφετέρου λόγω του γεγονότος ότι τα δεδομένα της υγειονομικής περίθαλψης περιορίζονται σε πακέτα δεδομένων χωρίς να πληρούνται οι απαιτήσεις των προτύπων, η ποιότητα και η ασφαλής ανταλλαγή τους. Επιπλέον, οι περισσότεροι από τους παραδοσιακούς αλγόριθμους εξόρυξης δεδομένων δεν είναι σε θέση να αντιμετωπίσουν την ετερογένεια των δεδομένων της υγειονομικής περίθαλψης [13], επομένως όλο και πιο συχνά χάνεται η δυνατότητα εντοπισμού πρώιμων συμπτωμάτων ασθενειών, καθιστώντας ακόμη πιο αποθαρρυντικό το έργο της διαλειτουργικότητας στον τομέα της υγείας. Καθίσταται λοιπόν ξεκάθαρο, πως το πρόβλημα δεν είναι μόνο η δυσκολία ανταλλαγής δεδομένων μεταξύ συστημάτων, αλλά και η ασυμβατότητα των δεδομένων. Η διαλειτουργικότητα είναι ο μόνος βιώσιμος τρόπος για να μπορούν οι φορείς που ενεργούν σε διάφορες τοποθεσίες, με διαφορετική εμπειρία και προοπτικές και χρησιμοποιούν ξεχωριστά συστήματα πληροφοριών, να συνεργάζονται και να παρέχουν ποιοτική υγειονομική περίθαλψη. Η παροχή και η έρευνα για την υγειονομική περίθαλψη σε πολλαπλά σημεία απαιτούν την αναδιάρθρωση των δεδομένων των EHR σε μια κοινή μορφή, με τυποποιημένες ορολογίες που συνδέονται με άλλες πηγές δεδομένων. Το κλειδί είναι να δημιουργηθεί μία προσέγγιση που θα φέρει τα διαφορετικά δεδομένα από μεμονωμένα συστήματα σε μια ενοποιημένη μορφή. Αυτή η ενοποίηση των δεδομένων έχει αποτελέσει τη βάση για τη δημιουργία της προτεινόμενης προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων της παρούσας Διδακτορικής Διατριβής.

1.2 Τομείς, Προκλήσεις και Όραμα Διαλειτουργικότητας Δεδομένων

Αναφορικά με τη διαλειτουργικότητα, υπάρχουν πολλαπλές κατηγορίες, ανάλογα με τον τομέα στον οποίο αυτή αναφέρεται, και προσπαθεί να επιλύσει το πρόβλημα της επικοινωνίας και ανταλλαγής των εκάστοτε δεδομένων. Ειδικότερα, ορισμένοι από τους πιο σημαντικούς τομείς διαλειτουργικότητας αναφέρονται στη συνέχεια:

- **Υγειονομική περίθαλψη:** Στο πλαίσιο του οικοσυστήματος της υγειονομικής περίθαλψης, η διαλειτουργικότητα χρησιμοποιείται για τη βελτιστοποίηση και την τυποποίηση της ποιότητας της ιατρικής περίθαλψης. Ο κλάδος αυτός βασίζεται σε πολλαπλές πηγές πληροφοριών, όπως δεδομένα από εργαστήρια, κλινικές, φαρμακεία, νοσοκομεία και φορείς παροχής πρωτοβάθμιας περίθαλψης, που χρησιμοποιούν πολλαπλά συστήματα για την καταγραφή αυτών των πληροφοριών. Ως εκ τούτου, τα συστήματα τεχνολογίας της πληροφόρησης για την υγεία πρέπει να είναι διαλειτουργικά, ώστε να επιτρέπεται η σχετική ανταλλαγή EHR και η παροχή ολοκληρωμένης υγειονομικής περίθαλψης στους ασθενείς.
- **Τηλεπικοινωνίες:** Στις τηλεπικοινωνίες, η διαλειτουργικότητα σχετίζεται με τις διάφορες υπηρεσίες που είναι σε θέση να λειτουργούν αποτελεσματικά και να δέχονται υπηρεσίες από άλλα συστήματα. Τα στοιχεία της συμβατότητας των τηλεπικοινωνιών αναφέρονται στη διαθεσιμότητα σήματος, στη κλίμακα του δικτύου, στις συχνότητες και στην κάλυψη του εξοπλισμού. Για παράδειγμα, απαιτείται διαλειτουργικότητα για την επιτυχή επικοινωνία αμφίδρομων ραδιοφωνικών σταθμών. Το ITU-T [14] είναι το πρότυπο που χρησιμοποιείται στις διεθνείς τηλεπικοινωνίες για την επίτευξη διαλειτουργικότητας.
- **Λογισμικό:** Η διαλειτουργικότητα στο λογισμικό αναφέρεται στη λειτουργικότητα διαφορετικών προγραμμάτων για την ανταλλαγή πληροφοριών, την κοινή χρήση αρχείων και τη χρήση των ίδιων πρωτοκόλλων. Η Java [15] θεωρείται συχνά ως μία εξαιρετικά διαλειτουργική γλώσσα προγραμματισμού για το λογισμικό, καθώς μπορεί να εκτελεστεί σε οποιοδήποτε πρόγραμμα με εικονική μηχανή Java (Java Virtual Machine - JVM) [16].

- **Αναζήτηση πληροφοριών:** Η διαλειτουργικότητα αναζήτησης πληροφοριών επιτρέπει στους χρήστες να συλλέγουν με ακρίβεια πληροφορίες από πολλές πηγές χρησιμοποιώντας μόνο ένα επερώτημα (query). Αυτό επιτρέπει στους οργανισμούς να κάνουν πιο αποτελεσματική τη χρήση μεγάλων ποσοτήτων δεδομένων, υποστηρίζοντας παράλληλα την εξόρυξη δεδομένων.
- **Στρατιωτική και Δημόσια ασφάλεια:** Στο πλαίσιο αυτό, η διαλειτουργικότητα ορίζεται ως η ικανότητα πολλαπλών τομέων επιβολής του νόμου να επικοινωνούν αποτελεσματικά κατά τη διάρκεια καταστάσεων έκτακτης ανάγκης, να τυποποιούν τις προσπάθειες κατάρτισης ή να συντονίζουν την εκτέλεση των καθηκόντων.
- **Διακυβέρνηση:** Η ανταλλαγή ηλεκτρονικών δεδομένων μεταξύ πολλών κυβερνητικών οργανισμών σε όλο τον κόσμο μπορεί να είναι δύσκολη. Η διαλειτουργικότητα συμβάλλει στην κάλυψη του χάσματος που δημιουργείται από τα γλωσσικά εμπόδια και τις διάφορες μορφολογικές προδιαγραφές που βοηθούν τις δημόσιες διοικήσεις να συνεργαστούν.

Παρόλο τους πολλαπλούς τομείς διαλειτουργικότητας, μεγάλη έμφαση τόσο ερευνητικά όσο και οικονομικά έχει αποδοθεί στην επίτευξη της διαλειτουργικότητας στον τομέα της υγείας. Το πρώτο EHR αναπτύχθηκε το 1972, καθιστώντας το ένα επαναστατικό βήμα προς την αλλαγή του τρόπου με τον οποίο επεξεργάζονται οι πληροφορίες των ασθενών. Ταυτόχρονα, ήταν πολύ υποτυπώδες και χρειάστηκε να περάσει ο χρόνος μέχρι τα μέσα της δεκαετίας του '80 ώστε τα συστήματα υγείας να αρχίσουν να κατανοούν την αξία που θα μπορούσαν να παρέχουν τα EHR και να ξεκινήσουν την ευρεία υιοθέτηση της τεχνολογίας. 49 χρόνια αργότερα δεν υπάρχει καμία αμφισβήτηση ότι έχουν γίνει τεράστια άλματα με την τεχνολογία της υγειονομικής περιθαλψης, η οποία βελτιώνεται καθημερινά μέσω πολλαπλών κλινικών καινοτομιών. Ωστόσο, με όλες αυτές τις εξελίξεις το επίμονο πρόβλημα της διαλειτουργικότητας μεταξύ των συστημάτων παραμένει. Τα προηγμένα κλινικά συστήματα παρέχουν πληθώρα γνώσεων κατ' απαίτηση για τους παρόχους, αλλά αυτή η γνώση είναι πολύ συχνά απομονωμένη και διαμοιρασμένη. Σύμφωνα με το Blackbook [17], το 70% των νοσοκομείων δεν χρησιμοποιεί δεδομένα εκτός του δικού τους EHR. Το ερώτημα που παραμένει είναι το γιατί με όλες τις τρέχουσες τεχνολογικές εξελίξεις ακόμα δεν έχει δημιουργηθεί ένας δομημένος τρόπος παρακολούθησης των πληροφοριών των ασθενών. Η απάντηση στον εν λόγω ερώτημα είναι πως μπορεί να διατίθεται απολύτως η τεχνολογία για την υλοποίησή του, αλλά υπάρχουν ορισμένες

προκλήσεις που καθιστούν αργή την υλοποίηση και την υιοθέτησή του. Η πρώτη πρόκληση αναφέρεται στην προτυποποίηση των πληροφοριών. Στις πρώτες ημέρες της ουσιαστικής χρήσης των διαφόρων προτύπων, όλα τα πρότυπα αποφεύγονταν στο χώρο της υγειονομικής περίθαλψης. Μπορεί να υπήρχαν ελάχιστα πρότυπα, αλλά δεν υπήρχε καμία προτυποποίηση. Τα συστήματα είχαν διαφορετικό τρόπο ταυτοποίησης των ασθενών, αποθήκευσης των δεδομένων, ή δημιουργίας σχεδίων φροντίδας. Τη σημερινή εποχή, οι εταιρείες συχνά χρησιμοποιούν πολλαπλά συστήματα για διαφορετικές πληροφορίες που συλλέγουν ή παρακολουθούν, και η προσπάθεια να συνδεθούν όλα τα συστήματα μεταξύ τους μπορεί να φαίνεται συντριπτική. Η δεύτερη πρόκληση αναφέρεται στην επιφυλακτικότητα σχετικά με τη χρήση των EHR. Ενώ ο αριθμός των διαφορετικών συστημάτων συνεχώς αυξάνεται, ακόμη και αν η ενσωμάτωση των συστημάτων μεταξύ τους δεν παράγει κάποιους τεχνικούς φραγμούς, υπάρχουν συχνά πολιτικά εμπόδια. Ορισμένοι πάροχοι υπηρεσιών υγείας συχνά αποφεύγουν την ιδέα της διασύνδεσης με εξωτερικά συστήματα, επειδή δε θεωρούν αξιόλογη την προκύπτουσα αξία στον οργανισμό τους. Ωστόσο, τα EHR αποτελούν ένα αναπόσπαστο κομμάτι του μέλλοντος αναφορικά με την υγειονομική περίθαλψη, μιας και επιτρέπουν την ευκολότερη πρόσβαση στα δεδομένα υγείας, βοηθούν στη μείωση των ιατρικών σφαλμάτων, έχουν τη δυνατότητα να συγκεντρώνουν και να αναλύουν ιατρικά δεδομένα υγείας από έναν ή και περισσότερους ασθενείς, και είναι ευκολότερο να διαμοιραστούν ταυτόχρονα με πολλαπλές υπηρεσίες υγείας. Αυτό έχει ως αποτέλεσμα, αργά ή γρήγορα να πρέπει να ενσωματωθεί και να χρησιμοποιηθεί από όλους. Η τρίτη πρόκληση και ο κύριος λόγος που το κενό της διαλειτουργικότητας υπάρχει ακόμα σχετίζεται με τον ίδιο λόγο που και τα ίδια τα EHR καθυστέρησαν να υιοθετηθούν, αναφερόμενοι στο ιδιαίτερα μεγάλο οικονομικό κόστος της ενσωμάτωσής τους. Τα παραδοσιακά μοντέλα ενσωμάτωσης είναι προσβάσιμα μόνο από τους μεγαλύτερους οργανισμούς υγειονομικής περίθαλψης. Ωστόσο, με τα νέα μοντέλα τιμολόγησης και τις στρατηγικές ενσωμάτωσης, η ενσωμάτωση καθίσταται πλέον ευκολότερη και πιο προσιτή από ποτέ.

Παρόλο που κάθε ιατρός και πάροχος υπηρεσιών υγείας έχουν διαφορετικούς στόχους, το όραμα οδηγεί πάντα στην φροντίδα και τη θεραπεία του ασθενούς. Μία από τις σημαντικότερες απαιτήσεις όσον αφορά την παροχή υψηλής ποιότητας φροντίδας είναι επίσης μία από τις πιο δύσκολες για επίτευξη, αναφερόμενοι στην αληθινή διαλειτουργικότητα μεταξύ των παρόχων, συστημάτων, τεχνολογιών και πληροφοριών υγειονομικής περίθαλψης, όπου κάθε ένα από αυτά τα μέρη πρέπει να επιτύχει τους αντίστοιχους στόχους. Η διαλειτουργικότητα αποτελεί ένα βασικό πρόβλημα για όλη την υγειονομική

περίθαλψη, καθώς οι διαφορετικοί προμηθευτές συστημάτων, οι πολύπλοκες διαδικασίες και τα κλινικά πρότυπα δημιουργούν τεράστια εμπόδια στην παροχή βελτιωμένων εμπειριών και καλύτερων αποτελεσμάτων περίθαλψης ασθενών. Το αυξανόμενο κόστος, η αυξημένη πίεση για βελτίωση της εμπειρίας των ασθενών και η μετάβαση σε φροντίδα με βάση την αξία, υποδηλώνουν ότι οι πάροχοι (και ολόκληρη η βιομηχανία υγειονομικής περίθαλψης) πρέπει να κινηθούν προς τη βελτίωση της διαλειτουργικότητας της υγειονομικής περίθαλψης. Η τεχνολογία της πληροφορικής στον τομέα της υγείας, οι δημιουργοί των πληροφοριακών συστημάτων, τα υπόλοιπα εμπλεκόμενα μέρη της υγειονομικής περίθαλψης (συμπεριλαμβανομένων των πληρωτών) και οι πάροχοι, πρέπει να μπορούν να επικοινωνούν και να συνδυάζουν τα υφιστάμενα δεδομένα για την επίτευξη ενός διαλειτουργικού συστήματος υγειονομικής περίθαλψης. Επιπρόσθετα από τις προκλήσεις που αντιμετωπίζει η διαλειτουργικότητα στον τομέα της υγείας, υπάρχουν επιπλέον προκλήσεις που αφορούν αποκλειστικά την υγειονομική περίθαλψη.

Η πρώτη πρόκληση αναφέρεται στην έλλειψη συνέπειας κατά την ταυτοποίηση των ασθενών. Επί του παρόντος, δεν υπάρχει συνεπής τρόπος ταυτοποίησης ενός ασθενούς σε όλο το φάσμα της υγειονομικής περίθαλψης, σε όλα τα μεγάλα συστήματα υγειονομικής περίθαλψης ή ακόμη και σε ένα δίκτυο παρόχων. Γενικότερα, οι ασθενείς εντοπίζονται και ταυτοποιούνται με το όνομά τους, την ημερομηνία γέννησης, ή τον αριθμό κοινωνικής ασφάλισής τους. Το πρόβλημα που προκύπτει είναι πως κάθε περίπτωση αυτών των πληροφοριών αποθηκεύεται με διάφορους τρόπους σε διαφορετικά συστήματα, πράγμα που σημαίνει ότι τα σφάλματα ταυτοποίησης του ασθενούς είναι πολύ πιθανό να πραγματοποιηθούν. Σύμφωνα με πολλές ομάδες υποστήριξης ασθενών, μια πιθανή λύση στο πρόβλημα αυτό είναι η δημιουργία ενός εθνικού, μοναδικού αναγνωριστικού ασθενούς. Το αναγνωριστικό ασθενούς θα είναι παρόμοιο με τον αριθμό κοινωνικής ασφάλισης ενός ατόμου, δεδομένου ότι θα είναι δικός του καθ' όλη τη διάρκεια ζωής του και θα χρησιμοποιείται σε κάθε σημείο φροντίδας. Το βασικό όφελος ενός αναγνωριστικού ασθενούς θα έγκειτο στο γεγονός ότι θα ήταν ένας κωδικός που ταξινομεί, κατηγοριοποιεί και αναγνωρίζει ένα άτομο ανεξάρτητα από το σύστημα ή τον πάροχο που χρησιμοποιείται. Ωστόσο, οι προσπάθειες για την προώθηση μιας τέτοιας προσπάθειας οδηγεί σε ένα αδιέξοδο. Για χρόνια, ποικίλοι οργανισμοί, συμπεριλαμβανομένων των Healthcare Information and Management Systems Society (HIMSS) [18] και College of Healthcare Information Management Executives (CHIME) [19], έχουν επιμείνει για τη δημιουργία ενός εθνικού αναγνωριστικού ασθενούς. Θεωρούν ότι καθώς οι ανταλλαγές δεδομένων για την υγεία συνεχίζουν να

εξελίσσονται, η ανάγκη για έναν συνεπή και ακριβή τρόπο ταυτοποίησης των αρχείων υγείας των ασθενών καθίσταται όλο και πιο απαραίτητη. Είναι δεδομένο πως η έλλειψη ενός εθνικά ομοιόμορφου συστήματος ταυτοποίησης ασθενών έχει οδηγήσει σε αυξημένα κόστη υγειονομικής περίθαλψης, στην αναποτελεσματικότητα της ανταλλαγής δεδομένων υγείας και στις απειλές για την ασφάλεια των ασθενών, μεταξύ άλλων προβλημάτων. Αυτήν την προσπάθεια ήρθε να ενισχύσει η δημιουργία του Νόμου περί Φορητότητας και Υπευθυνότητας Ασφάλισης Υγείας του 1996 (Health Insurance Portability and Accountability Act - HIPAA) που ζήτησε τη δημιουργία ενός μοναδικού αναγνωριστικού ασθενούς. Ωστόσο, το Κογκρέσο απέρριψε στη συνέχεια αυτή την εντολή λόγω ανησυχιών σχετικά με την ιδιωτικότητα των ασθενών. Μια μειοψηφία ανθρώπων φοβάται ότι οι κυβερνήσεις ή η βιομηχανία θα εκμεταλλευτούν τα δεδομένα των ασθενών και ότι η ιδιωτική ζωή θα μπορούσε να είναι ακόμη πιο δύσκολο να προστατευθεί αν έπρεπε να χρησιμοποιηθεί ένας αναγνωριστικός κωδικός. Ωστόσο, χωρίς ένα μοναδικό αναγνωριστικό, είναι δύσκολο να συνδεθούν τα διαφορετικά δεδομένα για την απόκτηση μιας ολοκληρωμένης εικόνας ασθενούς αναφορικά με την υγειονομική περίθαλψή του. Έτσι, προτού να μπορέσει να γίνει πραγματικότητα η βιομηχανική διαλειτουργικότητα, οι ομοσπονδιακές υπηρεσίες θα πρέπει να δώσουν προτεραιότητα στην τυποποίηση όλων των πτυχών της ανταλλαγής αρχείων υγείας των ασθενών.

Η δεύτερη πρόκληση αναφέρεται στην έλλειψη προτύπων για την αποστολή, τη λήψη και τη διαχείριση πληροφοριών μεταξύ των διαφορετικών συστημάτων υγείας. Επί του παρόντος, η τεχνολογία υγειονομικής περίθαλψης που προσφέρουν οι σημερινοί προμηθευτές καθιστά δύσκολη την απλή αντιγραφή ή ανταλλαγή πληροφοριών από ένα ηλεκτρονικό λογισμικό καταγραφής υγείας ή άλλη τεχνολογία υγειονομικής περίθαλψης σε άλλη. Οι λανθασμένες γραμματοσειρές, τα εξωτερικά πεδία δεδομένων και οι προσωποποιημένες μορφές αναπαράστασης των δεδομένων υποδηλώνουν πως τα δεδομένα πρέπει να υφίστανται επεξεργασία πριν εισαχθούν σε κάποιο άλλο σύστημα. Η υιοθέτηση και η χρήση των προτύπων δεδομένων για την υγεία αποτελούν τη βάση για τη διευκόλυνση της διαλειτουργικότητας μεταξύ των οργανισμών και μεταξύ των ηλεκτρονικών συστημάτων υγείας. Σύμφωνα με το Γραφείο του Εθνικού Συντονιστή (Office of the National Coordinator - ONC), *«τα πρότυπα είναι συμφωνημένες μέθοδοι για τη σύνδεση συστημάτων. Τα πρότυπα μπορούν να αφορούν την ασφάλεια, τη μεταφορά δεδομένων, τη μορφή δεδομένων ή τη δομή ή τις έννοιες των κωδικών ή των όρων»*. Με βάση τα πρότυπα, οι ανταλλαγές πληροφοριών για την υγεία, τα προγνωστικά συστήματα ανάλυσης και άλλες προσπάθειες ανταλλαγής πληροφοριών μπορούν να

γίνουν πραγματικότητα. Σε ολόκληρη την υγειονομική περίθαλψη, διάφοροι οργανισμοί ανάπτυξης διαφορετικών προτύπων δημιουργούν, καθορίζουν, επικαιροποιούν και διατηρούν πρότυπα δεδομένων για την υγεία μέσω συνεργατικών διαδικασιών που περιλαμβάνουν χρήστες πληροφορικής για την υγεία, χωρίς όμως να έχουν δημιουργήσει κάποιο ενιαίο πρότυπο. Με τις μεγάλες δαπάνες που πραγματοποιήθηκαν για την εφαρμογή των EHR, ένα σύστημα υγείας πρέπει να βρει πιο αποτελεσματικούς τρόπους για να συνδέσει τα κατακερματισμένα δεδομένα των ασθενών. Ωστόσο, εφόσον δεν υπάρχει κάποιο ενιαίο πρότυπο διαλειτουργικότητας, προκαλούνται πολλές καθυστερήσεις και ανακρίβειες, κάτι που καθίσταται ένα σημαντικό εμπόδιο στις προσπάθειες επίτευξης διαλειτουργικότητας.

Η τρίτη πρόκληση αναφέρεται στη δυσκολία στη μέτρηση, στην ανάλυση και στη βελτίωση της διαλειτουργικότητας μεταξύ των συστημάτων υγειονομικής περίθαλψης. Συνοπτικά, δεν μπορεί να βελτιωθεί κάτι που δε μπορεί να μετρηθεί. Ως εκ τούτου, είναι δύσκολο να ποσοτικοποιηθεί το κόστος, το ποσοστό σφάλματος και άλλα ζητήματα που δημιουργούνται όταν τα συστήματα υγειονομικής περίθαλψης δε μιλούν μεταξύ τους με ουσιαστικό τρόπο. Χωρίς τη δυνατότητα μέτρησης ή παρακολούθησης των αποτελεσμάτων, τα συστήματα υγείας και οι υπάρχοντες οργανισμοί υγειονομικής περίθαλψης δεν μπορούν να βελτιώσουν τις σημαντικότερες διαδικασίες, καθιστώντας μη επιτεύξιμη την πραγματική διαλειτουργικότητα. Η πρόκληση που αντιμετωπίζουν οι πάροχοι και οι προμηθευτές είναι η έλλειψη ενός τυποποιημένου τρόπου μέτρησης του αντίκτυπου των καθυστερήσεων και της αποτυχίας στη διαλειτουργικότητα. Μόλις οι πάροχοι θα είναι σε θέση να μετρήσουν τα ζητήματα σε ολόκληρη την αλυσίδα υγειονομικής περίθαλψης, τότε και μόνο τότε θα μπορούν να αρχίσουν να αναλύουν τυχόν προβληματικούς τομείς και να πραγματοποιούν αλλαγές ή ακόμα και να παρακολουθούν πώς αυτές οι αλλαγές μπορούν να βελτιώσουν την ποιότητα της φροντίδας και της παρακολούθησης των ασθενών. Παράλληλα, η ανάλυση δεδομένων έγινε γρήγορα μια από τις πιο δύσκολες επιχειρήσεις για την υγειονομική περίθαλψη. Οι πάροχοι εξακολουθούν να προσπαθούν να εξοικειωθούν με τη διαχείριση και τη λειτουργία των ηλεκτρονικών αρχείων υγείας τους, ενώ ταυτόχρονα καλούνται να αντλήσουν στοιχεία από τα κλινικά δεδομένα. Για τους οργανισμούς που μπορούν να ενσωματώσουν με επιτυχία τις πληροφορίες που βασίζονται σε δεδομένα και να τις φέρνουν στις κλινικές και επιχειρησιακές διαδικασίες τους, οι ανταμοιβές είναι άφθονες, αναφερόμενοι σε περισσότερους υγιείς ασθενείς, χαμηλότερο κόστος, αυξημένη απόδοση και υψηλότερο επίπεδο ικανοποίησης του προσωπικού και των ασθενών.

Η τέταρτη πρόκληση αναφέρεται στις πληροφορίες, των οποίων η μετάδοση και η μεταφορά εμποδίζονται από τους ίδιους του παρόχους. Ο διαμοιρασμός των δεδομένων στην υγειονομική περίθαλψη ή, καλύτερα, η έλλειψη διαμοιρασμού δεδομένων στην υγειονομική περίθαλψη είναι ένα πρόβλημα. Ένα σημαντικό εμπόδιο στη διαλειτουργικότητα της υγειονομικής περίθαλψης είναι η πρακτική της παρεμπόδισης των πληροφοριών - από ορισμένους παρόχους υγείας. Σε περίπτωση βέβαια που οι πληροφορίες αυτές μοιράζονται, ορισμένοι από τους παρόχους χρεώνουν υψηλά τέλη για τη μετάδοση των δεδομένων τους εκτός του προσωπικού τους συστήματος. Παρόλο που η πρακτική αυτή προκαλεί έντονες διαταραχές στη ροή των δεδομένων περί της υγειονομικής περίθαλψης, πολλοί προμηθευτές συστημάτων συνεχίζουν την ίδια πρακτική παρά τις προσπάθειες για να την θέσουν σε παύση. Για παράδειγμα, ο Joel White, εκτελεστικός διευθυντής της Health IT Now [20], μιας συμμαχίας εταιρειών υγειονομικής περίθαλψης και τεχνολογίας, ζητά τη δημοσίευση οποιουδήποτε κανονισμού σχετικά με την παρεμπόδιση των πληροφοριών, όπως απαιτείται από τη νομοθεσία του 21ου αιώνα για τη θεραπεία του καρκίνου. Σε μια δημοσίευση που δημοσιεύθηκε πρόσφατα στο STAT [21], ο Joel White έγραψε πως *«περισσότερο από 600 ημέρες μετά την έναρξη ισχύος του νόμου περί θεραπειών, δεν έχει εκδοθεί ενιαία ρύθμιση σχετικά με την παρεμπόδιση των πληροφοριών»*. Η δημοσίευση του κανονισμού σχετίζεται άμεσα με το ONC, το οποίο έχει αναλάβει την εκπλήρωση συγκεκριμένων καθηκόντων. Κύρια στοιχεία της νομοθεσίας περιλαμβάνουν την ενθάρρυνση της διαλειτουργικότητας των ηλεκτρονικών αρχείων υγείας και την πρόσβαση των ασθενών στα δεδομένα υγείας, ενώ αποθαρρύνουν την παρεμπόδιση των πληροφοριών. Ο νόμος λαμβάνει μεγάλη προσοχή λόγω της δυναμικής ικανότητάς του να απελευθερώνει τα εμπόδια που δημιουργούν πρόβλημα στην πραγματοποίηση της διαλειτουργικότητας των δεδομένων.

1.3 Αντικείμενο και Καινοτομία Διατριβής

Η παρούσα Διδακτορική Διατριβή στοχεύει στη δημιουργία μιας προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων αυτοματοποιημένου μετασχηματισμού ανομοιογενών δεδομένων, η οποία θα μετασχηματίζει τόσο σε σημασιολογικό όσο και σε συντακτικό επίπεδο, δεδομένα διαφορετικής αναπαράστασης και μορφής, προκειμένου αυτά να μπορούν να ταυτοποιηθούν, να αντιστοιχηθούν σε μία ενιαία γλώσσα, και στο τέλος να συγχωνευτούν σε ένα κοινό επίπεδο. Ειδικότερα, η παρούσα έρευνα αφορά σε αρχικό στάδιο τη μελέτη των διαφόρων τεχνικών που έχουν υλοποιηθεί σχετικά με το πώς θα συλλέγονται και θα «γίνονται εκμεταλλεύσιμα» τα δεδομένα που παράγονται από ετερογενείς πηγές, για διαφορετικούς σκοπούς από αυτούς που είχαν παραχθεί να ικανοποιούν. Στα αρχικά στάδια της έρευνας, αποδόθηκε περισσότερο ενδιαφέρον στο πώς είναι εφικτό να συγκεντρωθούν δεδομένα από μία συγκεκριμένη πηγή, πώς σχετίζονται με τον αντίστοιχο κύκλο ζωής τους, και πώς χρησιμοποιούν τις εξαγόμενες πληροφορίες σε συνδυασμό με τα ακατέργαστα δεδομένα άγνωστων πηγών, ώστε τα άγνωστα δεδομένα να ταυτοποιηθούν και να χρησιμοποιηθούν για διαφορετικό σκοπό από αυτόν για τον οποίο δημιουργήθηκαν. Ωστόσο, ύστερα από μελέτη σχετικών ερευνών, περισσότερο ενδιαφέρον αποδόθηκε στον τομέα της «ηλεκτρονικής υγείας» (electronic Health - eHealth). Για το λόγο αυτό, έγινε εμβάθυνση στους μηχανισμούς που έχουν υλοποιηθεί ώστε δεδομένα υγείας και πληροφορίες να μπορούν να ανταλλάσσονται διαλειτουργικά μεταξύ συστημάτων που ανήκουν σε διαφορετικές ομάδες ή υπακούουν σε διαφορετικά πρότυπα υγείας. Κατόπιν σχετικής έρευνας και πειραματικών δοκιμών, υλοποιήθηκε μία αρχική προσέγγιση η οποία ήταν ικανή να διαχωρίζει τα δεδομένα υγείας σε κατηγορίες ανάλογα με τον τρόπο αναπαράστασής τους, και στη συνέχεια εξάγοντας γνώσεις, πληροφορίες και μεταδεδομένα, τα δεδομένα αυτά μετασχηματίζονταν σε μία ενιαία γλώσσα, η οποία διατηρούσε συνδυαστικά στοιχεία από άλλες ήδη υπάρχουσες προτυποποιημένες γλώσσες αναπαράστασης δεδομένων υγείας. Ακολούθως, ο συγκεκριμένος μηχανισμός μετάφρασης προσαρμόστηκε ώστε να παράγει οντολογίες μέσω συγκεκριμένων επιπέδων μεταμοντέλου και οντολογιών, με απώτερο στόχο την κοινή αναπαράσταση των δεδομένων σε μία ενιαία μορφή. Χρησιμοποιώντας ως γνώμονα τις παραχθείσες οντολογίες, η επόμενη προσέγγιση αφορούσε αφενός τη δημιουργία των οντολογιών, και αφετέρου τη σύγκρισή τους με τις οντολογίες των επιμέρους πόρων (resources) του προτύπου υγείας Health Level 7 (HL7) Fast Healthcare Interoperable Resources (FHIR) [22], λόγω της παγκόσμιας διάδοσης και υιοθέτησής του. Η σύγκριση αυτή αφορούσε τόσο τη συντακτική όσο και τη σημασιολογική απεικόνιση των εκάστοτε οντολογιών,

ώστε στο τέλος να καθίσταται εφικτή η αντιστοίχιση των οντολογιών των δεδομένων υγείας με τις οντολογίες των πόρων HL7 FHIR. Η τελική προσέγγιση βασίστηκε στην προαναφερθείσα λύση, όπου τόσο οι μηχανισμοί υλοποίησης των οντολογιών, αλλά και της συντακτικής και της σημασιολογικής σύγκρισης και αντιστοίχισης βελτιώθηκαν, βάσει σχετικών πειραματικών δοκιμών.

Βάσει λοιπόν όλων των προαναφερθέντων, η συνεισφορά της παρούσας Διδακτορικής Διατριβής έγκειται στο γεγονός πως προτείνεται μία αυτοματοποιημένη προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων, ώστε δεδομένα διαφορετικής αναπαράστασης και μορφής να μπορούν να συγχωνεύονται σε ένα ενιαίο επίπεδο. Εστιάζοντας στο χώρο της υγείας, η ενιαία γλώσσα και επίπεδο που χρησιμοποιήθηκαν είναι η HL7 FHIR. Συνοπτικά, η HL7 FHIR συνδυάζει τα χαρακτηριστικά προηγούμενων προτύπων υγείας (HL7 version 2 (v2) [23], HL7 version 3 (v3) [24], και HL7 Clinical Document Architecture (CDA) [25]), μαζί με τις πιο προηγμένες τεχνολογίες υπηρεσιών ιστού. Βασίζεται σε συγκεκριμένους πόρους, οι οποίοι μπορούν να συνδυαστούν ώστε να επιλύσουν κλινικά και διαδικαστικά προβλήματα κατά έναν πρακτικό τρόπο. Οι περισσότερες από τις ήδη υπάρχουσες πρακτικές και τεχνικές διαλειτουργικότητας δεδομένων υγείας, στηρίζονται σε πρότυπα υγείας όπως και σε μηχανισμούς οντολογικών μετασχηματισμών και αντιστοιχίσεων, ωστόσο είναι σχεδιασμένες για συγκεκριμένες περιπτώσεις χρήσης με ήδη προκαθορισμένη τη μορφή με την οποία πρέπει να εισέρχονται τα δεδομένα, χωρίς να έχουν τη δυνατότητα προσαρμογής, μη μπορώντας έτσι να θεωρηθούν ως ολιστικές λύσεις. Το συγκεκριμένο κενό έρχεται να καλύψει η παρούσα Διδακτορική Διατριβή, αποδίδοντας μια καινοτόμα γενικευμένη προσέγγιση η οποία μπορεί να εφαρμοστεί αυτοματοποιημένα σε οποιοδήποτε σύνολο δεδομένων υγείας, χωρίς την ανάγκη προ-επεξεργασίας τους, αποδίδοντας και αντιστοιχίζοντάς τα σε ένα από τα πιο ισχυρά, ευρέως διαδεδομένο, και πολλά υποσχόμενο πρότυπο υγείας, το HL7 FHIR.

1.4 Δομή Διατριβής

Λαμβάνοντας υπόψιν τις προκλήσεις στο χώρο της υγείας, της διαλειτουργικότητας των δεδομένων, αλλά και των σχετικών ερευνών, η παρούσα Διδακτορική Διατριβή έχει δομηθεί ως ακολούθως:

- Στο **ΚΕΦΑΛΑΙΟ 1**, πραγματοποιείται μία εισαγωγή στην έννοια της διαλειτουργικότητας δεδομένων σε γενικό πλαίσιο, εκφράζοντας την ανάγκη ύπαρξής της στο χώρο της υγείας. Ακολούθως, αναφέρονται οι σχετικές προκλήσεις σε συνδυασμό με τους τομείς που συναντάται η έννοια της διαλειτουργικότητας των δεδομένων, επισημαίνοντας το όραμα του μέλλοντος για έναν κόσμο αποτελούμενο από διαλειτουργικές υπηρεσίες που βασίζονται και χρησιμοποιούν δεδομένα που παράγονται από διαφορετικές και ετερογενείς πηγές. Παράλληλα, πραγματοποιείται αναφορά στην καινοτομία, στη συνεισφορά και στον τελικό στόχο της Διδακτορικής Διατριβής, πραγματοποιώντας μία σύντομη ανασκόπηση σε παλαιότερες και κομβικές προσωπικές έρευνες, που μπορούν να θεωρηθούν ως ορόσημα για το τελικό αποτέλεσμα της έρευνας.
- Στο **ΚΕΦΑΛΑΙΟ 2**, αναφέρεται η σχετική μελέτη που πραγματοποιήθηκε στο χώρο της διαλειτουργικότητας δεδομένων, και των σχετικών τεχνικών της που οδήγησαν στην τελική προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων. Για το λόγο αυτό, γίνεται αναφορά αρχικά στο χώρο της διαλειτουργικότητας δεδομένων, αναφέροντας τις κατηγορίες στις οποίες χωρίζεται, καθώς και ορισμένα από τα πιο γνωστά πρότυπα και οργανισμούς προτυποποίησης. Σε δεύτερο στάδιο, γίνεται αναφορά στη διαλειτουργικότητα στο χώρο της υγείας – η οποία θα μας απασχολήσει στην παρούσα έρευνα. Αναφορικά με τον τομέα αυτό, επισημαίνονται η ανάγκη ύπαρξής της, οι σχετικές κατηγορίες της, ορισμένα από τα πιο γνωστά πρότυπα υγείας για τη διαλειτουργική μεταφορά δεδομένων και κωδικοποιήσεων, ενώ πραγματοποιείται εκτενής μελέτη σε αρκετές έρευνες, επισημαίνοντας και αναλύοντας τις τεχνικές οι οποίες επηρέασαν την παρούσα Διδακτορική Διατριβή.
- Στο **ΚΕΦΑΛΑΙΟ 3**, αναφέρονται οι βασικές τεχνολογίες που χρησιμοποιήθηκαν προς την υλοποίηση της παρούσας προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων, αναφορικά με την ύπαρξη των οντολογιών, της ανάγκης τους και της χρήσης τους στο χώρο της υγείας. Παράλληλα, αναλύονται και τεχνικές οι οποίες στοχεύουν στη δημιουργία οντολογιών. Ακολούθως, γίνεται αναφορά στον τομέα της επεξεργασίας

φυσικής γλώσσας, αναφέροντας τη σημασία των τεχνικών συντακτικής και σημασιολογικής ομοιότητας μεταξύ λεκτικών και κειμένων, ένας κλάδος ο οποίος επίσης επηρέασε την τελική προσέγγιση για την επίτευξη των τελικών οντολογικών αντιστοιχίσεων. Εν τέλει, επισημαίνονται και παρουσιάζονται τεχνικές εύρεσης τόσο συντακτικής όσο και σημασιολογικής ομοιότητας μεταξύ λεκτικών και κειμένων.

- Στο **ΚΕΦΑΛΑΙΟ 4**, αναλύεται η προτεινόμενη προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων συμπεριλαμβανομένων των αρχικών ερευνών που πραγματοποιήθηκαν στο πλαίσιο της Διδακτορικής Διατριβής, σε συνδυασμό με τα αποτελέσματά τους. Περιγράφεται κάθε τεχνική της παρούσας προσέγγισης, αναφορικά με τη δημιουργία των οντολογιών, την αποθήκευσή τους, τη συντακτική και σημασιολογική αντιστοίχισή τους, καθώς και την τελική απεικόνισή τους. Το σύνολο αυτών των τεχνικών οδηγεί εν τέλει στη λειτουργικότητα της προσέγγισης, και την τελική αντιστοίχιση των δεδομένων υγείας στο πρότυπο υγείας HL7 FHIR.
- Στο **ΚΕΦΑΛΑΙΟ 5**, παρατίθεται η αξιολόγηση της προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων, μέσω δύο διαφορετικών περιπτώσεων χρήσης. Η πρώτη περίπτωση χρήσης σχετίζεται με ένα σύνολο δεδομένων υγείας από ανώνυμους ασθενείς ενός Ινστιτούτου Υγείας, στο οποίο περιγράφονται όλα τα εξαγόμενα αποτελέσματα της προσέγγισης, ακολουθούμενα από σχετική συζήτηση και ανάλυσή τους. Η δεύτερη περίπτωση χρήσης σχετίζεται πάλι με ένα σύνολο δεδομένων από έναν ανώνυμο ασθενή ο οποίος προσφέρει τα δεδομένα του, όπως προήλθαν από μια φορητή συσκευή. Τα βήματα της προσέγγισης επίσης αναλύονται, καταλήγοντας σε παρόμοια συμπεράσματα, ενώ παράλληλα πραγματοποιείται και αναλύεται μία συγκριτική αξιολόγηση της παρούσας προσέγγισης με μία άλλη προσέγγιση μετασχηματισμού και αντιστοίχισης δεδομένων.
- Στο **ΚΕΦΑΛΑΙΟ 6**, πραγματοποιείται η σύνοψη της Διδακτορικής Διατριβής, και των βασικών συνεισφορών της, ενώ περιγράφονται τα ανοιχτά ερευνητικά θέματα και οι μελλοντικοί στόχοι όπως προέκυψαν από την παρούσα έρευνα, αλλά και από το χώρο της υγείας γενικότερα, αναφορικά τόσο με τον τομέα της διαλειτουργικότητας, όσο και τον τομέα της ένταξης και χρήσης των οντολογιών στο χώρο της υγείας.

ΚΕΦΑΛΑΙΟ 2

Διαλειτουργικότητα Δεδομένων

Περιεχόμενα

- 2.1. Δεδομένα και Διαλειτουργικότητα Δεδομένων
 - 2.1.1. Ανάγκη Διαλειτουργικότητας και Δεδομένων
 - 2.1.2. Κατηγορίες Διαλειτουργικότητας
 - 2.1.3. Πρότυπα Διαλειτουργικότητας
 - 2.1.4. Τεχνικές Διαλειτουργικότητας
- 2.2. Διαλειτουργικότητα στον τομέα της Υγείας
 - 2.2.1. Ανάγκη Διαλειτουργικότητας στον τομέα της Υγείας
 - 2.2.2. Κατηγορίες Διαλειτουργικότητας στον τομέα της Υγείας
 - 2.2.3. Πρότυπα Διαλειτουργικότητας στον τομέα της Υγείας
 - 2.2.4. Τεχνικές Διαλειτουργικότητας στον τομέα της Υγείας

Περίληψη

Στο **ΚΕΦΑΛΑΙΟ 2**, αναφέρεται η σχετική μελέτη που πραγματοποιήθηκε στο χώρο της διαλειτουργικότητας δεδομένων, και των σχετικών τεχνικών της που οδήγησαν στην τελική προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων. Για το λόγο αυτό, γίνεται αναφορά αρχικά στο χώρο της διαλειτουργικότητας δεδομένων, αναφέροντας τις κατηγορίες στις οποίες χωρίζεται, ορισμένα από τα πιο γνωστά πρότυπα και οργανισμούς προτυποποίησης, καθώς και ορισμένες κομβικές τεχνικές επίτευξης διαλειτουργικότητας. Παράλληλα, λόγω της αναφοράς στο χώρο των δεδομένων, περιγράφεται το φαινόμενο των μεγάλων δεδομένων στην καθημερινότητά μας. Σε δεύτερο στάδιο, γίνεται αναφορά στη διαλειτουργικότητα στο χώρο της υγείας. Αναφορικά με τον τομέα αυτό, επισημαίνονται η ανάγκη ύπαρξής της, οι σχετικές κατηγορίες της, ορισμένα από τα πιο γνωστά πρότυπα υγείας για τη διαλειτουργική μεταφορά δεδομένων και κωδικοποιήσεων, ενώ γίνεται εκτενής μελέτη σε αρκετές έρευνες, επισημαίνοντας και αναλύοντας τις τεχνικές οι οποίες επηρέασαν την παρούσα Διδακτορική Διατριβή.

2.1 Δεδομένα και Διαλειτουργικότητα Δεδομένων

2.1.1 Ανάγκη Διαλειτουργικότητας και Δεδομένων

Για κάθε διαφορετική κατηγορία διαλειτουργικότητας (όπως αναλύθηκαν στο Κεφάλαιο 1), η ίδια διαλειτουργικότητα μπορεί να προσφέρει χαμηλότερο κόστος μιας και η επίτευξή της απαιτεί λιγότερους πόρους και πρόσθετη συντήρηση. Παράλληλα, η πρόσβαση στις πληροφορίες μπορεί να παρέχεται σε όλους τους ενδιαφερόμενους, με την ποιότητα των δεδομένων να βελτιώνεται καθώς μπορούν να συγκεντρωθούν περισσότερες πηγές, ενώ ελαχιστοποιείται ο χρόνος που απαιτείται για την επεξεργασία των δεδομένων, αυξάνοντας έτσι την οργανωσιακή αποτελεσματικότητα και αποδοτικότητα. Ωστόσο, όπως φαίνεται και στην Εικόνα 2-1, η επίτευξη διαλειτουργικότητας μεταξύ των ενδιαφερομένων είναι δύσκολο, κάτι το οποίο οφείλεται σε πολλαπλούς παράγοντες [26].

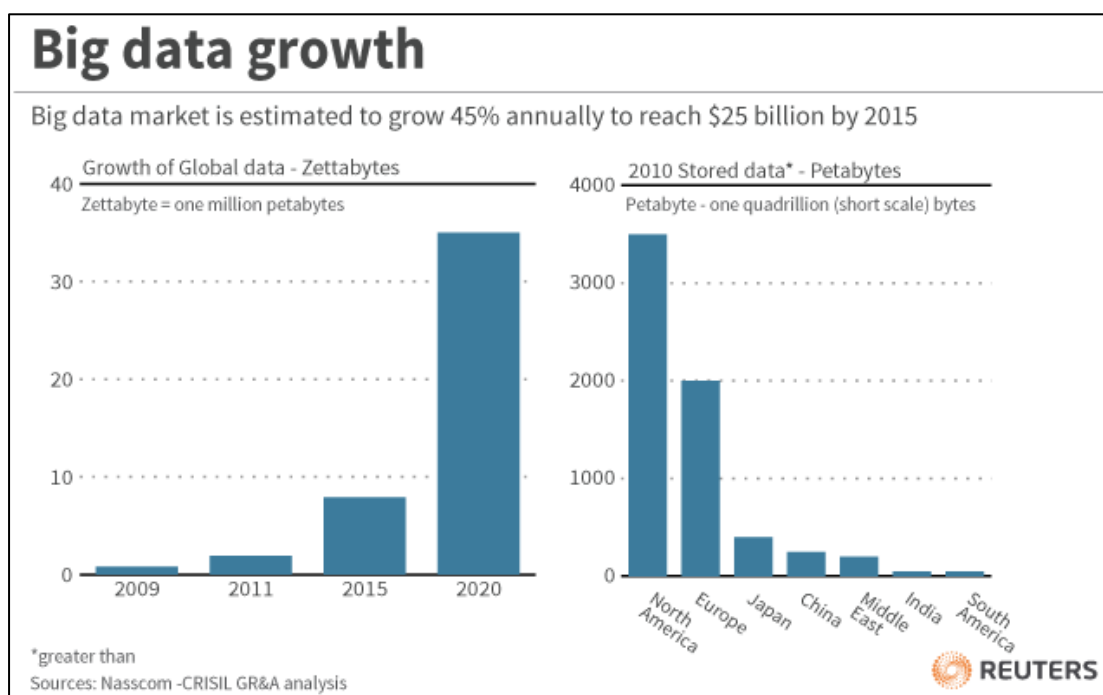
Ατέλειες	Οι υπολογιστές λειτουργούν συγκρίνοντας λέξεις Υπάρχουν ανάγκες τελειότητας Οι απαιτήσεις δεν ορίζονται σωστά
Μετάφραση	Οι υπολογιστές στέλνουν δεδομένα σε διαφορετικούς τρόπους Πραγματοποιείται διπλή μετάφραση πληροφοριών Δεν υπάρχει προτυποποιημένη γλώσσα
Προγραμματιστές	Κάθε σύστημα μιλάει τη δική του γλώσσα Επανα-χρησιμοποιείται ήδη υπάρχον λογισμικό Υπάρχει έλλειψη γνώσης Οι προγραμματιστές και οι χρήστες δεν συνεννοούνται σωστά
Χρήστες	Κάθε ειδικότητα μιλάει τη δική της γλώσσα Οι χρήστες δε γνωρίζουν τι πραγματικά θέλουν Υπάρχει συνεχής απαίτηση για αλλαγές Δεν υπάρχει κατανόηση της ανάπτυξης λογισμικού

Εικόνα 2-1 – Προκλήσεις επίτευξης διαλειτουργικότητας

Αναφορικά με τα δεδομένα, από την εποχή της ανάκαμψης της ψηφιοποίησης, οι επιχειρήσεις από διάφορους τομείς, έχουν αρχίσει να συσσωρεύουν τεράστια ποσά ψηφιακών δεδομένων, καταγράφοντας έτσι, δεδομένα τρισεκατομμύριων bytes, που αφορούν πληροφορίες των πελατών τους, των προμηθευτών, αλλά και των λειτουργιών τους. Ταυτόχρονα, με την ανάπτυξη νέων τεχνολογιών και εφαρμογών, όπως είναι η εξάπλωση των κοινωνικών δικτύων, η εκτεταμένη χρήση των έξυπνων τηλεφώνων και η εγκατάσταση αισθητήρων, έχει ως αποτέλεσμα ο όγκος και η μορφή των δεδομένων να έχει

αλλάξει δραματικά, ενώ οι δυνατότητες ανάλυσης και επεξεργασίας αυτών των δεδομένων γίνονται όλο και πιο εντυπωσιακές. Αυτό επέφερε μεγάλες αλλαγές στο πως αντλαμβάνονται τη λήψη αποφάσεων οι εταιρίες και οργανισμοί.

Όπως είναι λογικό, αυτή η αύξηση των δεδομένων δεν πρόκειται ποτέ να σταματήσει. Σύμφωνα με έρευνες που έχουν πραγματοποιηθεί, το 2009, είχαν δημιουργηθεί και αποθηκευτεί δεδομένα συνολικού μεγέθους 1 zettabyte, εν συνεχεία το 2015, το ποσό αυτό αυξήθηκε στα 8 zettabyte, ενώ προβλέπεται το ποσό αυτό να έχει αυξηθεί στα 35 zettabyte μέχρι το 2020, όπως φαίνεται και στην Εικόνα 2-2 [27].



Εικόνα 2-2 - Αύξηση όγκου δεδομένων

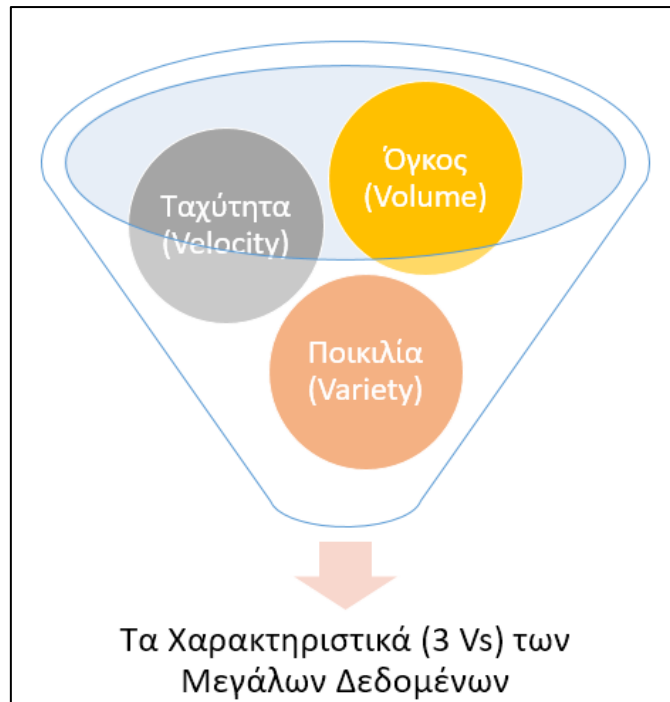
Η αύξηση αυτή των δεδομένων, έχει φέρει στα σκαριά το φαινόμενο των μεγάλων δεδομένων (Big Data) [28], ένα τεχνολογικό φαινόμενο που δημιουργήθηκε λόγω της συνεχούς και ραγδαίας αύξησης των δεδομένων, αλλά και των τεχνολογικών εξελίξεων που προκάλεσαν τη δημιουργία ενός οικοσυστήματος από προϊόντα λογισμικών και υλικών που επιτρέπουν στους χρήστες να αναλύουν τα δεδομένα αυτά, προκειμένου να είναι σε θέση να παράγουν καινούρια και πιο ακριβή επίπεδα αποτελεσμάτων (insights). Με πιο απλά λόγια, τα Big Data αποτελούν ένα φαινόμενο που προήλθε από την ραγδαία αύξηση των πληροφοριών, τους νέους τύπους πολύπλοκων δεδομένων και τις παράλληλες εξελίξεις στην τεχνολογία. Με τη βοήθεια του φαινομένου αυτού, οι παράγοντες αυτοί μπορούν να συνδυαστούν ώστε να

δίνουν τη δυνατότητα στους εμπλεκόμενους να αναλύουν εκτενέστερα και να διαχειρίζονται ευκολότερα τις εκάστοτε πληροφορίες.

Για τη διαχείριση των Big Data, καινούριες τεχνολογίες πρέπει να εφαρμοστούν προκειμένου αυτό το φαινόμενο να διαχειριστεί με τον πλέον κατάλληλο τρόπο. Στην ουσία, πρόκειται για ένα φαινόμενο που προσδιορίζει τις τεχνολογίες σαν μία νέα γενιά τεχνολογιών και αρχιτεκτονικών, που έχουν σχεδιαστεί, ώστε να εξάγουν εκτιμήσεις από ένα πολύ μεγάλο όγκο διαφόρων ειδών δεδομένων, επιτρέποντας την υψηλής ποιότητας σύλληψη της ταχύτητας, της ανακάλυψης, αλλά και της ανάλυσης αυτών.

Ειδικότερα, χρησιμοποιώντας τον όρο Big Data, αναφερόμαστε σε μεγάλα και πολύπλοκα σύνολα δεδομένων, των οποίων το μέγεθος ξεπερνά την ικανότητα των τυπικών βάσεων δεδομένων ή παραδοσιακών εφαρμογών επεξεργασίας δεδομένων, ώστε να καταγραφούν, να αποθηκευτούν και να αναλυθούν από αυτές. Ωστόσο, δεν υπάρχει κάποιος ακριβής ορισμός, για το πώς ένα μεγάλο σύνολο δεδομένων θα έπρεπε να ήταν, προκειμένου να καταταχθεί στην κατηγορία Big Data. Με πιο απλά λόγια, τα Big Data, αναφέρονται στα εργαλεία, τις διαδικασίες και τις διεργασίες, που μας δίνουν τη δυνατότητα να δημιουργούμε, να επεξεργαζόμαστε, να αποθηκεύουμε, να ανακτούμε και να διαχειριζόμαστε πολύ μεγάλα σύνολα δεδομένων.

Ωστόσο, τα Big Data δεν σχετίζονται απλά με το μέγεθος (volume) των δεδομένων, αλλά βρίσκουν αποτελέσματα (insights) από πολύπλοκα, θορυβώδη, ετερογενή, διαφορετικού μήκους και όγκου δεδομένων, όπως σχεσιακά δεδομένα, εικόνες, ήχοι, ή βίντεο. Δηλαδή, περιλαμβάνουν το είδος (variety) των δεδομένων, και έρχονται να απαντήσουν σε ερωτήματα, που πρωτίτερα ήταν αδύνατο να απαντηθούν. Επιπλέον, τα Big Data περιλαμβάνουν και την ταχύτητα (velocity) της συλλογής και της αποθήκευσης των δεδομένων, ένα χαρακτηριστικό που συνδυάζεται με τα δύο προαναφερθέντα χαρακτηριστικά, σχηματίζοντας τα 3Vs των Big Data (Εικόνα 2-3).



Εικόνα 2-3 - Τα 3Vs των Big Data

Πιο συγκεκριμένα, το χαρακτηριστικό του όγκου (volume) των δεδομένων, είναι στην ουσία συνώνυμο με την έννοια «big», του όρου Big Data. Ο όγκος των δεδομένων, είναι σχετικός όρος, καθώς οι μικρές επιχειρήσεις είναι πιθανό να έχουν μόνο gigabytes ή terabytes δεδομένων, σε αντίθεση με τις αρκετά πιο μεγάλες επιχειρήσεις οι οποίες καλούνται να διαχειρίζονται καθημερινά petabytes ή exabytes δεδομένων. Όπως είναι λογικό, ο όγκος των δεδομένων συνεχώς αυξάνεται, ασχέτως του μεγέθους μίας επιχείρησης. Γενικότερα, υπάρχει μία φυσική τάση από μέρους των επιχειρήσεων να αποθηκεύουν δεδομένα όλων των ειδών, όπως οικονομικά δεδομένα, ιατρικά δεδομένα, περιβαλλοντικά δεδομένα κτλ., είτε αυτά προέρχονται από το εσωτερικό τους περιβάλλον, είτε από το εξωτερικό. Πολλές από αυτές τις επιχειρήσεις μέχρι και σήμερα, μπορεί να καλούνται να διαχειρίζονται μόνο terabytes δεδομένων, αλλά στο μέλλον, είναι σίγουρο πως θα φτάσουν τα petabytes και τα exabytes. Οι αισθητήρες που είναι ενσωματωμένοι σε καθημερινά αντικείμενα, όπως οι έξυπνες συσκευές, σύντομα θα οδηγήσουν στη δημιουργία δισεκατομμυρίων νέων συνεχώς ενημερωμένων δεδομένων, συμπεριλαμβανομένων και πληροφοριών που σχετίζονται με το περιβάλλον, την τοποθεσία και άλλα. Για παράδειγμα, το Facebook, λαμβάνει καθημερινά 500 terabytes νέων δεδομένων, ένα ποσό το οποίο είναι σίγουρο, πως με το πέρασμα του χρόνου συνεχώς θα αυξάνεται και θα απαιτεί τις τεχνολογίες των Big Data προκειμένου να καθίσταται διαχειρίσιμο.

Όσον αφορά το χαρακτηριστικό της ταχύτητας (velocity) των δεδομένων, αυτό αναφέρεται στη συχνότητα με την οποία παράγονται και διανέμονται τα δεδομένα, δηλαδή στο πόσο γρήγορα αυτά στέλνονται και αποθηκεύονται, καθώς και πόσο εύκολα μπορεί να ανακτηθούν. Στα πλαίσια των Big Data, η ταχύτητα θα μπορούσε να αναφέρεται και στην κίνηση των δεδομένων, δηλαδή στην ταχύτητα με την οποία αυτά ρέουν. Τα διάφορα πληροφοριακά ρεύματα (streams) και η αύξηση της ανάπτυξης των αισθητήρων, έχουν οδηγήσει στη συνεχή ροή των δεδομένων, με τέτοιο ρυθμό, όπου καθιστάται αδύνατη η διαχείριση δεδομένων από παραδοσιακά συστήματα και εργαλεία που δεν χρησιμοποιούν Big Data. Ειδικότερα, η ταχύτητα των δεδομένων είναι ένα αρκετά σημαντικό χαρακτηριστικό των Big Data, μιας και καθημερινά με τη βοήθεια της ταχύτητας, καταγράφονται τα κλικ που πραγματοποιούν όλοι οι χρήστες του διαδικτύου, έχοντας σαν αποτέλεσμα να αποθηκεύεται η συμπεριφορά τους, σύμφωνα με εκατομμύρια κινήσεις που πραγματοποιούν ανά δευτερόλεπτο. Επιπλέον, οι υψηλής συχνότητας χρηματιστηριακές αλλαγές επηρεάζουν και καθορίζουν την κίνηση των αγορών μέσα σε microseconds, βασιζόμενοι πάντα στην ταχύτητα των δεδομένων, ενώ οι ηλεκτρονικές διαδικασίες ανταλλάσσουν δεδομένα μεταξύ δισεκατομμυρίων συσκευών. Πρόκειται δηλαδή για περιπτώσεις όπου παράγονται πολλαπλές εισοδοί δεδομένων ανά δευτερόλεπτο, απαιτώντας συνεχή παρακολούθηση και ανάλυση, είτε πρόκειται για δεδομένα πραγματικού χρόνου είτε όχι.

Τέλος, όσον αφορά το χαρακτηριστικό του είδους (variety) των δεδομένων, αυτό υποδηλώνει πως τα Big Data δεν είναι απλά αριθμοί, ημερομηνίες και αλφαριθμητικά στοιχεία. Τα Big Data μπορούν επίσης να είναι δεδομένα ήχου, εικόνας, βίντεο, 3D, αλλά και μη δομημένα κείμενα, συμπεριλαμβανομένων και των κοινωνικών δικτύων. Ειδικότερα, οι παραδοσιακές βάσεις δεδομένων είχαν σχεδιαστεί ώστε να διαχειρίζονται μικρότερο όγκο δομημένων δεδομένων, λιγότερο αναβαθμισμένων, με πιο προβλέψιμη, συνεπή και δομημένη μορφή. Επιπλέον, είχαν σχεδιαστεί για να λειτουργούν σε ατομικούς εξυπηρετητές, καθιστώντας την απαιτούμενη αυξανόμενη χωρητικότητα πολύ ακριβή και περιορισμένη. Καθώς λοιπόν οι εφαρμογές έχουν εξελιχθεί ώστε να εξυπηρετούν μεγάλο όγκο χρηστών και δεδομένων, και καθώς η ανάπτυξη των εφαρμογών έχει γίνει ευκίνητη (agile), οι παραδοσιακές βάσεις δεδομένων δεν εξασφαλίζουν τη βιωσιμότητα και ομαλή λειτουργία των επιχειρήσεων, οπότε απαιτούν την άμεση αντικατάστασή τους. Οι Big Data βάσεις δεδομένων, δηλαδή οι μη σχεσιακές βάσεις δεδομένων (Non Structured Query Language - NoSQL) [29] όπως η MongoDB [30], λύνουν ακριβώς αυτό το πρόβλημα της διαχείρισης των διαφόρων ειδών Big Data, παρέχοντας στις επιχειρήσεις τη δυνατότητα να διαχειρίζονται τα πολύ μεγάλου όγκου δεδομένα.

Καθίσταται ξεκάθαρο πως τα δεδομένα αυξάνονται με ιδιαίτερα γρήγορο ρυθμό, με αποτέλεσμα αρκετά συστήματα να έχουν οδηγηθεί στην αποθήκευση και ανάλυση τέτοιων μεγάλων ποσοτήτων δεδομένων. Τα εν λόγω συστήματα, αποκαλούμενα ως «Συστήματα Big Data», εξελίσσονται με αρκετά γοργούς ρυθμούς. Παραδείγματα τέτοιων συστημάτων περιλαμβάνουν συστήματα χρήσης NoSQL, αλγορίθμους Hadoop Map-Reduce [31], πλατφόρμες για εφαρμογή ανάλυσης δεδομένων (Data Analytics), πλατφόρμες αναζήτησης, καθώς και υποδομές μεταφοράς μηνυμάτων. Τα συγκεκριμένα συστήματα, έχουν την ανάγκη για δομημένα και αδόμητα δεδομένα, μέσα σε ένα σύνολο πηγών δεδομένων από τα οποία περιτριγυρίζονται, όπως είναι το διαδίκτυο, τα κοινωνικά δίκτυα, οι επιχειρήσεις, οι υποδομές νεφοϋπολογιστικής (Cloud Computing), τα κινητά τηλέφωνα, τα συστήματα υγείας, τα συστήματα μεταφοράς, οι πηγές ενέργειας, οι επιστήμες της βιοπληροφορικής και άλλα.

Μερικά παραδείγματα οργανισμών και συστημάτων που χρησιμοποιούν Big Data, για την ανάλυση και εξόρυξη γνώσεων από αυτά, είναι η Wal-Mart η οποία υπολογίζοντας πως το 2004, είχε τη μεγαλύτερη αποθήκη δεδομένων με 500 terabytes αποθηκευτικού χώρου. Μερικά χρόνια αργότερα, υπολογιζόταν πως το eBay, είχε αποθηκευτικό χώρο περίπου 8 petabytes, ενώ δύο χρόνια πιο μετά, η αποθήκη δεδομένων της Yahoo, αποτελούταν από 170 petabytes αποθηκευτικού χώρου. Παρατηρούμε δηλαδή πως, με την πάροδο των χρόνων, οι αριθμοί αυτοί πολλαπλασιάζονται, με αποτέλεσμα σήμερα να μιλάμε για εκατοντάδες petabytes δεδομένων ανά σύστημα.

Γενικότερα, σε ένα συνεχώς αυξανόμενο τεχνολογικά περιβάλλον, οι κατευθύνσεις των επιχειρήσεων είναι αλληλοεπιδρώμενες, με την έννοια πως ότι συμβαίνει σε μία μοναδική κατεύθυνση, αυτό έχει άμεσο αντίκτυπο και στις υπόλοιπες κατευθύνσεις. Σε έναν οργανισμό, αυτή η πολυπλοκότητα δυσκολεύει τις επιχειρήσεις, οι οποίες πρέπει να βασίζονται μόνο στην εμπειρία τους για να λαμβάνουν αποφάσεις. Για τον λόγο αυτό, χρειάζονται να βασίζονται σε έμπιστες και αποτελεσματικές υπηρεσίες δεδομένων, προκειμένου να λαμβάνουν τις αποφάσεις τους. Ενσωματώνοντας τα Big Data στην «καρδιά» της λειτουργικότητας των επιχειρήσεων, παρέχεται η πρόσβαση σε νέα αποτελέσματα, μετατρέποντάς τες σε περισσότερο ανταγωνιστικές και αποδοτικές. Η συμβολή λοιπόν των Big Data σε όλους τους τομείς, και ειδικότερα σε αυτούς των επιχειρήσεων, κρίνεται υψίστης σημασίας.

Για πολλά χρόνια, οι επιχειρήσεις κατέγραφαν τα δεδομένα των συναλλαγών τους, και χρησιμοποιούσαν απλοποιημένες διαδικασίες, προκειμένου να διαθέτουν περιλήψεις και πληροφορίες σχετικές με τα προϊόντα τους,

αποθηκευμένες σε παραδοσιακές σχεσιακές βάσεις δεδομένων. Τα τελευταία χρόνια, νέες τεχνολογίες με λιγότερα κόστη έχουν βελτιώσει την καταγραφή των δεδομένων αυτών, την αποθήκευση αλλά και την ανάλυσή τους. Οι επιχειρήσεις μπορούν πλέον να καταγράψουν δεδομένα από πολλές και διαφορετικές πηγές δεδομένων, όπως από ιστολόγια (blogs), από κοινωνικά δίκτυα και άλλα. Οι επιλογές αυτές, οι οποίες δίνουν τη δυνατότητα αποθήκευσης και επεξεργασίας των δεδομένων με το βέλτιστο τρόπο, έχουν επεκταθεί δραματικά, με αποτέλεσμα οι τεχνολογίες όπως η MapReduce, να παρέχουν ιδιαίτερα βελτιστοποιημένες δυνατότητες για διαφορετικούς επιχειρηματικούς σκοπούς.

Η ανάλυση των δεδομένων μπορεί να γίνει είτε σε πραγματικό χρόνο είτε σε σχετικά πραγματικό χρόνο (close to real time), εφαρμόζοντάς την σε μεγάλα σύνολα δεδομένων, παρά σε συγκεκριμένα στοιχεία. Επιπρόσθετα, ο αριθμός των επιλογών προκειμένου να εξηγούνται και να αναλύονται τα δεδομένα, έχει επίσης αυξηθεί, με τη χρήση τεχνολογιών απεικόνισης (visualisation). Η τεχνική ανάλυσης των δεδομένων, αναφέρεται στις ποιοτικές και στις ποσοτικές τεχνικές και διαδικασίες που χρησιμοποιούνται, ώστε να αυξηθεί η παραγωγικότητα και το κέρδος των επιχειρήσεων. Τα δεδομένα αποσυμπίεζονται και κατηγοριοποιούνται προκειμένου να ταυτοποιηθούν, να αναλυθούν επιπλέον δεδομένα και πρότυπα (patterns) δεδομένων, καθώς και τεχνικές, οι οποίες ποικίλουν ανάλογα με τις απαιτήσεις του εκάστοτε οργανισμού.

Ο όρος ανάλυση των δεδομένων (Data Analytics), είναι γνωστός και ως Data Analysis. Ωστόσο, αποτελεί μία λέξη γύρω από το χώρο της επιχειρησιακής νοημοσύνης (Business Intelligence), την οποία κάθε οργανισμός και επιχείρηση, χρησιμοποιούν για να περιγράψουν πολύ διαφορετικές λειτουργίες. Ειδικότερα, ύστερα από έρευνα που πραγματοποιήθηκε σχετικά με το τι είναι analytics, οι απαντήσεις ποίκιλαν, αναφέροντας είτε πως αποτελούν μια αναλυτική διαδικασία που πραγματοποιείται διαδικτυακά, είτε μια διαδικασία εξόρυξης δεδομένων.

Προσπαθώντας να δώσουμε έναν κάπως πιο αποτελεσματικό και σωστό ορισμό, καταλήγουμε πως ο όρος Data Analytics, αποτελεί την επιστήμη που εξετάζει ακατέργαστα δεδομένα, με στόχο τη δημιουργία πληροφοριών, σχετικά με αυτά. Τα Data Analytics, χρησιμοποιούνται σε πολλές βιομηχανίες, με στόχο να βοηθούν τις εταιρείες και τους οργανισμούς, να λαμβάνουν καλύτερες επιχειρηματικές αποφάσεις, και τις εκάστοτε επιστήμες να επιβεβαιώνουν ή να απορρίπτουν, ήδη υπάρχοντα μοντέλα και θεωρίες.

Τα Data Analytics, διαχωρίζονται από τον όρο Εξόρυξη Δεδομένων (Data Mining), σε ότι αφορά τον σκοπό, το λόγο και την επικέντρωση που δίνεται κατά την ανάλυση. Πιο συγκεκριμένα, στην Εξόρυξη Δεδομένων, αναζητούνται δεδομένα και πληροφορίες, μεταξύ τεράστιων όγκων δεδομένων, με τη χρήση εξελιγμένων λογισμικών, ώστε να αποκαλύπτονται κρυφές συσχετίσεις και ανεύρετα πρότυπα (patterns). Από την άλλη μεριά, τα Data Analytics, επικεντρώνονται στα συμπεράσματα, βάσει της διαδικασίας δημιουργίας μιας κατάληξης που βασίζεται αποκλειστικά σε αυτό που είναι ήδη γνωστό από τον ερευνητή. Γενικότερα, ο στόχος των Data Analytics (είτε σε μεγάλη ποσότητα, είτε σε μικρή ποσότητα δεδομένων), είναι να χρησιμοποιούν τα παραγόμενα αποτελέσματα ώστε να λαμβάνουν πιο έξυπνες αποφάσεις και καλύτερα συνολικά αποτελέσματα για τις επιχειρήσεις. Ο τρόπος με τον οποίο γίνεται η αρχιτεκτονική των επιχειρησιακών τεχνολογιών και η σχεδίαση των διαδικασιών των Data Analytics, ώστε να βγουν πολύτιμα αποτελέσματα (insights), ποικίλει. Κρίνεται κρίσιμο να σχεδιάζεται και να κτίζεται μία αποθήκη δεδομένων ή μία αρχιτεκτονική επιχειρησιακής νοημοσύνης, που να παρέχει ένα ευέλικτο, πολύπλευρο σύστημα, με στόχο την αποτελεσματική ανάλυση μεγάλων και διαφορετικών συνόλων από δεδομένα.

Γενικότερα, τα Data Analytics υπόσχονται πως θα επιφέρουν επανάσταση στους χώρους των επιχειρήσεων, των επιστημών, της έρευνας και της παιδείας. Ισχυρότατοι αλγόριθμοι βοηθούν στην αναγνώριση προσώπων σε συστήματα κοινωνικών υπηρεσιών, εντοπίζουν τις απάτες, προβλέπουν τα αποτελέσματα των φυσικών καταστροφών, αναγνωρίζουν πρότυπα σε επιστημονικές έρευνες και ανακαλύπτουν τις ανάγκες των καταναλωτών. Τα Data Analytics, παίζουν έναν πολύ σημαντικό ρόλο στο να διαχειρίζονται θέματα που αφορούν μία ολόκληρη κοινωνία, τα οποία αφορούν από θέματα κατανόησης βιολογικών εννοιών, μέχρι και θέματα που σχετίζονται με τη βελτίωση της εκπαίδευσης και της διαχείρισης των πηγών ενέργειας. Καθώς η χρήση των Data Analytics μπορεί να παρατηρηθεί ήδη από τις αρχές του 1800, σήμερα η χρήση τους συνοδεύεται από τον όρο Big Data, τα οποία όπως αναφέρθηκε, αποτελούν τεράστιες ποσότητες πληροφοριών που συλλέγονται από παραδοσιακές πηγές (δεδομένα υγείας, δεδομένα συναλλαγών, οικονομικά δεδομένα), καθώς και από νέες πηγές δεδομένων (κείμενα, δεδομένα διαδικτύου, χρόνος, δεδομένα τοποθεσίας, δεδομένα κοινωνικών δικτύων).

Παραδοσιακά τα Data Analytics, χρησιμοποιούνταν προκειμένου να βρίσκονται απαντήσεις σε προκαθορισμένες ερωτήσεις, ενώ πλέον η εφαρμογή τους στα Big Data βοηθά στην εξερεύνηση των πληροφοριών, προκειμένου να εξαχθεί κάποιο είδος γνώσης, ώστε να καθοριστούν συνδέσεις και συσχετίσεις

που δεν είχαν προβλεφθεί προηγουμένως. Έτσι, ο στόχος των Big Data να εμφανίζουν απρόβλεπτα αποτελέσματα, το ιδιαίτερα χαμηλό κόστος της αποθήκευσης των πληροφοριών, αλλά και η ταχύτερη ανάπτυξη της ισχύος των αλγορίθμων, σε συνδυασμό με τη χρήση μεθοδολογιών Data Analytics, μετατόπισαν τις προτεραιότητες των οργανισμών και των επιχειρήσεων, ώστε να συλλέγουν όσο πιο πολλά δεδομένα είναι δυνατό, προσπαθώντας να εξάγουν αξιολογικά και αντιπροσωπευτικά αποτελέσματα και γνώσεις από αυτά. Όλες αυτές οι εξελίξεις, αποτελούν το πλαίσιο εντός του οποίου τοποθετούνται τα Big Data, με αποτέλεσμα να κρίνεται απαραίτητη η ύπαρξη και η εφαρμογή τους.

Λειτουργώντας με βάσεις δεδομένων που χρησιμοποιούν Big Data, οι επιχειρήσεις μπορούν να εξοικονομήσουν χρήματα, να αποκτήσουν κέρδος, και να επιτύχουν πολλούς διαφορετικούς επιχειρησιακούς στόχους, σε κάθε τομέα. Πιο συγκεκριμένα, τα Big Data μπορούν να συμβάλλουν στα ακόλουθα:

- **Δημιουργία νέων εφαρμογών:** Τα Big Data μπορεί να επιτρέπουν σε μια επιχείρηση να συλλέγει δισεκατομμύρια δεδομένα πραγματικού χρόνου σχετικά με τα προϊόντα της, τις πηγές της, ή τους πελάτες της – και στη συνέχεια να επεξεργαστεί τα δεδομένα αυτά, ώστε να αποκτήσει μία εικόνα για τους χρήστες των προϊόντων ή για τη χρήση των πόρων της. Για παράδειγμα, μία ανεπτυγμένη πόλη των Ηνωμένων Πολιτειών της Αμερικής, χρησιμοποιεί τη μη σχεσιακή βάση δεδομένων MongoDB, προκειμένου να μειώσει την εγκληματικότητα και να βελτιώσει τις κοινωνικές υπηρεσίες της, συλλέγοντας και αναλύοντας γεωγραφικά δεδομένα σε πραγματικό χρόνο, από 30 διαφορετικά διαμερίσματα.
- **Βελτίωση της αποτελεσματικότητας και μείωση του κόστους των υπαρχόντων εφαρμογών:** Οι τεχνολογίες των Big Data, μπορούν να αντικαταστήσουν συστήματα που είναι πολύ ακριβά και μη παραμετροποιήσιμα, με μία λύση η οποία λειτουργεί αποκλειστικά με υλικά αγαθά. Επειδή πολλές τεχνολογίες Big Data είναι ανοιχτού λογισμικού, μπορούν να χρησιμοποιηθούν με πολύ πιο χαμηλό κόστος, από τις υπόλοιπες τεχνολογίες. Για παράδειγμα, μεταφέροντας τη διαχείριση των δεδομένων μίας εφαρμογής, στη μη σχεσιακή βάση δεδομένων MongoDB, μία συγκεκριμένη τράπεζα, μείωσε δραματικά τα κόστη για την άδεια και τη λειτουργία της, ενώ παράλληλα, η εφαρμογή που χρησιμοποιούνταν βελτιώθηκε σύμφωνα με τις καθημερινές ανάγκες των χρηστών.

- **Κατανόηση νέων πόρων για ανταγωνιστικό πλεονέκτημα:** Τα Big Data μπορούν να βοηθήσουν τις επιχειρήσεις να δράσουν περισσότερο έξυπνα, επιτρέποντάς τες να προσαρμοστούν στις αλλαγές πιο γρήγορα από τους ανταγωνιστές τους. Για παράδειγμα, η μη σχεσιακή βάση δεδομένων MongoDB, επέτρεψε στους παρόχους λύσεων μιας εταιρείας, να δημιουργήσουν με πολύ γρήγορους ρυθμούς εφαρμογές για κινητά τηλέφωνα, που χρησιμοποιούσαν δεδομένα από μία ευρεία ποικιλία, ξεχωριστών πηγών.
- **Αύξηση της εμπιστευτικότητας των πελατών:** Η αύξηση του ποσού των δεδομένων που διαμοιράζονται μεταξύ των οργανισμών - σε συνδυασμό με την ταχύτητα που αναβαθμίζονται - επιτρέπουν στις επιχειρήσεις και σε άλλους οργανισμούς, να ανταποκρίνονται πιο γρήγορα και με περισσότερη ακρίβεια, στις απαιτήσεις των πελατών τους. Για παράδειγμα, ένας πάροχος ασφάλειας, χρησιμοποιεί τη μη σχεσιακή βάση δεδομένων MongoDB, προκειμένου να συγκεντρώνει πληροφορίες πελατών, από περισσότερες από εβδομήντα διαφορετικές πηγές, ώστε να τις παρέχει σε ένα μοναδικό, ταχέως εξελισσόμενο περιβάλλον.

Συνοπτικά, καταλήγουμε πως τα δεδομένα βρίσκονται παντού, σε κάθε πηγή που μπορεί οποιοσδήποτε να σκεφτεί. Στα κινητά τηλέφωνα, στα τρέινα, στις μηχανές των αεροπλάνων, στα αυτοκίνητα και φυσικά στο διαδίκτυο. Η αξία λοιπόν των Big Data αυξάνεται με την πάροδο του χρόνου όλο και περισσότερο, καθώς εξάγονται όλο και περισσότερα αποτελέσματα από την εφαρμογή των analytics σε αυτά. Παρόλο, που τα analytics από μόνα τους δεν αποτελούν κάτι νέο, με την ύπαρξη των Big Data, έχουμε στη διάθεσή μας έναν περισσότερο ευρύ πίνακα πηγών δεδομένων, που μπορεί να αυξήσουν τις γνώσεις και τα αποτελέσματα. Νέες τεχνολογίες όπως το Hadoop, αποθήκες αντικειμένων και βάσεις δεδομένων εσωτερικής μνήμης, σε συνδυασμό με την αύξηση της οικονομίας, βοηθούν στην άμεση επεξεργασία απρόβλεπτων όγκων και ποικιλιών δεδομένων. Τώρα πλέον, μπορούμε να ισχυριζόμαστε, πως η δυνατότητα να εξάγονται γνώσεις και αποτελέσματα από δεδομένα, έχει περάσει σε πολύ ανώτερο επίπεδο.

2.1.2 Κατηγορίες Διαλειτουργικότητας

Η διαλειτουργικότητα, περιλαμβάνει και εισχωρεί γενικά σε πολλούς διαφορετικούς κλάδους με αποτέλεσμα να αναφερόμαστε σε διαλειτουργικότητα δεδομένων, διαδικασιών, κανόνων, αντικειμένων, συστημάτων, πολιτισμού, γνώσης ή υπηρεσιών. Τα διασυνδεδεμένα συστήματα μπορούν να διαφέρουν σε βαθμό διαλειτουργικότητας. Ορισμένοι τύποι διαλειτουργικότητας που συναντώνται πιο συχνά περιλαμβάνουν τις ακόλουθες κατηγορίες διαλειτουργικότητας:

- **Συντακτική διαλειτουργικότητα [32]:** Αναφέρεται σε συστήματα που μπορούν να επικοινωνούν με επιτυχία μέσω συμβατών μορφών και πρωτοκόλλων. Εργαλεία που διευκολύνουν τη συντακτική διαλειτουργικότητα είναι αναγνωρισμένα πρότυπα μορφοποίησης, όπως οι γλώσσες Extensible Markup Language (XML) και Structured Query Language (SQL). Η συντακτική διαλειτουργικότητα αποτελεί προϋπόθεση για τη σημασιολογική διαλειτουργικότητα. Η συντακτική διαλειτουργικότητα αναφέρεται στους μηχανισμούς συσκευασίας και μετάδοσης δεδομένων. Το τρέχον πρότυπο διαδικτύου για τη σήμανση εγγράφων είναι η XML, η οποία χρησιμοποιεί τον «<>» ως οριοθέτη δεδομένων. Οι οριοθέτες δεδομένων δεν μεταφέρουν κανένα νόημα στα δεδομένα εκτός από τη δομή των δεδομένων. Χωρίς λεξικό δεδομένων για τη μετάφραση του περιεχομένου των οριοθετών, τα δεδομένα παραμένουν χωρίς νόημα. Ενώ υπάρχουν πολλές προσπάθειες για τη δημιουργία λεξικών δεδομένων και μοντέλων πληροφοριών για τη συσχέτιση με αυτούς τους μηχανισμούς συσκευασίας δεδομένων, κανένας δεν είναι ολοκληρωτικά πρακτικός ώστε να εφαρμοστεί.
- **Σημασιολογική διαλειτουργικότητα [33]:** Αναφέρεται στην ικανότητα των συστημάτων να ανταλλάσσουν και να ερμηνεύουν με ακρίβεια τις πληροφορίες αυτόματα. Η σημασιολογική διαλειτουργικότητα επιτυγχάνεται όταν η δομή και η κωδικοποίηση των δεδομένων είναι ομοιόμορφη μεταξύ όλων των εμπλεκόμενων συστημάτων. Ουσιαστικά είναι η ικανότητα των συστημάτων ηλεκτρονικών υπολογιστών να ανταλλάσσουν δεδομένα με σαφή και κοινή έννοια. Η σημασιολογική διαλειτουργικότητα είναι μια απαίτηση για να καταστεί δυνατή η υπολογιστική λογική της μηχανής, η συσχέτιση, και η ανακάλυψη γνώσεων μεταξύ των πληροφοριακών συστημάτων. Επομένως, η σημασιολογική διαλειτουργικότητα αφορά όχι μόνο τη συσκευασία των δεδομένων (σύνταξη), αλλά την ταυτόχρονη μετάδοση της έννοιας των δεδομένων

(σημασιολογία). Αυτό επιτυγχάνεται προσθέτοντας επιπλέον δεδομένα σχετικά με τα δεδομένα (μεταδεδομένα), συνδέοντας κάθε στοιχείο δεδομένων σε ένα ελεγχόμενο κοινόχρηστο λεξιλόγιο. Η έννοια των δεδομένων μεταδίδεται με τα ίδια τα δεδομένα, σε ένα αυτο-περιγραφόμενο «πακέτο πληροφοριών» που είναι ανεξάρτητο από οποιοδήποτε σύστημα πληροφοριών. Συνοπτικά, είναι αυτό το κοινό λεξιλόγιο και οι σχετικοί σύνδεσμοί του με μια οντολογία, που παρέχουν τη βάση και την ικανότητα της μηχανικής ερμηνείας, συμπερασμάτων και λογικής.

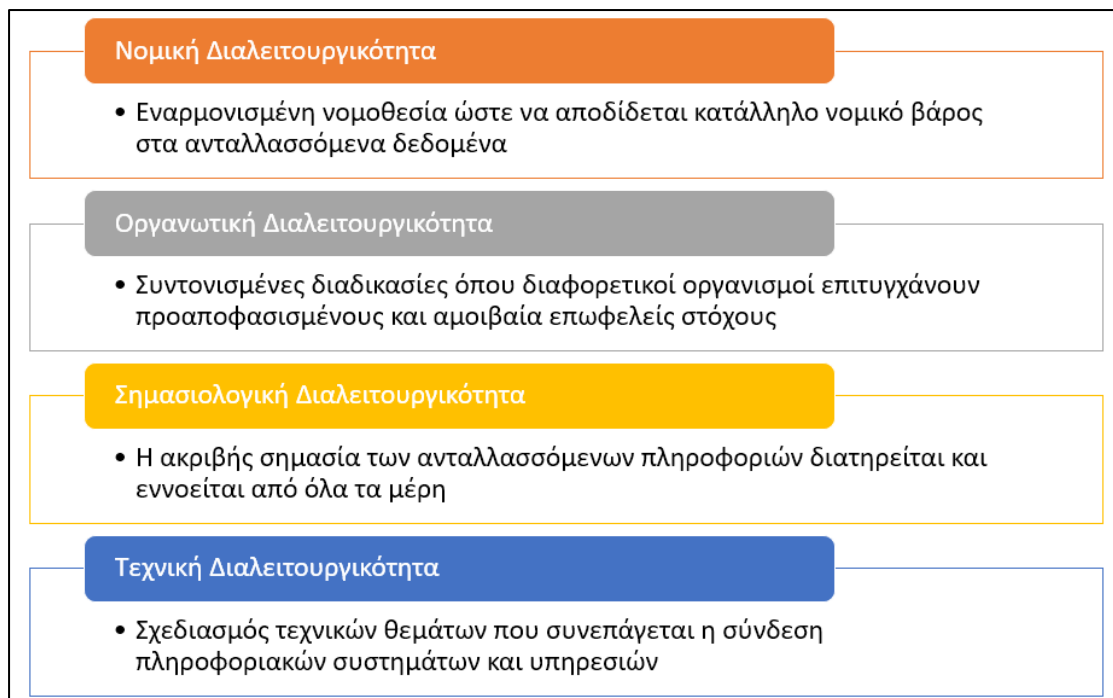
- **Διαλειτουργικότητα μεταξύ διαφόρων τομέων ή διασταυρούμενων οργανισμών [34]:** Αναφέρεται στην τυποποίηση πρακτικών, πολιτικών, θεμελίων και απαιτήσεων διαφόρων συστημάτων. Αντί να σχετίζεται με τους μηχανισμούς που βρίσκονται πίσω από την ανταλλαγή δεδομένων, αυτός ο τύπος επικεντρώνεται μόνο στις μη τεχνικές πτυχές ενός διαλειτουργικού οργανισμού. Ο εν λόγω τύπος διαλειτουργικότητας εντοπίζεται στις περιπτώσεις που οργανισμοί ή συστήματα από διαφορετικούς τομείς αλληλοεπιδρούν στην ανταλλαγή πληροφοριών, υπηρεσιών ή/και αγαθών για να επιτύχουν δικούς τους ή κοινούς στόχους. Ένας τομέας σε αυτήν την περίπτωση είναι μια κοινότητα με σχετική υποδομή που δεσμεύεται από κοινούς σκοπούς και συμφέροντα, με σταθερές αμοιβαίες αλληλεπιδράσεις ή κανόνες δέσμευσης που μπορούν να διαχωριστούν από άλλες κοινότητες με κοινωνικά, τεχνικά, γλωσσικά, επαγγελματικά, νομικά ή κυριαρχικά όρια. Η δυνατότητα διαλειτουργικότητας μεταξύ τομέων αποκτά ολοένα και μεγαλύτερη σημασία καθώς οι επιχειρήσεις και οι κυβερνητικές επιχειρήσεις μετατρέπονται σε παγκόσμιες και αλληλεξαρτώμενες. Η διαλειτουργικότητα μεταξύ τομέων επιτρέπει τη συνεργασία, επεκτείνει τη χρησιμότητα του προϊόντος και επιτρέπει στους χρήστες να είναι πιο αποτελεσματικοί και επιτυχείς στους δικούς τους τομείς και στη συνδυασμένη προσπάθεια. Η διαλειτουργικότητα μεταξύ τομέων χαρακτηρίζεται από κοινή κατανόηση και συμφωνίες και από τις δύο πλευρές, επιτρέποντας σε μεμονωμένους οργανισμούς να προσαρμόζονται ή να κάνουν τα προϊόντα, τα περιουσιακά στοιχεία ή τις υπηρεσίες τους διαλειτουργικά μέσα στην ευρύτερη κοινότητα. Κάθε συμμετέχων αποδέχεται και επιβάλλει τη χρήση αμοιβαίων, και παγκόσμιων προτύπων και πρωτοκόλλων διεπαφής. Συνεπώς, οι διασυνδέσεις μεταξύ τομέων ενδέχεται να μην ελέγχονται από κανένα στοιχείο ή αρχή - σε αντίθεση με ένα περιβάλλον ολοκληρωμένου συστήματος συστημάτων όπου ένας τομέας ή η αρχή του μπορεί να ελέγχει τις διεπαφές που πρέπει να χρησιμοποιούνται μεταξύ τομέων. Ορισμένοι οργανισμοί, επιχειρήσεις και ιδρύματα εργάζονται για την τεχνολογία και τις πολιτικές για να καταστήσουν πραγματικότητα τη διαλειτουργικότητα

μεταξύ τομέων, συμπεριλαμβανομένου του Εθνικού Ινστιτούτου Προτύπων και Τεχνολογίας, του Υπουργείου Άμυνας των Ηνωμένων Πολιτειών, του North Atlantic Treaty Organization (NATO) και της Network Centric Operations Industry Council Interoperability (NCOIC).

Εκτός των προαναφερθέντων κατηγοριών διαλειτουργικότητας, βάσει ορισμού, οι βασικότερες κατηγορίες στις οποίες μπορεί να χωριστεί είναι οι ακόλουθες [35]:

- **Νομική Διαλειτουργικότητα** (Legal Interoperability)
- **Οργανωτική Διαλειτουργικότητα** (Organizational Interoperability)
- **Σημσιολογική Διαλειτουργικότητα** (Semantic Interoperability)
- **Τεχνική Διαλειτουργικότητα** (Technical interoperability)

Στην Εικόνα 2-4, μέσα από μια σχηματική αναπαράσταση, βλέπουμε το εύρος επιρροής των βασικών ειδών της διαλειτουργικότητας. Η νομική μπορεί να θεωρηθεί πως αποτελεί υπερσύνολο της οργανωτικής, η οποία με τη σειρά της είναι υπερσύνολο της σημσιολογικής, που έχει υποσύνολο την τεχνική. Καθίσταται προφανές ότι η διαλειτουργικότητα στο σύνολό της, δεν έχει πολλές πιθανότητες επιτυχίας, εάν δεν ληφθούν σοβαρά υπόψη όλες οι πτυχές της, και εάν δεν δοθεί σε κάθε επίπεδο της η αντίστοιχη βαρύτητα, που του αρμόζει.



Εικόνα 2-4 – Βασικές κατηγορίες διαλειτουργικότητας

2.1.2.1 Νομική Διαλειτουργικότητα

Γενικότερα, κατά την ανταλλαγή οποιωνδήποτε πληροφοριών μεταξύ των διαφόρων ενδιαφερομένων με σκοπό την παροχή υπηρεσιών, είναι απαραίτητο να διατηρείται αφενός η νομική εγκυρότητα των πληροφοριών, και αφετέρου πρέπει να λαμβάνεται υπόψη και να γίνεται σεβαστή η εκάστοτε νομοθεσία περί προστασίας των δικαιωμάτων μεταξύ των ενδιαφερομένων που συναναστρέφονται. Οι ασυμβατότητες λοιπόν, μεταξύ των κανόνων των ενδιαφερομένων, μετατρέπουν τη μεταξύ τους συνεργασία σε κάτι πιο δύσκολο και σύνθετο, καθιστώντας πολλές φορές αδύνατη τη συνεργασία, ακόμη και στις περιπτώσεις που οι κανόνες προκύπτουν σαν αποτέλεσμα της μεταφοράς των οδηγιών στο εθνικό δίκαιο.

Ακολούθως, εντοπίζονται μερικά από τα θέματα νομικής φύσης που μπορεί να συναντηθούν και χρειάζεται να ληφθούν υπόψη κατά την ανταλλαγή διαλειτουργικών δεδομένων. Αυτά είναι:

- Ο εντοπισμός και η προσαρμογή των νομικών διατάξεων που διέπουν την εκτέλεση των διαδικασιών και την παροχή των υπηρεσιών, έτσι ώστε να προβλέπεται η χρήση ηλεκτρονικών μέσων.
- Η διασφάλιση της εγκυρότητας και της νομικής ισχύος των αποτελεσμάτων των διαδικασιών/υπηρεσιών, όταν αυτά παρέχονται με ηλεκτρονικά μέσα.
- Η προστασία της ιδιωτικότητας και των προσωπικών και ευαίσθητων δεδομένων των αποδεκτών των ηλεκτρονικών υπηρεσιών.

Σαν αποτέλεσμα αυτών, για την αποφυγή των προαναφερθέντων προβληματικών καταστάσεων και δυσκολιών, ορίζεται η έννοια της νομικής διαλειτουργικότητας, η οποία ξεκινά από το πρόβλημα της ετερογένειας σε επίπεδο νομοθεσίας και κανονιστικών διατάξεων που διέπουν τη λειτουργία δύο ή περισσότερων φορέων που επιθυμούν να συνεργαστούν, για τη μεταξύ τους ανταλλαγή πληροφοριών ή/και την παροχή ολοκληρωμένων ηλεκτρονικών υπηρεσιών προς πολίτες, επιχειρήσεις και άλλους φορείς. Τελικός σκοπός της είναι να διασφαλίσει ότι οι πληροφορίες που ανταλλάσσονται ηλεκτρονικά έχουν την ίδια νομική ισχύ για όλους τους εμπλεκόμενους φορείς, σε οποιαδήποτε χώρα και κράτος βρίσκονται.

Πρόκειται λοιπόν, για την εναρμόνιση των νομικών πλαισίων, τα οποία καθορίζουν τον τρόπο οργάνωσης και λειτουργίας των φορέων και των

υπηρεσιών, και καθορίζεται από την μεριά της με επιπλέον νομοθεσίες και κανονιστικές διατάξεις. Αξίζει να σημειωθεί, πως η νομική διαλειτουργικότητα μπορεί να επηρεάσει σε μεγάλο βαθμό είτε θετικά είτε αρνητικά όλα τα επόμενα στρώματα διαλειτουργικότητας που θα αναλυθούν, μιας και η μη τήρηση ή ύπαρξη νομοθετικών κανόνων, μπορεί να οδηγήσουν στη διακοπή της επικοινωνίας και της ανταλλαγής υπηρεσιών μεταξύ διαφορετικών κρατών μελών. Αυτό θα έχει το δυσάρεστο αποτέλεσμα της ανάγκης ανασχεδιασμών, αλλά και επιπλέον κόστος. Αυτός είναι στην ουσία και ο βασικότερος λόγος, που το στρώμα της νομικής διαλειτουργικότητας αποτελεί και το σημαντικότερο από τα υπόλοιπα στρώματα διαλειτουργικότητας. Είναι σαφές δηλαδή, ότι η μη επίτευξη θεσμικής διαλειτουργικότητας μεταξύ δύο ή περισσότερων φορέων καθιστά προβληματική ή αδύνατη τη μεταξύ τους διαλειτουργικότητα σε οποιοδήποτε επίπεδο.

2.1.2.2 Οργανωτική Διαλειτουργικότητα

Η οργανωσιακή διαλειτουργικότητα, αφορά την προετοιμασία των οργανωτικών δομών και των επιχειρησιακών διαδικασιών, ώστε να είναι σε θέση να συνυπάρχουν με τις αυξανόμενες απαιτήσεις της επικοινωνίας σε επίπεδο δεδομένων, πληροφορίας και γνώσης. Το συγκεκριμένο επίπεδο διαλειτουργικότητας, περιλαμβάνει τον σχεδιασμό των διαδικασιών, των ροών εργασίας και της πληροφορίας των οργανισμών οι οποίοι επιθυμούν τη συνεργασία με διαφορετικούς οργανισμούς, αλλά έχουν διαφορετική δομή και λειτουργία σε σχέση με τους υπολοίπους. Σαν αποτέλεσμα αυτού, οφείλουν να διεξαχθούν οι απαραίτητες ενέργειες, ώστε οι λειτουργίες και οι διαδικασίες των συνεργαζομένων οργανισμών να εναρμονιστούν, ώστε να μπορέσουν να επικοινωνήσουν και να συνεργαστούν εποικοδομητικά, δρώντας σαν μια ενιαία οντότητα. Ωστόσο, προκειμένου να επιτευχθεί η διάσπαση του τρόπου λειτουργίας και των ορίων κάθε επιχείρησης, απαιτείται η αξιολόγηση όλων των διαδικασιών και των τρόπων επικοινωνίας της κάθε μίας, έτσι ώστε να ανασχεδιαστούν και να προσαρμοστούν όλες οι διαδικασίες και να αντιμετωπιστούν νομικές και διοικητικές δυσκολίες ή εμπόδια.

Μέσω του επιπέδου αυτού της διαλειτουργικότητας, οφείλεται να εντοπιστούν οι απαιτήσεις που χρειάζεται να αποφευχθούν, καθώς και να υποδειχθεί το τι πρέπει να τροποποιηθεί ή να εξαλειφθεί.

Σαν αποτέλεσμα των παραπάνω, καταλήγουμε στο συμπέρασμα πως η οργανωτική διαλειτουργικότητα έχει να κάνει σχέση με τα εξής θέματα:

- Τον τρόπο που παρέχονται οι υπηρεσίες.
- Τα κοινά σημεία με άλλες διαδικασίες.
- Τα μέσα που χρησιμοποιούνται για την παροχή κοινών λειτουργιών.
- Τους μηχανισμούς που απαιτούνται για τη μείωση των εξαρτήσεων μεταξύ των υπηρεσιών.
- Τον τρόπο που πρέπει να διαχειριστούν οι υπηρεσίες.
- Την αναγκαιότητα και το σκοπό της συνεργασίας μεταξύ των φορέων για την παροχή ολοκληρωμένων υπηρεσιών.
- Τα καθήκοντα κάθε εμπλεκόμενου και τον τρόπο εκτέλεσης των καθηκόντων του.
- Την πληροφορία που ανταλλάσσεται μεταξύ των φορέων.
- Τους κανόνες της συνεργασίας, οι οποίοι περιλαμβάνουν αφενός τις νομικές διατάξεις που διέπουν τις διαδικασίες παροχής της υπηρεσίας, και αφετέρου τις επιμέρους συμφωνίες μεταξύ των φορέων σχετικά με τις υποχρεώσεις τους.

Τέλος, αξίζουν να σημειωθούν τα απαραίτητα βήματα που χρειάζονται να πραγματοποιηθούν, για την επίτευξη της οργανωτικής διαλειτουργικότητας:

- **Βήμα 1:** Μοντελοποίηση και διαγραμματική απεικόνιση των υπηρεσιών, με απώτερο στόχο τη συνεχή διοίκηση και βελτιστοποίησή τους.
- **Βήμα 2:** Καθορισμός των ρόλων και των αρμοδιοτήτων για την εφαρμογή και την παρακολούθηση της εκτέλεσης των βημάτων παροχής υπηρεσιών.
- **Βήμα 3:** Ανάδειξη των κοινών χαρακτηριστικών των υπηρεσιών και των λειτουργιών τους.
- **Βήμα 4:** Ευθυγράμμιση των επιμέρους διαδικασιών, ώστε να είναι δυνατή η διασύνδεση και ο συνδυασμός τους για την παροχή ολοκληρωμένων υπηρεσιών.

- **Βήμα 5:** Υλοποίηση των ηλεκτρονικών υπηρεσιών «μιας στάσης», με την αξιοποίηση αυτόνομων διαλειτουργικών πληροφοριακών συστημάτων.
- **Βήμα 6:** Σύνδεση των προτύπων ηλεκτρονικών υπηρεσιών με τα επιχειρησιακά πρότυπα κάθε φορέα, ώστε αυτές να αποτελούν μέρος της καθημερινής λειτουργίας του.
- **Βήμα 7:** Καθορισμός της συμμετοχής των χρηστών στη διεκπεραίωση των υπηρεσιών, σε συνδυασμό με τα κατάλληλα τεχνολογικά πρότυπα.
- **Βήμα 8:** Προτυποποίηση κανόνων για τη διαχείριση της ανταλλαγής των πληροφοριών και των μηνυμάτων.
- **Βήμα 9:** Παροχή υπηρεσιών από πολλαπλές πηγές.
- **Βήμα 10:** Προσαρμογή του θεσμικού πλαισίου για την οριοθέτηση των ευθυνών και κατανομή των αρμοδιοτήτων μεταξύ των φορέων.

Μπορούμε να καταλήξουμε στο συμπέρασμα πως, με απλά λόγια, η οργανωτική διαλειτουργικότητα διασφαλίζει πως οι φορείς γνωρίζουν τον λόγο για το οποίο κάθε φορά επικοινωνούν, ποιες μονάδες ή ρόλοι τους είναι κυρίως υπεύθυνοι για την επικοινωνία και τι είδους πληροφορία ανταλλάσσεται κάθε φορά.

2.1.2.3 Σημασιολογική Διαλειτουργικότητα

Η σημασιολογική διαλειτουργικότητα, έρχεται να αντιμετωπίσει την ανάγκη για την κοινή συμφωνία οργανισμών, διαδικασιών και συστημάτων γύρω από τα δεδομένα που αυτά διαχειρίζονται. Αυτό που προσπαθούμε να επιτευχθεί γενικότερα, δεν είναι απλώς και μόνο η ανταλλαγή της πληροφορίας, αλλά και η ικανότητα αποδοτικής χρήσης και κατανόησής της.

Για την επίτευξη της διαλειτουργικότητας σε σημασιολογικό επίπεδο, απαιτείται η υιοθέτηση τεχνολογιών και μεθόδων σημασιολογικής ολοκλήρωσης, πάντα σε συνεργασία με τα άλλα δύο επίπεδα διαλειτουργικότητας.

Για το λόγο αυτό, η πληροφορία αυτή οφείλει να ερμηνεύεται με τον ίδιο τρόπο από όλους τους εμπλεκόμενους, χωρίς την ύπαρξη ασαφειών και δυσκολιών κατά την ερμηνεία της. Τον ρόλο αυτό, έρχεται να καλύψει η σημασιολογική

διαλειτουργικότητα, μιας και ακόμα και αν φτάσουν δύο φορείς να ανταλλάσσουν πληροφορίες και υπηρεσίες με διαλειτουργικό τρόπο, η μη κατανοητή και η μη αποδοτική χρήση των πληροφοριών αυτών, οδηγεί σίγουρα σε προβληματικές καταστάσεις. Για αυτό άλλωστε, έχει δοθεί τόσο μεγάλη σημασία στο συγκεκριμένο επίπεδο διαλειτουργικότητας.

Έτσι, προκειμένου η πληροφορία να εντοπίζεται σε μία «κοινή» για όλους γλώσσα, η ερμηνεία της γίνεται από τα ίδια τα συστήματα χωρίς την παρέμβαση του χρήστη, γεγονός που αποτελεί την ουσία της σημασιολογικής διαλειτουργικότητας.

Τα εννοιολογικά προβλήματα σχετίζονται με τις συντακτικές και σημασιολογικές διαφορές, τα οποία παρουσιάζονται τόσο σε υψηλό επίπεδο (π.χ. κατά την διαδικασία της μοντελοποίησης των διαδικασιών του οργανισμού) όσο και σε χαμηλό επίπεδο (τρόπος αναπαράστασης και ερμηνείας των δεδομένων). Γενικότερα, συντακτική ασυμβατότητα παρουσιάζεται όταν χρησιμοποιούνται διαφορετικοί τρόποι για την αναπαράσταση της πληροφορίας, ενώ σημασιολογική ασυμβατότητα δημιουργείται όταν δεν υπάρχει κοινός τρόπος ερμηνείας της πληροφορίας η οποία έχει ανταλλαχθεί.

Το πρόβλημα της σημασιολογικής ετερογένειας παράγεται από τις σημασιολογικές αντιθέσεις, οι οποίες προκύπτουν όταν η σημασία των δεδομένων και της πληροφορίας μπορεί να εκφραστεί με διαφορετικούς τρόπους και ερμηνείες. Ακόμα και όταν χρησιμοποιείται ένας κοινός τρόπος αναπαράστασης των δεδομένων, δεν είναι πάντοτε εξασφαλισμένο ότι οι διαφορετικοί φορείς θα αντιλαμβάνονται το νόημα των δεδομένων αυτών με τον ίδιο τρόπο. Η δυσκολία επομένως της επίτευξης σημασιολογικής διαλειτουργικότητας μεταξύ των φορέων, αφορά στη διασφάλιση ότι η ακριβής σημασία των ανταλλασσόμενων πληροφοριών είναι το ίδιο κατανοητή από τον κάθε φορέα.

Το πρόβλημα αυτό αναφέρεται ως πρόβλημα σημασιολογικής ολοκλήρωσης, και για τον λόγο αυτό έχουν δημιουργηθεί τρεις (3) προσεγγίσεις αναφορικά με την αναπαράσταση της σημασιολογίας [36]: η προσέγγιση μονού μοντέλου, η προσέγγιση πολλαπλών μοντέλων και η υβριδική προσέγγιση.

- Στην προσέγγιση **μονού** μοντέλου, ένα καθολικό σημασιολογικό μοντέλο παρέχει ένα διαμοιρασμένο λεξιλόγιο για τον ορισμό των σημασιών αυτόνομων πηγών δεδομένων, οι οποίες σχετίζονται με αυτό.

- Στην προσέγγιση **πολλαπλών** μοντέλων, κάθε τοπικό σύστημα δεδομένων περιγράφεται από ένα ξεχωριστό τοπικό σημασιολογικό μοντέλο. Η απουσία ενός καθολικού μοντέλου διευκολύνει την αυτόνομη ανάπτυξη τοπικών οντολογιών, οι οποίες εκφράζουν αναλυτικά και με συνέπεια τις έννοιες και τις σχέσεις κάθε τοπικού συστήματος δεδομένων.
- Στην **υβριδική** προσέγγιση, συνδυάζονται χαρακτηριστικά από τις δύο προαναφερθείσες προσεγγίσεις. Κάθε τοπική πηγή περιγράφεται από ένα ξεχωριστό τοπικό μοντέλο, το οποίο είτε έχει εξ αρχής δημιουργηθεί με βάση τις πρωτογενείς έννοιες ενός καθολικού μοντέλου είτε έχει μετατραπεί ώστε να είναι συμβατό με αυτές.

Προκειμένου να επιτευχθεί η σημασιολογική διαλειτουργικότητα, απαιτείται η ανάπτυξη συγκεκριμένων στοιχείων που να αναπαριστούν, να κωδικοποιούν και να ταξινομούν με ενιαίο τρόπο όλες τις πληροφορίες που λαμβάνονται και ανταλλάσσονται μεταξύ των διαφόρων φορέων. Τα στοιχεία αυτά είναι:

- **Κωδικολόγια** (Code Lists): Αποτελούν λίστες κωδικοποιημένων και ταξινομημένων ορισμών για κύριες κοινά χρησιμοποιούμενες οντότητες σε διαδικασίες, έγγραφα και συστήματα (π.χ. λίστα χωρών, λίστα νομισμάτων, λίστα υπηρεσιών).
- **Δομικά Στοιχεία Δεδομένων** (Core Data Components): Αποτελούν σύνολο δεδομένων για κύριες οντότητες που χρησιμοποιούνται για τη συστηματική δόμηση πρότυπων ηλεκτρονικών εγγράφων και αρχείων δεδομένων (π.χ. ταυτότητα προσώπου, διεύθυνση).
- **Τύποι Δεδομένων** (Data Types): Αποτελούν κωδικοποιήσεις των τύπων που μπορούν να έχουν τα πεδία ή οι περιγραφές δεδομένων (π.χ. ακέραιος, αλφαριθμητικά, ημερομηνία).
- **Πρότυπα Σχήματα Ηλεκτρονικών Εγγράφων** (Standard XML Schemas): Αποτελούν δομημένες περιγραφές εκφρασμένες σε XML, για την αποτύπωση της δομής και του περιεχομένου των ηλεκτρονικών εγγράφων. Περιέχουν αναφορές σε κωδικολόγια, τύπους δεδομένων ή δομικά στοιχεία δεδομένων. Επιπλέον, μπορούν να περιέχουν άλλα XML σχήματα, ενώ διακρίνονται σε τρεις (3) κατηγορίες, αυτές των οριζόντιων προτύπων XML, των κλαδικών προτύπων XML και των Διεθνών προτύπων XML.

- **Μεταδεδομένα (Metadata):** Αποτελούν σύνολα πληροφορίας που αφορούν τις διάφορες οντότητες και επιτρέπουν την ανεύρεση και κατανόησή τους από ανθρώπους ή συστήματα (π.χ. φορείς, υπηρεσίες, έγγραφα).
- **Οντολογίες (Ontologies):** Αποτελούν σύνολα συσχετισμένων αναπαραστάσεων, τα οποία περιλαμβάνουν ταξινομημένες οντότητες, τις μεταξύ τους σχέσεις και τα μεταδεδομένα τους, σε ένα συνολικό, ισχυρά ορισμένο και τυπικά περιγραφόμενο σύνολο. Οι οντολογίες χρησιμεύουν στην καλύτερη κατανόηση των οντοτήτων και των συσχετίσεών τους, συμβάλλοντας ουσιαστικά στην κοινή κωδικοποίηση από συσχετιζόμενους φορείς.
- **Ληξιαρχείο Διαλειτουργικότητας (Interoperability Registry):** Αποτελεί ένα σύστημα, το οποίο υποστηρίζει τον ορισμό, την καταχώρηση και την εξαγωγή προτύπων δεδομένων, κωδικολογίων, υπηρεσιών και μεταδεδομένων.

Για την ανάπτυξη στοιχείων σημασιολογικής διαλειτουργικότητας, είναι απαραίτητο να ακολουθούνται τα εξής βήματα από κάθε φορέα:

- **Βήμα 1:** Κάθε φορέας αναζητά την υπάρχουσα πληροφορία (πρότυπα), ξεκινώντας από τα XML πρότυπα σχήματα, συνεχίζοντας στα δομικά στοιχεία σχημάτων, έπειτα οδηγείται στην κωδικοποίηση και καταλήγει στους τύπους δεδομένων.
- **Βήμα 2:** Σε περίπτωση μη ανεύρεσης κάποιου υπάρχοντος στοιχείου, εξετάζεται η δυνατότητα σύνθεσής του από τα υπάρχοντα στοιχεία.
- **Βήμα 3:** Ελέγχεται η δυνατότητα επαναχρησιμοποίησης κάποιου σχετιζόμενου διεθνούς, γενικού ή κλαδικού προτύπου.
- **Βήμα 4:** Δημιουργείται το επιθυμητό στοιχείο, δηλαδή το XML σχήμα.
- **Βήμα 5:** Υλοποιούνται τα μεταδεδομένα περιγραφής του νέου στοιχείου και ενημερώνονται τα αντίστοιχα σημεία δημοσιοποίησης της πληροφορίας.
- **Βήμα 6:** Το νέο στοιχείο που έχει δημιουργηθεί, υποβάλλεται για έγκριση στο Ληξιαρχείο Διαλειτουργικότητας.

Τέλος, να σημειωθεί πως γενικότερα η επαναχρησιμοποίηση έτοιμων μοντέλων αναπαράστασης σημασιολογικής πληροφορίας (XML σχημάτων και οντολογιών) που έχουν αναπτυχθεί στο πλαίσιο άλλων συναφών εφαρμογών, αποτελεί βέλτιστη πρακτική, καθώς οδηγεί σε σημαντική μείωση του χρόνου και της προσπάθειας ανάπτυξης νέων μοντέλων. Επιπλέον, η υιοθέτηση έτοιμων μοντέλων διευκολύνει κατά πολύ την προσπάθεια επίτευξης σημασιολογικής διαλειτουργικότητας μεταξύ των εκάστοτε φορέων.

2.1.2.4 Τεχνική Διαλειτουργικότητα

Το τελευταίο επίπεδο διαλειτουργικότητας, η τεχνική διαλειτουργικότητα, αναφέρεται και καλύπτει τις τεχνικές πτυχές του ζητήματος της διαλειτουργικότητας (τεχνολογίες, πρότυπα, πρωτόκολλα, διασυνδέσεις υπηρεσιών, ενοποίηση δεδομένων, ενδιάμεσο λογισμικό, παρουσίαση και ανταλλαγή δεδομένων, προσβασιμότητα και ασφάλεια υπηρεσιών). Απώτερος στόχος είναι να διασφαλιστεί ότι τα Πληροφοριακά συστήματα των οργανισμών μπορούν να επικοινωνήσουν, να ανταλλάξουν πληροφορίες και να επιτύχουν την συνεργατική εργασία.

Πιο συγκεκριμένα, η τεχνική διαλειτουργικότητα σχετίζεται με τις τεχνικές λεπτομέρειες που αφορούν την διασύνδεση των πληροφορικών συστημάτων για την ανταλλαγή πληροφοριών. Αναφέρεται σε πρότυπα και προδιαγραφές που οδηγούν στην ορθή ανταλλαγή πληροφοριών μεταξύ των υπολογιστικών συστημάτων και περιέχουν «ρυθμίσεις, πρότυπα και κατευθυντήριες γραμμές για ένα κοινό μηχανισμό λειτουργίας, καθώς και την ανάπτυξη τυποποιημένων μεταδεδωμένων και τη χρήση μιας κοινής γλώσσας». Οι ηλεκτρονικές υπηρεσίες διαδικτύου είναι διαθέσιμες σε χιλιάδες μορφές και εμφανίσεις και προσφέρουν διάφορων ειδών τύπους αλληλεπίδρασης από απλές ιστοσελίδες μέχρι διαδραστικές εφαρμογές. Οι υπηρεσίες αυτές χωρίζονται σε τέσσερα στάδια, ανάλογα με τις υπηρεσίες που προσφέρουν:

- Στο **πρώτο στάδιο**, βρίσκονται οι ηλεκτρονικές υπηρεσίες που προσφέρονται για παρουσίαση πληροφοριών.
- Στο **δεύτερο στάδιο**, συμπεριλαμβάνονται οι υπηρεσίες που διαθέτουν online φόρμες και έγγραφα, τα οποία μπορεί να κατεβάσει ο χρήστης και να τα στείλει μέσω ταχυδρομείου fax ή ηλεκτρονικού ταχυδρομείου (email).

- Στο **τρίτο στάδιο**, βρίσκονται οι υπηρεσίες στις οποίες μπορούν να γίνουν συναλλαγές μεταξύ των διοικήσεων και των πολιτών μέσω υπολογιστή, όπως η συμπλήρωση και η αποστολή των φορμών ηλεκτρονικά (online).
- Τέλος, στο **τέταρτο στάδιο**, βρίσκονται οι υπηρεσίες οι οποίες καθιστούν δυνατές οποιοδήποτε είδους συναλλαγές.

Για την επίτευξη της τεχνικής διαλειτουργικότητας, έχει δημιουργηθεί ένα Ευρωπαϊκό Πλαίσιο Διαλειτουργικότητας (European Interoperability Framework - EIF) [37], το οποίο εμπεριέχει τεχνικές προδιαγραφές και περιγράφει αναλυτικά τρόπους αποθήκευσης, μεταφοράς και παρουσίασης δεδομένων. Στα πλαίσια του προτύπου αυτού, διαλειτουργικά δίκτυα και βάσεις δεδομένων συντονίζονται σε περιφερειακό επίπεδο ώστε να είναι διαχειρίσιμα, εύκολα και αποτελεσματικά.

Έχουν δημιουργηθεί πολλαπλά μοντέλα για την διασφάλιση της τεχνικής διαλειτουργικότητας, από ποικίλους οργανισμούς. Ορισμένα από τα μοντέλα αυτά είναι τα ακόλουθα:

- Το μοντέλο Levels of Information System Interoperability (LISI) της DARPA [38], μέσω του οποίου καθίσταται εφικτή η κατηγοριοποίηση της ωριμότητας της τεχνικής διαλειτουργικότητας των πληροφοριακών συστημάτων.
- Το μοντέλο NATO Consultation, Command and Control (C3) Technical Architecture (NC3TA) [39] της NATO, το οποίο περιλαμβάνει αρχιτεκτονικές περιγραφές, μοντέλα αναφοράς και έτοιμες Off-The-Shelf (OTS) [40] τεχνολογίες.
- Το μοντέλο Organisational Interoperability Maturity (OIM) [41], που αποτελεί την εξέλιξη του μοντέλου LISI, κατά το οποίο ο διαχωρισμός των επιπέδων γίνεται δίνοντας έμφαση στην τεχνολογία.

Καθίσταται σαφές ότι η επίτευξη της τεχνικής διαλειτουργικότητας αποτελεί τη βάση για την επίτευξη των άλλων επιπέδων διαλειτουργικότητας, δεδομένου ότι στο επίπεδο αυτό καθορίζονται τα πρότυπα μέσω των οποίων επιτυγχάνεται η διαλειτουργικότητα από τεχνικής άποψης. Επίσης, θα πρέπει να επισημανθεί, ότι στο επίπεδο αυτό καθορίζονται όλα τα σχετικά τεχνικά πρότυπα στα οποία βασίζονται τα άλλα επίπεδα.

2.1.3 Πρότυπα Διαλειτουργικότητας

Με τον προτυποποίηση (standardization), ή απλούστερα πρότυπα (standards), χαρακτηρίζεται γενικά η ανάπτυξη ειδικών προτύπων βάσει των οποίων ακολουθείται στη συνέχεια η παραγωγή μεγάλων ποσοτήτων προϊόντων που να μπορούν μεταξύ τους να επικοινωνούν, χωρίς άλλες προσαρμογές. Με απλούστερα λόγια, το ακολουθούμενο κάθε φορά πρότυπο αποτελεί ένα συγκεκριμένο υπόδειγμα προδιαγραφών επί του οποίου και ακολουθεί η βιομηχανική γραμμή παραγωγής του συγκεκριμένου προϊόντος. Το πρότυπο μπορεί να αφορά όργανα, μηχανήματα που χρησιμοποιούνται για έλεγχο και μετρήσεις διαφόρων χαρακτηριστικών των προϊόντων όπως μέγεθος, χρώμα, βάρος ή ακόμα και τύπους, σκίτσα, σύμβολα χρήσης, μεταφοράς, ή ενεργειών που θα πρέπει να ακολουθηθούν. Ως πρότυπο λοιπόν, ορίζεται ένα σύνολο προδιαγραφών που περιγράφει ακριβώς το πώς μεταφέρονται ή αποθηκεύονται τα δεδομένα μέσα από ένα σύνολο κανόνων, συνθηκών, απαιτήσεων και τρόπων αναπαράστασης της πληροφορίας. Τα πρότυπα είναι επί της ουσίας το κλειδί για την πρόσβαση στα δεδομένα που περιέχει ένα σύστημα. Αν δεν είναι γνωστός ο τρόπος αποθήκευσης και μεταφοράς των δεδομένων αυτών, δεν εξασφαλίζεται η δυνατότητα πρόσβασης. Με τη χρήση των προτύπων αποφεύγονται ασυμβατότητες, ασάφειες και παρερμηνείες των δεδομένων.

Στο εν λόγω κεφάλαιο, αναφερόμαστε σε πρότυπα τα οποία χρησιμοποιούνται προκειμένου να επιτευχθούν και να συντονιστούν τα βασικότερα είδη διαλειτουργικότητας σε πολλαπλούς τομείς με έμφαση στις διαδικασίες των ηλεκτρονικών δημόσιων προμηθειών. Για τον λόγο αυτό, μερικοί από τους πιο γνωστούς οργανισμούς όπως ο ISO, ο UN/CEFACT και άλλοι, έχουν αφιερώσει ένα κομμάτι του έργου και των ερευνών τους στη δημιουργία προτύπων και πλαισίων σχετικών με τη νομική, οργανωτική, σημασιολογική, και τεχνική διαλειτουργικότητα.

2.1.3.1 ISO (International Organization for Standardization)

Ο ISO [42] είναι ο μεγαλύτερος οργανισμός στον κόσμο ανάπτυξης και έκδοσης διεθνών προτύπων. Αποτελεί ένα μη κυβερνητικό οργανισμό που γεφυρώνει τους ιδιωτικούς με τους δημόσιους τομείς, δίνοντας τη δυνατότητα μιας συναίνεσης που θα επιτευχθεί με λύσεις που ανταποκρίνονται τόσο στις απαιτήσεις των επιχειρήσεων όσο και στις ευρύτερες ανάγκες της κοινωνίας. Σε ότι αφορά το πεδίο της



διαλειτουργικότητας, δύο από τα βασικότερα πρότυπα που έχει εγκρίνει είναι το ISO 15000 [43] και το ISO/IEC 11179-1 [44].

ISO 15000 Electronic business eXtensible Markup Language (ebXML)

Το ISO 15000, αποτελεί ένα σύνολο από προδιαγραφές που παρέχει τη δυνατότητα σε επιχειρήσεις κάθε είδους ανά όλο τον κόσμο, να διεξάγουν εργασίες μέσω διαδικτύου. Οι επιχειρήσεις, χρησιμοποιώντας την ebXML, μπορούν να ανταλλάσσουν επιχειρηματικά μηνύματα, να πραγματοποιούν εμπορικές συναλλαγές, να επικοινωνούν με δεδομένα υπό κοινούς όρους, να καθορίζουν και να αρχειοθετούν επιχειρηματικές διαδικασίες.

Πέντε δυνατότητες αναγνωρίστηκαν ως μέρος του αρχικού στόχου της ebXML, που παραμένουν σημαντικές για τη διεξαγωγή σχέσεων ηλεκτρονικού εμπορίου ανάμεσα σε επιχειρηματικές πλευρές. Το ebXML περιλαμβάνει **πέντε (5) τύπους προδιαγραφών** που μπορούν να εφαρμοστούν μεμονωμένα ή από κοινού:

- **Επιχειρηματική Διαδικασία (Business Process):** Καθορίζει το πώς περιγράφονται τα βασικά στοιχεία που εμπεριέχουν πληροφορία και χρησιμοποιούνται σε επιχειρησιακά μηνύματα, ενώ περιλαμβάνει την ανταλλαγή πληροφοριών μεταξύ δύο ή περισσότερων εταιρών.
- **Προφίλ Πρωτοκόλλου Συνεργασίας και Συμφωνίας (Collaboration Protocol Profile and Agreement):** Καθορίζει το πώς εμπορικοί εταίροι εμπλέκονται σε συνεργασίες ηλεκτρονικού επιχειρείν μέσω ανταλλαγής ηλεκτρονικών μηνυμάτων. Είναι ένα από τα βασικά τμήματα, το οποίο καθιστά τους χρήστες ικανούς να ρυθμίσουν γρήγορα τις αλληλεπιδράσεις και τις σχέσεις με τους εταίρους, έχοντας λεπτομερή αναφορά για τις μεθόδους επικοινωνίας, την ταυτότητα των εταιρών, τα τελικά σημεία, την ασφάλεια και τις επιλογές αξιοπιστίας.
- **Υπηρεσίες Μηνυμάτων (Messaging Services):** Καθορίζουν τη μεταφορά, δρομολόγηση και πακετοποίηση ηλεκτρονικών επιχειρηματικών μηνυμάτων, χρησιμοποιώντας πρότυπες τεχνολογίες του διαδικτύου. Το διαχειρίσιμα ebXML Messaging Service Specification (ebMS) περιγράφει μία ουδέτερη μέθοδο επικοινωνίας πρωτοκόλλου για ανταλλαγή ηλεκτρονικών μηνυμάτων επιχειρήσεων. Καθορίζει συγκεκριμένες δομές φακέλων που υποστηρίζουν αξιόπιστη και ασφαλή μεταφορά επιχειρηματικών πληροφοριών. Ο πρωταρχικός στόχος του ebMS είναι να διευκολύνει την ανταλλαγή ηλεκτρονικών μηνυμάτων επιχειρήσεων σε ένα

πλαίσιο XML που αξιοποιεί κοινά πρότυπα του διαδικτύου. Αυτά τα μηνύματα μπορεί να χρησιμοποιηθούν με διαφορετικούς τρόπους που είναι εκτός του πεδίου εφαρμογής αυτής της προδιαγραφής. Πιο συγκεκριμένα, μπορούν να συνδέονται με υπάρχοντες εφαρμογές, με υπηρεσίες, να τοποθετούνται σε ουρά, να εισέρχονται σε διαδικασίες ροών εργασιών, να αναμένονται από ήδη εκτελούμενες επιχειρηματικές διαδικασίες, να δεσμεύονται για μετέπειτα επεξεργασία, να δρομολογούνται από έναν διάυλο υπηρεσιών της επιχείρησης πριν φτάσουν στην εφαρμογή κατανάλωσής τους, κλπ.

- **Μητρώα και Αρχεία Φύλαξης (Registries and Repositories):** Καθορίζουν διαλειτουργικά μητρώα (registries) και αρχεία φύλαξης (repositories), σε περιβάλλον που δίνει τη δυνατότητα υποβολής, ερωτήματος και ανάκτησης περιεχομένων.
- **Βασικά Στοιχεία (Core Components):** Παρουσιάζουν μία μεθοδολογία για την ανάπτυξη ενός κοινού συνόλου δημιουργίας κοινών μπλοκ για την αναπαράσταση γενικών τύπων που χρησιμοποιούνται σήμερα σε επιχειρηματικά δεδομένα. Επίσης παρέχουν μεθοδολογίες για τη δημιουργία καινούργιων επιχειρηματικών λεξικών, αλλά και την αναδόμηση των υπαρχόντων.

ISO/IEC 11179 – Information Technology: Metadata Registries

Το ISO/IEC 11179 είναι ένα πρότυπο δεδομένων και μεταδεδομένων που αποτελείται από έξι μέρη και εστιάζει στην κοινή χρήση στοιχείων δεδομένων σε συστήματα μέσα σε μία εταιρεία και μεταξύ επιχειρήσεων. Το πρότυπο καθορίζει τι είναι ένα αρχείο μεταδεδομένων, πως μπορεί να κατηγοριοποιηθεί και πως μπορεί να περιγραφεί, να ονομαστεί, να αναγνωρισθεί, να αποθηκευτεί, να ανακτηθεί.

- **Μέρος 1 - Πλαίσιο (Framework):** Το 1^ο μέρος, καθορίζει και περιγράφει τις βασικές έννοιες στοιχείων δεδομένων, πεδίων τιμών, εννοιολογικούς, φυσικούς ή λογικούς τομείς και συστήματα ταξινόμησης που είναι αναγκαίες για την κατανόηση των υπόλοιπων μερών της προδιαγραφής. Εισάγεται η έννοια του αρχείου μεταδεδομένων, δηλαδή ο ρόλος του στη σημασιολογία των δεδομένων, η έννοια της απεικόνισης των δεδομένων και την καταγραφή των περιγραφών των δεδομένων.
- **Μέρος 2 - Κατηγοριοποίηση (Classification):** Το 2^ο μέρος, περιέχει ένα λεπτομερές σύνολο από χαρακτηριστικά που αποτελούν ένα σχήμα

συνεκτικής κατηγοριοποίησης. Υποστηρίζει την καταγραφή και την διαχείριση του συνόλου ή μέρους ενός σχήματος κατηγοριοποίησης και παρέχει ένα μηχανισμό για την κατάταξη των διαχειριζόμενων στοιχείων σε ένα αρχείο μεταδεδομένων.

- **Μέρος 3 - Ληξιαρχείο μεταμοντέλου και βασικές ιδιότητες (Registry metamodel and basic attributes):** Το 3^ο μέρος, καθορίζει τη δομή και το εννοιολογικό μοντέλο ενός αρχείου μεταδεδομένων. Καθορίζει βασικές δομές μεταδεδομένων σε όρους εννοιολογικών μοντέλων και περιέχει έναν αριθμό βασικών χαρακτηριστικών που είναι απαραίτητα για τον ορισμό στοιχείων μεταδεδομένων.
- **Μέρος 4 - Διατύπωση των ορισμών των δεδομένων (Formulation of data definitions):** Το 4^ο μέρος, παρέχει τις απαιτήσεις και την καθοδήγηση για τη δημιουργία σαφών ορισμών των στοιχείων δεδομένων και των συστατικών τους. Ο ορθός ορισμός είναι αναγκαίος για τη σημασιολογική κατανόηση όλων των αποθηκευμένων δεδομένων και μεταδεδομένων.
- **Μέρος 5 - Αρχές ονομασίας και προσδιορισμού (Naming and identification principles):** Το 5^ο μέρος, περιγράφει τον τρόπο με τον οποίο πρέπει να σχηματιστούν συμβάσεις για την ονομασία στοιχείων δεδομένων και τα συστατικά τους μέρη. Περιλαμβάνει ειδικές οδηγίες για την ανάπτυξη της ονομασίας και την ταυτοποίηση των εννοιών των στοιχείων δεδομένων.
- **Μέρος 6 - Εγγραφή (Registration):** Το 6^ο μέρος, ορίζει τον τρόπο με τον οποίο τα διαχειριζόμενα στοιχεία θα πρέπει να καταχωρούνται μέσω μιας κεντρικής αρχής καταγραφής. Δίνει οδηγίες για την κατανομή μοναδικών κλειδιών για κάθε στοιχείο δεδομένων, και παρέχει πληροφορίες για τη διατήρηση της διαχείρισης των στοιχείων.

2.1.3.2 CEN (European Committee for Standardization)

Ο CEN [45] είναι ένας πολυεθνικός μη κερδοσκοπικός οργανισμός που στοχεύει στην προώθηση της ευρωπαϊκής οικονομίας στο παγκόσμιο εμπόριο, στην εξασφάλιση της ευημερίας των ευρωπαίων πολιτών και του περιβάλλοντος, παρέχοντας μια αποτελεσματική υποδομή στα ενδιαφερόμενα μέρη, σχετικά με την ανάπτυξη, τη συντήρηση και τη διανομή των συνεκτικών συνόλων προτύπων και προδιαγραφών.



Ο CEN λειτουργεί κυρίως μέσω Τεχνικών Επιτροπών και εργαστηρίων (Workshops), όπου το Business Interoperability Interfaces (BII) Workshop (Information Systems and Business Interoperability Interfaces for public procurement), έχει ως βασικούς στόχους να παρέχει ένα βασικό πλαίσιο για τεχνική διαλειτουργικότητα στις πανευρωπαϊκές ηλεκτρονικές συναλλαγές, προκειμένου να επιβεβαιωθεί η παγκόσμια διαλειτουργικότητα.

Ειδικότερα, οι στόχοι των εργαστηρίων (workshops) καταγράφονται ως εξής:

- Να διευκολυνθεί η σύγκλιση μεταξύ των προτύπων United Nations Centre for Trade Facilitation and E-business (UN/CEFACT) [46] και Universal Business Language (UBL) [47], ώστε να εφαρμοστούν με πρακτικό τρόπο.
- Να αναγνωριστούν και να αρχειοθετηθούν οι απαραίτητες διαλειτουργικές επιχειρησιακές διεπαφές που σχετίζονται με τις πανευρωπαϊκές ηλεκτρονικές συναλλαγές στις δημόσιες προμήθειες.
- Να υποστηρίζονται πιλοτικά προγράμματα που εντάσσουν τις τεχνικές προδιαγραφές, προκειμένου να αφαιρεθούν οι τεχνικές δυσκολίες που εμποδίζουν τη διαλειτουργικότητα.

CEN/BII - Business Interoperability Interfaces for public procurement in Europe

Ο λόγος χρήσης του προτύπου CEN/BII [48], οφείλεται σε διαφορετικούς σκοπούς, ένας από τους οποίους είναι πως υπάρχει ισχυρή υποστήριξη από την Ευρωπαϊκή Επιτροπή. Επιπλέον, στις ημέρες μας, το μόνο πρότυπο που εντάσσεται και χρησιμοποιείται σε δημόσιες ηλεκτρονικές προμήθειες στην Ευρώπη είναι το UBL. Η UN/CEFACT έχει προτυποποιήσει αρκετές διαδικασίες αλλά δεν υπάρχει ακόμα κάποια ένταξή της. Σαν αποτέλεσμα αυτού, το CEN/BII workshop δείχνει να είναι το πλέον κατάλληλο να υιοθετηθεί σαν ένα πρότυπο, προκειμένου να χρησιμοποιείται σε πιλοτικά προγράμματα.

Το CEN/BII έχει ως επίκεντρο να συλλέγει ευρωπαϊκές επιχειρησιακές απαιτήσεις και να παρέχει καθοδήγηση για τη συνεπή εφαρμογή των απαιτήσεων αυτών, χρησιμοποιώντας ήδη υπάρχοντα διεθνή αναπτόγματα, όπως το UN/CEFACT ή το OASIS UBL. Το βασικό επιχείρημα αυτής της προσέγγισης είναι, ότι προκειμένου να επιτευχθεί ο στόχος της διαλειτουργικότητας, πρέπει να υπάρξει ταυτόχρονη συνεργασία από οργανισμούς ώστε να συνεισφέρει ο καθένας με το δικό του κομμάτι.

2.1.3.3 UN/CEFACT (United Nations/Centre for Trade Facilitation and Electronic Business)

Ο συγκεκριμένος οργανισμός, υποστηρίζει τις δραστηριότητες με στόχο τη βελτίωση της ικανότητας των επιχειρήσεων να ανταλλάζουν προϊόντα και σχετικές υπηρεσίες αποτελεσματικά, με στόχο τη διευκόλυνση των εθνικών και διεθνών συναλλαγών, μέσω της απλοποίησης και της εναρμόνισης διαδικασιών και ροών πληροφοριών. Ένα από τα κύρια πρότυπα στα οποία έχει συμβάλει ο UN/CEFACT για τη διαλειτουργικότητα είναι η προδιαγραφή ISO 15000 για την ebXML. Άλλες τεχνικές προδιαγραφές αποτελούν οι παρακάτω:



UN/CEFACT - XML Naming and Design Rules Technical Specification

Η προδιαγραφή UN/CEFACT - XML Naming and Design Rules Technical [49] Specification, περιγράφει και καθορίζει τους κανόνες και τις κατευθυντήριες γραμμές που θα εφαρμόζονται από τον UN/CEFACT όταν αναπτύσσει σχήματα XML. Παρέχει έναν τρόπο αναγνώρισης, καταγραφής και μεγιστοποίησης της επαναχρησιμοποίησης των επιχειρηματικών πληροφοριών που εκφράζονται ως συστατικά σχημάτων XML για την υποστήριξη και βελτίωση της διαλειτουργικότητας των πληροφοριών σε πολλαπλές επιχειρηματικές συνθήκες.

UN/CEFACT Core Components Technical Specification (CCTS) and UN/CEFACT Modelling Methodology (UMM)

Το πρότυπο UN/CEFACT metamodels for business information (CCTS) [50] and business process modelling (UMM) [50], παρέχει εννοιολογικά μεταμοντέλα τα οποία για να μπορούν να χρησιμοποιούνται με κοινά εργαλεία μοντελοποίησης πρέπει να αντιστοιχηθούν στο Unified Modelling Language (UML) [51] μεταμοντέλο.

UN/CEFACT - Core Components Data Type Catalogue

Η προδιαγραφή Core Components Technical Specification (CCTS) [52] που αναπτύχθηκε από τον UN/CEFACT παρέχει μία μεθοδολογία για μοντελοποίηση σημασιολογικών δεδομένων που επιτυγχάνει κοινή κατανόηση δομών δεδομένων και τύπων μηνυμάτων σε ανεξάρτητο συντακτικό επίπεδο. Για την υποστήριξη αυτών των τύπων δεδομένων, καθορίζεται επίσης ένα σύνολο θεμελιακών στοιχείων, που είναι ο βασικός ορισμός των πεδίων τιμών. Οι τιμές των πεδίων μπορούν να εκφραστούν χρησιμοποιώντας κωδικοποιημένη λίστα τιμών με αναγνωριστικά σχήματα.

2.1.3.4 W3C (World Wide Web Consortium)

Το World Wide Web Consortium (W3C) [53], είναι ένας οργανισμός δημιουργίας και καθιέρωσης τεχνολογικών προτύπων, που αναπτύσσει διαλειτουργικές τεχνολογίες (προδιαγραφές, οδηγίες, λογισμικό και εργαλεία) με σκοπό να αξιοποιήσει και να καθοδηγήσει την πορεία του παγκόσμιου δικτυακού



ιστού. Το W3C, αναπτύσσει διαλειτουργικές τεχνολογίες για να οδηγήσει το διαδίκτυο στην πλήρη αξιοποίηση της δυναμικής του. Το W3C προωθεί τη διαλειτουργικότητα ανάμεσα σε συσκευές και λογισμικό που έχουν πρόσβαση στο Web, σχεδιάζοντας και προωθώντας ανοιχτά και συμβατά μεταξύ τους πρωτόκολλα, αποφεύγοντας έτσι τον κατακερματισμό της αγοράς.

Οι τεχνολογίες του W3C ομαδοποιούνται σε τεχνολογίες Εφαρμογών Ιστού, Πρόσβασης στον παγκόσμιο ιστό μέσω κινητής συσκευής, φωνή, υπηρεσιών ιστού, σημασιολογικού ιστού και Ιδιωτικότητας και Ασφάλειας. Οι τεχνολογίες αυτές βασίζονται σε θεμελιώδεις γλώσσες και πρωτόκολλα του W3C, όπως είναι τα Uniform Resource Identifiers (URIs) [54], η Hypertext Transfer Protocol (HTTP) [55], η XML, και χρησιμοποιούνται στην ανάπτυξη εφαρμογών που ικανοποιούν τις απαιτήσεις Προσβασιμότητας, Διεθνοποίησης, Ανεξαρτησίας Συσκευών, Κινητής Πρόσβασης και Διασφάλισης Ποιότητας.

Ορισμένα από τα πιο γνωστά πρότυπα διαλειτουργικότητας και προδιαγραφές είναι τα ακόλουθα:

XML

Η XML, αποτελεί μία γλώσσα σήμανσης, που παρέχει ένα σύνολο από κανόνες για ηλεκτρονική κωδικοποίηση των εγγράφων, προκειμένου να είναι αναγνωρίσιμα και διαλειτουργικά.

XML Schema

Ένα XML Schema [56], είναι μια περιγραφή του τύπου ενός XML αρχείου, που τυπικά εκφράζει τους περιορισμούς της δομής και του περιεχομένου του εγγράφου XML. Οι περιορισμοί αυτοί, εκφράζονται χρησιμοποιώντας κάποιους συνδυασμούς από κανόνες γραμματικής που περιορίζουν το περιεχόμενο και αφορούν τον τρόπο εμφάνισης των στοιχείων του εγγράφου XML, σε ότι αφορά το πλήθος τους και τις ιδιότητές τους.

XSL

Η EXtensible Stylesheet Language (XSL) [57] λειτουργεί ώστε να μετασχηματίζει ένα XML έγγραφο σε ένα άλλο XML έγγραφο ή σε έναν άλλον τύπο εγγράφου, ενώ συνήθως χρησιμοποιείται για να μετασχηματιστεί ένα XML έγγραφο σε μια μορφή που να αναγνωρίζεται από έναν φυλλομετρητή (browser).

XPath

Η γλώσσα XML Path (XPath) [58] βασίζεται σε μια δεντρική αναπαράσταση ενός εγγράφου XML, και παρέχει τη δυνατότητα της πλοήγησης μέσα στο δέντρο αυτό, επιλέγοντας κόμβους βάσει μιας μεγάλης ποικιλίας κριτηρίων.

SOAP

Το Simple Object Access Protocol (SOAP) [59] αποτελεί ένα πρωτόκολλο προδιαγραφής για την ανταλλαγή δομημένης πληροφορίας στην εφαρμογή των υπηρεσιών Web σε δίκτυα υπολογιστών. Για τη μορφή των μηνυμάτων, χρησιμοποιεί XML και συνήθως βασίζεται σε άλλα πρωτόκολλα επιπέδου εφαρμογής, κυρίως HTTP ή Simple Mail Transfer Protocol (SMTP) [60], για τη διαπραγμάτευση και τη μετάδοση μηνυμάτων. Το SOAP, το ίδιο δεν προσδιορίζει καμία σημασιολογική εφαρμογή, όπως ένα μοντέλο προγραμματισμού ή την υλοποίηση συγκεκριμένης σημασιολογίας. Αντίθετα, ορίζει έναν απλό μηχανισμό για την έκφραση σημασιολογίας εφαρμογών, παρέχοντας ένα αρθρωτό μοντέλο συσκευασίας και έναν μηχανισμό για την κωδικοποίηση των δεδομένων στο πλαίσιο των ενοτήτων. Αυτό επιτρέπει στο SOAP να χρησιμοποιηθεί σε μία μεγάλη ποικιλία συστημάτων.

WSDL

Η Web Services Description Language (WSDL) [61] είναι ένα XML Schema για την περιγραφή δικτυακών υπηρεσιών σαν ένα σύνολο από τελικά σημεία που λειτουργούν σε μηνύματα, τα οποία περιέχουν πληροφορία είτε προσανατολισμένη στα έγγραφα είτε προσανατολισμένη στις διαδικασίες. Οι λειτουργίες και τα μηνύματα περιγράφονται περιληπτικά, και τότε δένονται σε ένα συγκεκριμένο πρωτόκολλο δικτύων και μορφή μηνυμάτων για να καθορίσουν ένα τελικό σημείο. Πολλά σχετικά τελικά σημεία συνδυάζονται σε υπηρεσίες (services).

2.1.3.5 EIF (European Interoperability Framework)

Η Ευρωπαϊκή Ένωση, στο πλαίσιο της πρωτοβουλίας διαλειτουργικής παράδοσης παν-Ευρωπαϊκών υπηρεσιών ηλεκτρονικών προμηθειών σε δημόσιους φορείς, επιχειρήσεις και πολίτες έχει εκδώσει ένα πλαίσιο διαλειτουργικότητας για τις ηλεκτρονικές προμήθειες, το EIF [35].



Βασικοί στόχοι του πλαισίου αυτού είναι:

- Ο ορισμός βασικών κατευθύνσεων και αρχών για την παροχή υπηρεσιών ηλεκτρονικών προμηθειών σε πανευρωπαϊκό επίπεδο και η διαλειτουργικότητα μεταξύ συστημάτων ηλεκτρονικών προμηθειών διαφορετικών χωρών.
- Η παροχή οδηγιών και συστάσεων προς τις χώρες της Ευρωπαϊκής Ένωσης ώστε να υιοθετούν το EIF κατά τη σχεδίαση των εγχώριων πλαισίων για να εξασφαλίζεται η διαλειτουργικότητα σε πανευρωπαϊκό επίπεδο.

Συνολικά, **τρεις (3) εκδόσεις** έχουν δημιουργηθεί σχετικά με το EIF.

Η **1^η έκδοση** του EIF ασχολείται κυρίως με οδηγίες για τεχνολογίες και αρχιτεκτονικές για πανευρωπαϊκές υπηρεσίες ηλεκτρονικών προμηθειών. Βασίζεται στην ιδέα ότι κάθε μέλος της Ευρωπαϊκής Ένωσης διαθέτει το δικό του πλαίσιο διαλειτουργικότητας (interoperability framework) και έχει ως στόχο, όχι να αντικαταστήσει το ήδη υπάρχον πλαίσιο κάθε μέλους, αλλά να προσθέσει επιπλέον χαρακτηριστικά σε αυτό. Όσον αφορά τη συγκεκριμένη έκδοση του EIF (EIF v1.0), εκδόθηκε το 2004. Ακολουθώντας τους στόχους του *Interoperable Delivery of European eGovernment Services to public Administrations, Businesses and Citizens (IDABC)* [62], στόχευσε στην εισαγωγή των πανευρωπαϊκών κυβερνητικών υπηρεσιών. Γρήγορα φάνηκε πως η διαλειτουργικότητα απείχε από αυτήν την προσέγγιση. Όταν το γεγονός αυτό έγινε αντιληπτό, η IDABC ξεκίνησε να αναπτύσσει τη **2^η έκδοση** του EIF, με τη χρήσιμη βοήθεια εξωτερικών μελών όπως βιομηχανίες, επιχειρήσεις και άλλα. Έτσι, εμφανίστηκαν οι επιπλέον αδυναμίες της πρώτης έκδοσης του EIF, όπως:

- Το πρωταρχικό μοντέλο διαλειτουργικότητας μπορούσε να αποτελείται από λιγότερα βήματα.

- Χρειαζόταν περισσότερη προσοχή σε νομικά θέματα και θέματα της εξέλιξης των προτύπων.
- Δεν ήταν ξεκάθαρες οι ευθύνες του καθενός.
- Δεν δινόταν επαρκής προσοχή στα νομικά εμπόδια σχετικά με τη διαλειτουργικότητα, τα οποία ήταν εξίσου πολυάριθμα και σοβαρά.

Όλα αυτά, οδήγησαν στη δημιουργία του EIF v2.0, που εκδόθηκε το 2008. Η **2η έκδοση** του EIF ενημερώνει αυτές τις οδηγίες και παράλληλα εισάγει τις έννοιες του πολιτικού και νομικού επιπέδου της διαλειτουργικότητας.

Οι στόχοι του συγκεκριμένου πλαισίου είναι:

- Η υποστήριξη και προώθηση της ανάπτυξης των πανευρωπαϊκών υπηρεσιών ηλεκτρονικής δημόσιας διοίκησης καθώς και της διαλειτουργικότητας μεταξύ τους.
- Η δημιουργία της δυνατότητας ανταλλαγής πληροφοριών μεταξύ των δημόσιων διοικήσεων και των κοινοτικών οργάνων.
- Η διευκόλυνση της παροχής πανευρωπαϊκών υπηρεσιών στις επιχειρήσεις και στους πολίτες, λαμβάνοντας υπόψη τις ανάγκες τους.
- Η επίτευξη της διαλειτουργικότητας μεταξύ των διάφορων τομέων πολιτικής, κυρίως βάση ενός ευρωπαϊκού διαλειτουργικού πλαισίου.
- Η προώθηση της διάδοσης ορθών πρακτικών και της ενθάρρυνσης της ανάπτυξης καινοτομικών τηλεματικών λύσεων στις δημόσιες διοικήσεις.

Στις 23 Μαρτίου 2017, το πρόγραμμα ISA2 [63] δημοσίευσε μια νέα έκδοση του EIF. Αυτή η έκδοση ονομάζεται «νέο EIF» [35] και περιλαμβάνει όλες τις πολιτικές αλλαγές των τελευταίων ετών

Αξιζει να σημειωθεί πως το EIF από μόνο του δεν αποτελεί κάποιο πρότυπο, ωστόσο παρέχει ένα πλαίσιο με οδηγίες και διατάξεις για μια καλή και δομημένη επιλογή των υπολοίπων προτύπων.

2.1.3.6 OASIS (Organization for the Advancement of Structured Information Standards)

Ο OASIS [64] είναι μία μη κερδοσκοπική κοινοπραξία που οδηγεί στην ανάπτυξη, τη σύγκλιση και την υιοθέτηση ανοικτών προτύπων για την παγκόσμια κοινωνία της πληροφορίας. Τα πρότυπα που παράγονται από τον OASIS είναι ανοικτά πρότυπα, πράγμα που σημαίνει ότι μπορεί να χρησιμοποιηθούν από οποιαδήποτε εταιρεία ή οργανισμό και έχουν ως στόχο τη μείωση του κόστους, την τόνωση της καινοτομίας, την αύξηση των παγκόσμιων αγορών, και την προστασία του δικαιώματος της ελεύθερης επιλογής της τεχνολογίας. Αυτό επιτρέπει σε πολλαπλές εταιρείες να αναπτύξουν προϊόντα με βάση το ίδιο πρότυπο, το οποίο προσφέρει υψηλού βαθμού διαλειτουργικότητα μεταξύ των διαφόρων συστημάτων πληροφορικής.



Ο οργανισμός φιλοξενεί δύο από τις πιο ευρέως γνωστές πύλες πληροφορίας σχετικά με την XML και τα πρότυπα υπηρεσιών Web, τη σελίδα Cover Pages [65] και την XML.org [66], ενώ διαθέτει τεχνικές επιτροπές για την ανάπτυξη διαλειτουργικών προτύπων.

OASIS eGovernment Member OASIS eGovernment Member (eGov MS)

Ένας από τους σημαντικότερους τομείς του OASIS, είναι ο OASIS eGovernment Member [67], που αφορά την προτυποποίηση ηλεκτρονικών επιχειρήσεων. Βασικοί στόχοι του τομέα αυτού είναι:

- Η προώθηση της υιοθέτησης και εφαρμογής ανοικτών προτύπων που διευκολύνουν τη διαλειτουργικότητα μέσα και έξω από κυβερνητικές υπηρεσίες και στα ενδιαφερόμενη μέρη.
- Τόνωση του επιμερισμού καλών πρακτικών και εξέταση περιπτώσεων που αφορούν την παροχή υπηρεσιών ηλεκτρονικής διακυβέρνησης, φέρνοντας μαζί όλους τους υπεύθυνους από τη δημόσια διοίκηση, επιχειρήσεις, εμπειρογνομόνων σε ΤΠΕ, ερευνητές, κλπ.
- Προώθηση της διαλειτουργικότητας και της εφαρμογής ανοικτών προτύπων για μετασχηματισμένες κυβερνήσεις.

- Ανάπτυξη οδηγιών ηλεκτρονικής διακυβέρνησης που εστιάζουν σε περιπτώσεις και απαιτήσεις διαφορετικών δικαιοδοσιών δημόσιας διοίκησης σε χώρες ποικίλων επιπέδων ανάπτυξης της οικονομίας και της τεχνολογίας.
- Δημιουργία μίας πλατφόρμας για μία σειρά ενεργειών που εστιάζουν στην τόνωση λύσεων μετασχηματισμού κυβέρνησης στραμμένων προς τους πολίτες.

OASIS Universal Business Language (UBL)

Η γλώσσα UBL παρέχει έναν ευκολότερο τρόπο μετάδοσης εγγράφων μεταξύ του αποστολέα και του παραλήπτη. Τα έγγραφα μεταδίδονται σε ηλεκτρονική μορφή, σε μορφή XML αρχείων που έχουν μια συγκεκριμένη δομή, αναγνωρίσιμη από όλες τις εφαρμογές. Με τον τρόπο αυτό, δεν υπάρχει επιπλέον κόστος για έγγραφα τυπωμένα σε χαρτί, και δεν υπάρχει πλέον η ανάγκη εισαγωγής περισσότερων δεδομένων σε κάθε πληροφοριακό σύστημα.

Η γλώσσα UBL επιτρέπει στις εφαρμογές να επικοινωνήσουν, δίνοντας τους μια κοινή γλώσσα επικοινωνίας. Παρέχει ένα πλαίσιο για διάφορες κοινότητες είτε απλών χρηστών είτε προμηθευτών, ώστε να δηλώνονται συγκεκριμένοι τρόποι κατανόησης και μετάδοσης των επιμέρους πληροφοριών. Η UBL 2.0 συμφωνήθηκε σαν ένα standard της OASIS τον Δεκέμβρη του 2007. Περιλαμβάνει μια σειρά από βιβλιοθήκες με τριάντα ένα (31) έγγραφα, τα οποία μπορούν να κατηγοριοποιηθούν ως εξής:

- **Προ - πώληση:** Αίτηση για προσφορά, προσφορά, αίτηση για κατάλογο, κατάλογος, διαγραφή καταλόγου, ενημέρωση στοιχείου, ενημέρωση των τιμών καταλόγου.
- **Παραγγελία:** Απάντηση παραγγελίας, απλή απάντηση παραγγελίας, ακύρωση παραγγελίας, αλλαγή παραγγελίας.
- **Παράδοση:** Φορτωτική εταιρεία, αποδεικτικό προέλευσης, οδηγίες προώθησης, κατάσταση μεταφοράς, συμβουλές παραλαβής.
- **Τιμολόγηση:** Υπενθύμιση, τιμολόγιο, πιστωτικό σημείωμα Invoice, χρεωστικό σημείωμα, κατάσταση.
- **Πληρωμή:** Συμβουλές εμβάσματος.

Η χρήση της γλώσσας UBL, βελτιώνει τον τρόπο της μετάδοσης των εγγράφων, με πρωταρχικό πλεονέκτημα να είναι πως η χρήση της XML προσφέρει την ανεξαρτησία από την πλατφόρμα και το σύστημα ανάπτυξης του εγγράφου. Στόχος της UBL δεν είναι να επαναπροσδιορίσει την επιχειρησιακή διαδικασία, ενώ η χρήση της δεν περιλαμβάνει την προσαρμογή του τρόπου που αποθηκεύεται η πληροφορία, μιας και η ανταλλαγή μεταξύ των εγγράφων είναι ανεξάρτητη από οποιοδήποτε σύστημα προέρχονται.

Τέλος, όπως το πρώτο πρότυπο που εντάχθηκε από την ebXML, η βιβλιοθήκη της UBL βασίζεται στα στοιχεία που είναι ειδικά για εννοιολογικά μοντέλα, γνωστά και ως οντότητες, οι οποίες εντάσσονται σε συγκεκριμένα έγγραφα όπως την παραγγελία ή το τιμολόγιο. Τα τελικά έγγραφα, μετατρέπονται με τη χρήση συγκεκριμένων κανόνων, που βρίσκονται εντός ενός XML Schema, με αποτέλεσμα να είναι εύκολο να αναγνωριστούν από οποιονδήποτε.

2.1.3.7 IDABC (Interoperable Delivery of European eGovernment Services to Public Administrations, Businesses and Citizens)

Το IDABC [62] είναι ένα πρόγραμμα ηλεκτρονικών διοικητικών υπηρεσιών που θεσπίστηκε για την περίοδο 2005-2009. Σύμφωνα με το πρόγραμμα αυτό, η διαλειτουργικότητα και τα ανοικτά πρότυπα παραμένουν πεδία δράσης με προτεραιότητα, στα οποία προστίθενται οι νέες πανευρωπαϊκές υπηρεσίες που πρόκειται να δημιουργηθούν.



Αποσκοπεί να υποστηρίξει και να προωθήσει την ανάπτυξη των πανευρωπαϊκών υπηρεσιών ηλεκτρονικής δημόσιας διοίκησης, καθώς και τα διαλειτουργικά τηλεματικά δίκτυα που τις υποστηρίζουν. Επιδιώκει επίσης, να δημιουργήσει τη δυνατότητα ανταλλαγής πληροφοριών μεταξύ των δημόσιων διοικήσεων, καθώς και μεταξύ των εν λόγω δημόσιων διοικήσεων και των κοινοτικών οργάνων.

Παράλληλα, έχει ως στόχο να διευκολύνει την παροχή πανευρωπαϊκών υπηρεσιών στις επιχειρήσεις και στους πολίτες, λαμβάνοντας υπόψη τις ανάγκες τους, αλλά και να επιτύχει τη διαλειτουργικότητα μεταξύ των διάφορων τομέων πολιτικής, κυρίως βάσει ενός ευρωπαϊκού διαλειτουργικού πλαισίου.

Περιλαμβάνει σχέδια κοινού ενδιαφέροντος που επιτρέπουν την εφαρμογή της κοινοτικής νομοθεσίας και τη βελτίωση της συνεργασίας, ενώ εμπεριέχει, επίσης, οριζόντια μέτρα που αποσκοπούν στη δημιουργία πανευρωπαϊκών οριζόντιων υπηρεσιών ηλεκτρονικής διακυβέρνησης και υπηρεσιών υποδομής, ιδίως υπηρεσιών που υποστηρίζουν τη διαλειτουργικότητα.

Η εφαρμογή των σχεδίων κοινού ενδιαφέροντος και των οριζόντιων μέτρων πρέπει να τηρεί ορισμένες αρχές, όπως:

- Να στηρίζεται σε μια τομεακή νομική βάση.
- Να συνεπάγεται τη συμμετοχή του μεγαλύτερου δυνατού αριθμού κρατών μελών.
- Να περιλαμβάνει, ενδεχομένως, ένα προκαταρκτικό στάδιο, ένα στάδιο μελέτης σκοπιμότητας, ένα στάδιο ανάπτυξης και επικύρωσης, καθώς και ένα στάδιο εκτέλεσης.
- Να λαμβάνει υπόψη τα άλλα κοινοτικά προγράμματα ώστε να μην επαναλαμβάνονται ταυτόσημες εργασίες. Πρόκειται ιδίως για τα προγράμματα έρευνας και τεχνολογικής ανάπτυξης, όπως τα προγράμματα [68] eTen, eContent, e-Inclusion και eLearning.
- Να λαμβάνει υπόψη το ευρωπαϊκό πλαίσιο διαλειτουργικότητας που παρέχεται από το πρόγραμμα IDABC.
- Να χρησιμοποιεί, στο μέτρο του δυνατού, τις πανευρωπαϊκές οριζόντιες υπηρεσίες ηλεκτρονικής δημόσιας διοίκησης και υποδομής.
- Να πραγματοποιεί μια εξέταση με στόχο την παρακολούθηση των σχεδίων και των μέτρων εντός του έτους που ακολουθεί την ολοκλήρωση του σταδίου εκτέλεσης.

2.1.4 Τεχνικές Διαλειτουργικότητας

Η διαλειτουργικότητα δεν είναι μια νέα έννοια. Προς το τέλος της δεκαετίας του '80 η έλλειψη διαλειτουργικότητας ήταν ένα κοινό πρόβλημα ενός διαχειριστή συστήματος που διαχειριζόταν ένα ευρύ και ετερογενές δίκτυο υπολογιστών, όπου τα τμήματα ενός ίδιου οργανισμού χρησιμοποιούσαν διαφορετικά λειτουργικά συστήματα, εργαλεία λογισμικού και βάσεις δεδομένων. Γενικότερα, η διαλειτουργικότητα εξαρτάται από την αίτηση και τη λήψη υπηρεσιών, οι οποίες μπορεί να θεωρηθούν ως μια ανταλλαγή πόρων. Ένα παράδειγμα ενός πόρου είναι μια περιγραφή. Ειδικότερα, στις περιγραφές περιβαλλοντικών συστημάτων πληροφοριών περιλαμβάνονται: (i) οι περιγραφές των προβλημάτων, (ii) οι περιγραφές του περιβαλλοντικού συστήματος, (iii) οι περιγραφές των μοντέλων ή άλλων εργαλείων, (iv) οι περιγραφές των δεδομένων παρατήρησης, και (v) οι περιγραφές των γενικών γνώσεων. Άλλοι πόροι περιλαμβάνουν (εικονικές) μηχανές ικανές να επεξεργάζονται αυτές τις περιγραφές και άλλα γενικά εργαλεία.

Αναφορικά με τις διαφορετικές τεχνικές διαλειτουργικότητας, η εστίασή τους στην έρευνα διαλειτουργικότητας αφορούσε την ικανότητα των διαφορετικών λειτουργικών συστημάτων να επικοινωνούν σε ένα δικτυωμένο περιβάλλον. Αυτό το τεχνικό πρόβλημα διαλειτουργικότητας επιλύθηκε χάρη στην ευρεία υιοθέτηση κοινών πρωτοκόλλων για δίκτυα επικοινωνιών, όπως το TCP/IP [69], στο οποίο βασίζεται το διαδίκτυο. Ωστόσο, η διαλειτουργικότητα ήταν εκτός άμεσης εμβέλειας, καθώς η απρόσκοπτη ανταλλαγή δεδομένων εξακολουθούσε να εμποδίζεται από συντακτικά προβλήματα, μιας και δεν ήταν δυνατή η πρόσβαση σε ένα έγγραφο που φυλασσόταν σε σταθμό εργασίας βασισμένο σε UNIX, από έναν υπολογιστή που βασιζόταν σε άλλο λειτουργικό σύστημα, εκτός εάν εκτελούταν η κατάλληλη προ-επεξεργασία.

Διαφορετικές ομάδες έρευνας εργάστηκαν για να ξεπεράσουν αυτόν τον περιορισμό, με τα αποτελέσματά τους να μπορούν να βρεθούν στα πακέτα λογισμικού που χρησιμοποιούμε καθημερινά. Μεταξύ των διαφόρων προϊόντων που ξεπέρασαν τη συντακτική διαλειτουργικότητα, αναφέρονται τα Open Database Connectivity (ODBC), Common Object Request Broker Architecture (CORBA) και XML.

- Το **ODBC** [70] είναι μια ευρέως αποδεκτή διεπαφή προγραμματισμού εφαρμογών, που αναπτύχθηκε από τη Microsoft, η οποία επιτρέπει την ομοιόμορφη πρόσβαση σε ένα ευρύ φάσμα συστημάτων βάσεων δεδομένων. Εισάγει ένα στρώμα μεταξύ της εφαρμογής πελάτη και της βάσης

δεδομένων, το οποίο αποκρύπτει τις λεπτομέρειες της βάσης δεδομένων πίσω από ένα γενικό και δημοσιευμένο επίπεδο αφαίρεσης. Το ODBC χρησιμοποιεί το SQL ως τεχνολογία ενεργοποίησης για να επικοινωνεί με την πηγή δεδομένων. Ειδικά προγράμματα οδήγησης έχουν σχεδιαστεί για τα περισσότερα συστήματα βάσεων δεδομένων. Το ODBC API έχει μεταφερθεί σε διάφορα λειτουργικά συστήματα και μπορεί να χρησιμοποιηθεί σε πολλές γλώσσες προγραμματισμού, υποστηρίζοντας έτσι την πραγματική διαλειτουργικότητα.

- Το **CORBA** [71] είναι μια αρχιτεκτονική λογισμικού και μια υποδομή που επιτρέπει σε εφαρμογές υπολογιστών που λειτουργούν σε διαφορετικά μηχανήματα, και σε διαφορετικά λειτουργικά συστήματα, να συνεργάζονται σε ένα δίκτυο. Το CORBA ήταν η πρώτη προσπάθεια να προχωρήσουμε πέρα από την έννοια της «τηλεφωνικής διαδικασίας κλήσης», την οποία εντοπίζουμε για πρώτη φορά στα δίκτυα UNIX. Στο CORBA ο προγραμματιστής περιστρέφει την εφαρμογή του δημοσιεύοντας τη διασύνδεσή του χάρη στην Interface Description Language (IDL) [72], η οποία ορίζεται γενικά αλλά εφαρμόζεται ειδικά για κάθε γλώσσα που υποστηρίζει το CORBA. Μια τέτοια εφαρμογή μπορεί να ανακοινώσει τις υπηρεσίες της σε ένα δίκτυο. Άλλες εφαρμογές που γνωρίζουν για αυτή τη δημοσιευμένη διεπαφή μπορούν συνεπώς να χρησιμοποιούν τις υπηρεσίες που παρέχονται από την απομακρυσμένη εφαρμογή. Ουσιαστικά ο Broker Request Object είναι αυτός που διαχειρίζεται τα αιτήματα που προέρχονται από εφαρμογές πελάτη και τα κατευθύνει στους διακομιστές. Η Microsoft ανέπτυξε τη δική της παραλλαγή του CORBA, πρώτα σε ένα τοπικό περιβάλλον χάρη στην αρχιτεκτονική Component Object Model (COM), στη συνέχεια το διένειμε σε ένα ομοιογενές δίκτυο Distributed COM (DCOM) και τελικά έκανε το βήμα για να ανοίξει την αρχιτεκτονική σε άλλα λειτουργικά συστήματα με την αρχιτεκτονική .Net. Η γλώσσα Java της Sun Microsystems έκανε την ίδια σχεδόν εφαρμογή στην αρχιτεκτονική Java RMI (Remote Method Invocation). Το CORBA και παρόμοιες προσεγγίσεις δεν έχουν φτάσει ακόμα στην ευρεία υιοθέτηση και δημοτικότητα του ODBC. Οι πιο σημαντικοί λόγοι που δεν έχει υιοθετηθεί ευρέως είναι ο υπολογιστικός φόρτος, η δυσκολία δημιουργίας ενός καταναμημένου δικτύου διαφανών εφαρμογών, η βραδύτητα του εφαρμοσμένου κώδικα και ενδεχομένως η έλλειψη πραγματικής ανάγκης για καταναμημένη υπολογιστική.

- Η **XML** αναπτύχθηκε και προωθήθηκε από το W3C, ασχολείται με τη διαλειτουργικότητα σε επίπεδο εγγράφων. Η XML είναι παράγωγο της Structured General Markup Language (SGML) [73]. Η XML εισήχθη ως εργαλείο διαλειτουργικότητας μεταξύ ενός πελάτη διαδικτύου (Web Client), ενός διακομιστή διαδικτύου (Web Server) και βάσεων δεδομένων. Ενώ η CORBA και παρόμοιες προσεγγίσεις επιτρέπουν σε κάποιον να επικαλείται εξ' αποστάσεως μια εφαρμογή, δεν αναφέρεται τίποτα σχετικά με τη μορφή δεδομένων που ζητείται από την απομακρυσμένη εφαρμογή, και ως εκ τούτου ένας τρόπος να προσδιοριστεί αυτό ήταν απαραίτητος. Στο πρόβλημα αυτό, η XML είναι η απάντηση.

Γενικότερα, προκειμένου να επιτευχθεί η διαλειτουργικότητα υπάρχουν τρεις προσεγγίσεις που συνήθως ακολουθούνται, όπως αναφέρονται παρακάτω:

- **Ολοκληρωμένη προσέγγιση (Integrated) [74]:** Στην προσέγγιση αυτή, υπάρχει κοινή μορφή για όλα τα μοντέλα, η οποία πρέπει να είναι τόσο λεπτομερής όσο και τα μοντέλα. Η κοινή μορφή δεν είναι αναγκαστικά ένα πρότυπο, αλλά πρέπει να συμφωνηθεί από όλα τα μέρη ώστε να επεξεργαστούν τα μοντέλα και να δημιουργήσουν συστήματα.
- **Ενοποιημένη προσέγγιση (Unified) [75]:** Στην προσέγγιση αυτή, υπάρχει μια κοινή μορφή αλλά μόνο σε μετα-επίπεδο. Αυτό το μετα-μοντέλο δεν αποτελεί μια εκτελέσιμη οντότητα, όπως είναι στην ολοκληρωμένη προσέγγιση, αλλά παρέχει ένα μέσο σημασιολογικής ισοδυναμίας για να επιτρέπεται η αντιστοίχιση μεταξύ των μοντέλων.
- **Ομοσπονδιακή προσέγγιση (Federated) [76]:** Στην προσέγγιση αυτή, δεν υπάρχει κάποια κοινή μορφή. Για να καθιερωθεί η διαλειτουργικότητα, τα συμβαλλόμενα μέρη πρέπει να φιλοξενηθούν σε πραγματικό χρόνο. Η χρήση ομοσπονδιακής προσέγγισης συνεπάγεται ότι κανένας εταίρος δεν επιβάλλει τα μοντέλα, τις γλώσσες και τις μεθόδους εργασίας, κάτι που υποδηλώνει τον διαμοιρασμό μίας κοινής οντολογίας.

Σύμφωνα με τις προαναφερθείσες τεχνικές και προσεγγίσεις, πολλαπλές έρευνες προς την επίτευξη διαλειτουργικότητας δεδομένων σε διαφορετικούς τομείς έχουν πραγματοποιηθεί. Γενικότερα, μία υποδομή για την επίτευξη διαλειτουργικότητας πρέπει να αποτελείται από τέσσερα επίπεδα:

- **1^ο Επίπεδο:** Στο χαμηλότερο στρώμα, να υπάρχει ένα κανάλι επικοινωνίας που διαμοιράζεται μεταξύ των ενεργών συστημάτων πληροφοριών.

- **2^ο Επίπεδο:** Στην κορυφή του καναλιού επικοινωνίας, να καθορίζονται υπηρεσίες που παρέχουν βασική λειτουργικότητα ανταλλαγής δεδομένων, όπως η ασφαλής ή η αξιόπιστη μετάδοση δεδομένων. Αυτά τα δύο επίπεδα αποτελούν τη βάση για την ανταλλαγή δεδομένων μεταξύ των συστημάτων πληροφοριών.
- **3^ο Επίπεδο:** Για να είναι πραγματικά κατανοητά τα συστήματα πληροφοριών, πρέπει να δημιουργηθεί μια κοινή σύνταξη στην οποία θα πρέπει να συμμορφώνονται όλα τα μηνύματα που μεταφέρονται μεταξύ τους. Αυτό αναφέρεται ως διαλειτουργικότητα σύνταξης που μπορεί να θεωρηθεί ως ένα τρίτο στρώμα που βρίσκεται πάνω από το στρώμα υπηρεσιών.
- **4^ο Επίπεδο:** Σε τελευταίο επίπεδο, πρέπει να υπάρχει ένα κοινό σημασιολογικό πλαίσιο το οποίο είναι απαραίτητο για την ερμηνεία κάθε μηνύματος με τη σωστή έννοια.

Εκτός από την προαναφερθείσα υποδομή, πρέπει να υπάρχει και ένα καθιερωμένο κοινό λεξιλόγιο και μια ενοποιημένη αναπαράσταση, τα οποία είναι απαραίτητα στοιχεία για τη συντακτική διαλειτουργικότητα. Για παράδειγμα, ένα μέρος στην επικοινωνία μπορεί να κατανοεί την αγγλική γλώσσα ενώ ένα άλλο μέρος τη γαλλική γλώσσα. Επίσης, μπορεί να κωδικοποιεί τα μηνύματα σε απλό κείμενο, ενώ άλλα μέρη να χρησιμοποιούν κωδικοποίηση XML. Η σημασιολογική διαλειτουργικότητα, από την άλλη πλευρά, χρειάζεται πιο λεπτομερή προσοχή, δεδομένου ότι από την φύση της είναι πιο δύσκολο να επιτευχθεί. Δεν είναι συνήθως αρκετό να χρησιμοποιούνται οι ίδιες λέξεις, αλλά και να ερμηνεύονται οι λέξεις με τη σωστή έννοια. Για παράδειγμα, ένας ιατρός που εργάζεται στην κλινική του μπορεί να χρειαστεί να έχει πρόσβαση στα αποτελέσματα των μετρήσεων του ασθενούς, στη βάση δεδομένων του εργαστηρίου. Από την άλλη πλευρά, μπορεί επίσης να ενδιαφέρεται να συγκρίνει ένα έγγραφο με ένα κανονικό τεστ ελέγχου που είναι αποθηκευμένο στη βάση δεδομένων του νοσοκομείου. Αν και αυτά τα δύο στοιχεία μπορούν να αναφέρονται και ως εργαστηριακές αναφορές που έχουν τα ίδια στοιχεία δεδομένων και την ίδια δομή, αντιπροσωπεύουν εντελώς διαφορετικά πράγματα και το σύστημα θα πρέπει να είναι σε θέση να διακρίνει και τα δύο.

2.2 Διαλειτουργικότητα στον τομέα της Υγείας

2.2.1 Ανάγκη Διαλειτουργικότητας στον τομέα της Υγείας

Σύμφωνα με το HIMSS [77], «η διαλειτουργικότητα περιγράφει την έκταση στην οποία τα συστήματα και οι συσκευές μπορούν να ανταλλάξουν δεδομένα και να ερμηνεύσουν αυτά τα δεδομένα. Για να είναι διαλειτουργικά δύο συστήματα, πρέπει να είναι σε θέση να ανταλλάσσουν δεδομένα και στη συνέχεια να παρουσιάζουν τα δεδομένα αυτά ώστε να μπορούν να γίνουν κατανοητά από έναν χρήστη». Όσον αφορά τη βιομηχανία της υγειονομικής περίθαλψης, η διαλειτουργικότητα έχει να κάνει με την ικανότητα διαφορετικών συστημάτων πληροφορικής και λογισμικού να ανταλλάσσουν και να μοιράζονται δεδομένα από μια σειρά ζωτικών πηγών, συμπεριλαμβανομένων των εργαστηρίων, των κλινικών, των φαρμακείων, των νοσοκομείων και των ιατρικών πρακτικών. Είναι χρήσιμο να θεωρήσουμε τη διαλειτουργικότητα ως μία φιλοσοφία αντί για μια «αλληλεπίδραση βασισμένη στα πρότυπα μεταξύ των συστημάτων ηλεκτρονικών υπολογιστών» [77]. Από την τεχνική πλευρά των πραγμάτων, η διαλειτουργικότητα στον τομέα της υγείας οδηγεί σε αυξημένη δέσμευση των ασθενών και καλύτερα αποτελέσματα, καθώς περισσότερος χρόνος μπορεί να δαπανηθεί για τις λεπτομέρειες της ασθένειας ή του τραυματισμού, καθώς οι επαγγελματίες του ιατρικού τομέα είναι σε θέση να σχεδιάζουν ένα προσωποποιημένο σχέδιο θεραπείας. Η διαλειτουργικότητα στην υγειονομική περίθαλψη φέρει ως αποτέλεσμα την ευκολότερη πρόσβαση σε πράγματα όπως οι λεπτομέρειες στα ηλεκτρονικά αρχεία υγείας των ασθενών, αλλά λειτουργεί μόνο στο βαθμό που οι πωλητές ιατρικού λογισμικού θα συμφωνήσουν να μοιραστούν τις συχνά ιδιόκτητες πληροφορίες ενός συστήματος, σημειώνει μια έκθεση από την SearchHealthIT [78]. Επιπλέον, οι ενδιαφερόμενοι και οι οργανισμοί υγειονομικής περίθαλψης πρέπει να είναι πρόθυμοι να μοιράζονται τις λεπτομέρειες των ασθενών μέσω δικτύων για πρόσβαση από διαφορετικές βάσεις δεδομένων.

Είναι απαραίτητο να αναγνωριστεί η πολυπλοκότητα των διαδικασιών που περιβάλλουν την υγειονομική περίθαλψη όταν επιδιώκεται η διαλειτουργικότητα των πληροφοριών, στις οποίες εμπλέκονται άμεσα ή έμμεσα διάφορες κατηγορίες ενδιαφερομένων:

- **Εταιρείες υγειονομικής περίθαλψης:** Περιλαμβάνει πωλητές κλινικών συστημάτων, διοικητικά συστήματα πληροφορικής και ιατρικές συσκευές. Αυτή η κατηγορία περιλαμβάνει συνήθως μεγάλες πολυεθνικές εταιρείες με παγκόσμιους στόχους και εταιρείες μικρού/μεσαίου μεγέθους με επίκεντρο μια αγορά σε μικρή γεωγραφική περιοχή ή στενό πεδίο εφαρμογής.

- **Πάροχοι υπηρεσιών υγείας:** Περιλαμβάνει επαγγελματίες του τομέα της υγείας, που ανήκουν σε ποικίλα επαγγέλματα π.χ. ιατροί, νοσηλευτές, τεχνικοί. Δυστυχώς, πολύ λίγοι έχουν όλες τις απαραίτητες και κατάλληλες δεξιότητες στην επιστήμη και την τεχνολογία υπολογιστών και τον απαραίτητο χρόνο για τη διάθεση πόρων για την αντιμετώπιση προβλημάτων διαλειτουργικότητας ή για την ενεργό συμμετοχή στην τυποποίηση της πληροφορικής και των προτύπων στον τομέα της υγείας.
- **Το προσωπικό πληροφορικής και διοικητικών στελεχών:** Και οι δύο διαδραματίζουν βασικούς ρόλους σε μεγάλα ιδρύματα παροχής υγειονομικής περίθαλψης, αλλά συχνά απουσιάζουν σε μικρές εγκαταστάσεις υγειονομικής περίθαλψης. Επιπλέον, πολύ λίγοι δραστηριοποιούνται σε οργανισμούς ανάπτυξης προτύπων ή σε τεχνολογική ολοκλήρωση.
- **Υγειονομικές αρχές και κυβερνήσεις:** Περιλαμβάνει τους σχετικούς φορείς που εποπτεύουν ή διαχειρίζονται το σύστημα υγείας σε εθνικό επίπεδο. Με τις εξελισσόμενες τεχνολογίες και πρότυπα, συνήθως, οι κυβερνήσεις και οι αρχές επιβάλλουν τη διαλειτουργικότητα μέσω εθνικών προτύπων και, στις περισσότερες περιπτώσεις, έχει αποδειχθεί αποτελεσματική.

Για να γίνει κατανοητή η σημασία της, είναι σημαντικό να κατανοήσουμε το σημερινό τοπίο. Σήμερα, οι πάροχοι αντιμετωπίζουν μια καταπληκτική λίστα προκλήσεων [79]: περιορισμοί στην κοινή χρήση δεδομένων, έλλειψη τυποποιημένων μορφών δεδομένων σε διαφορετικά συστήματα, συστήματα παλαιού τύπου που αποθηκεύουν σημαντικά δεδομένα, δαπανηρά τέλη διασύνδεσης, περικοπές τελών εργαστηρίου, μειωμένες επιστροφές και περιορισμούς χρόνου και πόρων. Συνδυάζοντας τα στοιχεία αυτά με την ανάγκη να βελτιωθεί η φροντίδα με βάση την αξία, η σημασία επίλυσης αυτών των προκλήσεων καθίσταται όλο και πιο αναγκαία. Για παράδειγμα, τα εργαστήρια και τα συστήματα υγείας πραγματοποιούν σημαντικές οικονομικές και επιχειρησιακές επενδύσεις σε κινητήρες διασύνδεσης και αναπτυξιακούς πόρους, ώστε το σύστημά τους να μπορεί να επικοινωνεί με εργαστήρια. Τα συστήματα υγείας έχουν ιδιαίτερα περίπλοκες απαιτήσεις, οι οποίες καθυστερούν περαιτέρω την υλοποίηση των διεπαφών τους και αυξάνουν τα έξοδα, προσλαμβάνοντας δαπανηρό προσωπικό για την ολοκλήρωση των έργων. Κάθε φορά, σχηματίζουν μια ομάδα για να αναπτύξουν αυτές τις διεπαφές και να εκτελέσουν αυτές τις υλοποιήσεις, ξοδεύοντας ένα τεράστιο χρονικό διάστημα και πόρους «επανεφεύροντας τον τροχό», παρόλο που αυτές οι διεπαφές έχουν ήδη αναπτυχθεί από κάποιον άλλο κάπου αλλού.

Τα οφέλη ωστόσο της διαλειτουργικότητας στον τομέα της υγείας είναι πολλά [80]. Ορισμένα από αυτά αναφέρονται παρακάτω:

- **Βελτιωμένη απόδοση:** Η διαλειτουργικότητα στην υγειονομική περίθαλψη έχει σχεδιαστεί για να αυξήσει την αποτελεσματικότητα. Όταν τα δεδομένα παρουσιάζονται σε σταθερή βάση ανεξάρτητα από την πηγή, είναι πιο εύκολο για τους επαγγελματίες να φτάσουν γρήγορα στην καρδιά του θέματος καθώς λαμβάνουν αποφάσεις σχετικά με τη θεραπεία.
- **Ασφαλέστερες μεταβάσεις φροντίδας:** Η συνεχής φροντίδα είναι ζωτικής σημασίας για τους ασθενείς, είτε για χρόνιες παθήσεις είτε για την αντιμετώπιση επικίνδυνων καταστάσεων μεταξύ πολλαπλών παρόχων υπηρεσιών υγείας. Η διαλειτουργικότητα επιτρέπει ασφαλέστερες μεταβάσεις της περίθαλψης, γεγονός που οδηγεί σε καλύτερα αποτελέσματα για όλους τους ασθενείς. Για παράδειγμα, ένας ασθενής που είναι σε διακοπές και αρρωσταίνει μπορεί να μην μπορεί να παράσχει όλες τις λεπτομέρειες του ιατρικού ιστορικού του, γεγονός που μπορεί να αποβεί μοιραίο για τον ίδιο τον ασθενή και για τον εκάστοτε ιατρό που τον έχει αναλάβει. Ένα πρόσφατο άρθρο στο *Medical Economics* [81] υπογράμμισε μια περίπτωση όπου ένας ασθενής μπορούσε μόνο να περιγράψει ότι είχε υποστεί «κάποια επέμβαση» χωρίς να γνωρίζει άλλες λεπτομέρειες. Στην περίπτωση αυτή, ο ιατρός θα μπορούσε να αντιμετωπίσει το πρόβλημα υγείας του ασθενούς γρηγορότερα και πιο ολοκληρωμένα, εάν τα δεδομένα από το EHR του ασθενούς στην πατρίδα του ήταν προσβάσιμα από τον ιατρό.
- **Μείωση κόστους:** Η διαλειτουργικότητα σημαίνει ότι μπορούν να κοινοποιηθούν εγκαίρως πιο χρήσιμες πληροφορίες. Έτσι, για παράδειγμα τα δεδομένα από έναν ασθενή που έκανε εξέταση αίματος την περασμένη εβδομάδα, μπορούν να χρησιμοποιηθούν σε έναν άλλο χώρο, εξοικονομώντας έτσι χρόνο και κόστος από την πραγματοποίηση νέων και περισσότερων μη απαραίτητων εξετάσεων στο νοσοκομείο. Σαν αποτέλεσμα, η βελτιωμένη αποτελεσματικότητα μέσω της μεγαλύτερης ανταλλαγής πληροφοριών, εξοικονομεί χρόνο και προσπάθεια για τους ενδιαφερομένους, γεγονός που οδηγεί σε μεγαλύτερη εξοικονόμηση κόστους. Η ανάγκη μείωσης του κόστους, όπου αυτό είναι δυνατό, είναι πρωταρχικής σημασίας για τη μελλοντική κερδοφορία και επιτυχία για τα συστήματα υγείας. Μια σημαντική ευκαιρία για μείωση του κόστους είναι μέσω της σωστής στρατηγικής διασύνδεσης. Οι σύγχρονοι μηχανισμοί διασύνδεσης υγειονομικής περίθαλψης που λειτουργούν ως κεντρικός

κόμβος όλων των κλινικών συναλλαγών διατίθενται σήμερα για τη διαλειτουργικότητα σε επίπεδο επιχειρήσεων. Τέτοιοι κινητήρες επιτρέπουν στα νοσοκομεία, τα εργαστήρια, τα ακτινολογικά κέντρα και τις κλινικές να σταματήσουν να ανακαλύπτουν τον ίδιο τροχό και να αρχίζουν να επαναχρησιμοποιούν ήδη αναπτυγμένες διεπαφές. Αντί να δημιουργούν και να υλοποιούν διασυνδέσεις, εστιάζουν στην εύκολη διαμόρφωση από τις αναπτυσσόμενες λίστες, επιτρέποντας τις ταχύτερες υλοποιήσεις - μερικές φορές μέσα σε λίγα λεπτά. Επιπλέον, μειώνουν σημαντικά το κόστος της πρόσληψης προγραμματιστών διεπαφών με υψηλούς μισθούς και δαπανηρές αμοιβές μηχανών διασύνδεσης.

Εμπόδια Διαλειτουργικότητας

Τρεις κατηγορίες εμποδίων (εννοιολογικών, τεχνολογικών και οργανωτικών) προσδιορίζονται κατά την επίτευξη της διαλειτουργικότητας στον τομέα της Υγείας [82]:

- **Εννοιολογικά εμπόδια:** Τα εμπόδια αυτά αφορούν τις συντακτικές και σημασιολογικές διαφορές των πληροφοριών που πρέπει να ανταλλάσσονται. Αυτά τα εμπόδια αφορούν τη μοντελοποίηση σε υψηλό επίπεδο αφαίρεσης (όπως για παράδειγμα τα επιχειρηματικά μοντέλα μιας επιχείρησης) καθώς και το επίπεδο του προγραμματισμού (για παράδειγμα μοντέλα XML).
- **Τεχνολογικά εμπόδια:** Τα εμπόδια αυτά αναφέρονται στην ασυμβατότητα των τεχνολογιών της πληροφορίας (αρχιτεκτονική και πλατφόρμες, υποδομή). Αυτά τα προβλήματα αφορούν τα πρότυπα για την παρουσίαση, αποθήκευση, ανταλλαγή, επεξεργασία και κοινοποίηση των δεδομένων μέσω της χρήσης υπολογιστών.
- **Οργανωτικά εμπόδια:** Τα εμπόδια αυτά σχετίζονται με τον ορισμό της ευθύνης (ποιος είναι υπεύθυνος για το τι) και την εξουσία (ποιος είναι εξουσιοδοτημένος να κάνει τι), καθώς και την ασυμβατότητα των οργανωτικών δομών.

Ανεξάρτητα από τις κατηγορίες αυτές, οι βασικές δυσκολίες και εμπόδια διαφαίνονται στα ακόλουθα [83]:

- **Ανεπαρκή πρότυπα δεδομένων:** Ενώ υπάρχουν πρότυπα για την ηλεκτρονική ανταλλαγή πληροφοριών μεταξύ των συστημάτων EHR, τα ενδιαφερόμενα μέρη δηλώνουν ότι τα πρότυπα αυτά δεν επαρκούν για την επίτευξη της διαλειτουργικότητας των EHR. Η επίτευξη ευρείας υιοθέτησης των προτύπων δεδομένων είναι το θεμέλιο της πραγματικής διαλειτουργικότητας των δεδομένων για την υγεία. Ωστόσο, ορισμένοι προμηθευτές και οργανισμοί υγείας εξακολουθούν να λειτουργούν κάτω από ξεπερασμένες ιδέες σχετικά με την ιδιοκτησία δεδομένων και τα οφέλη της τοπικής διατήρησης πληροφοριών ασθενών. Άλλοι είναι πρόθυμοι να μοιραστούν, αλλά απλά δεν έχουν τη δυνατότητα να το κάνουν. Πολλοί ειδικοί έχουν θέσει το HL7 FHIR ως το πρωτόκολλο που θα επιτρέψει στους πωλητές να αγκαλιάσουν την απρόσκοπτη ανταλλαγή δεδομένων στο εγγύς μέλλον. Άλλες ομάδες, όπως το DirectTrust [84], προσπαθούν να αντιμετωπίσουν τα τρέχοντα εμπόδια, επιτρέποντας γρήγορη και απλή ανταλλαγή πληροφοριών μέσω ασφαλών μηνυμάτων. Ανεξάρτητα από τη μέθοδο που επιλέγουν, οι προμηθευτές και οι ιδιωτικές εταιρικές σχέσεις κάνουν το καλύτερο για να απομακρυνθούν από τα πρότυπα δεδομένων και ασχολούνται με τη δημιουργία ενός διαλειτουργικού οικοσυστήματος προς όφελος των παρόχων και των ασθενών. Κάτι τέτοιο θα χρειαστεί χρόνο, καθώς και υπομονή, ώστε αυτές οι προσπάθειες να ωριμάσουν.
- **Ανώριμότητα EHR:** Γενικότερα, ένας τρόπος με τον οποίο τα συστήματα υγείας μπορούν να επιτύχουν πιο διαλειτουργικές ροές εργασίας είναι να χρησιμοποιήσουν ένα σύστημα παραπομπής. Τα συστήματα παραπομπής εξομαλύνουν τη ροή των παραπομπών των ασθενών μεταξύ των παρόχων υγειονομικής περίθαλψης, επιτρέποντας στενές σχέσεις μεταξύ όλων των παρόχων, ενώ εξορθολογούν και συντονίζουν τη φροντίδα του ασθενούς. Αυτά τα συστήματα παραπομπής εντάσσονται στα EHR, διευκολύνοντας τη διαλειτουργικότητα για την ανταλλαγή κλινικών δεδομένων. Ωστόσο, οι πάροχοι εξακολουθούν να αγωνίζονται με τη βελτιστοποίηση του EHR, τη διαχείριση της υγείας του πληθυσμού και την επιστροφή των εξόδων. Τα EHR είναι ανώριμα, ενώ οι πάροχοι περιορίζονται από τις απαιτήσεις υποβολής εκθέσεων ποιότητας, τις πρωτοβουλίες βελτίωσης των εννοιών, τους συναγερμούς και τις ειδοποιήσεις που διαταράσσουν τις ροές εργασίας και τους κανόνες ακεραιότητας δεδομένων που τους εμποδίζουν. Επιπλέον, τα δεδομένα είναι συχνά ελλιπή. Τόσο οι ασθενείς όσο και οι κλινικοί ιατροί διαμαρτύρονται για την αναντιστοιχία των φαρμάκων, τα δυσπρόσιτα αποτελέσματα των εξετάσεων και τις ενοχλητικά διοικητικές διαδικασίες που εμποδίζουν την παροχή προσωπικής, προσεκτικής και ουσιαστικής φροντίδας.

- **Διακυμάνσεις των κανόνων περί απορρήτου του κράτους:** Οι νομοθέτες θα μπορούσαν να βοηθήσουν τη διαδικασία παράλληλα με την εναρμόνιση των κανόνων περί προστασίας της ιδιωτικής ζωής σε ολόκληρη την πολιτεία. Οι αντιπρόσωποι και οι εκπρόσωποι δηλώνουν ότι η ανταλλαγή πληροφοριών για την υγεία με παρόχους σε άλλα κράτη, η οποία είναι απαραίτητη για τη διαλειτουργικότητα του EHR σε εθνικό επίπεδο, μπορεί να είναι δύσκολη. Σύμφωνα με έρευνες, οι πάροχοι ενδέχεται να διστάζουν να ανταλλάξουν πληροφορίες για την υγεία με άλλους παρόχους, στην περίπτωση που οι πάροχοι δεν έχουν τη διαβεβαίωση ότι οι ασθενείς έχουν συγκατατεθεί ρητά στην ανταλλαγή. Ορισμένες από τις σημαντικότερες πληροφορίες για ανταλλαγή είναι επίσης ευαίσθητες. Πολλά κράτη απαιτούν πρόσθετη συγκατάθεση για τους παρόχους ώστε να μοιράζονται δεδομένα ή πληροφορίες υγείας, ενώ τα ηλεκτρονικά εργαλεία δεν είναι αρκετά περίπλοκα για να διαχειριστούν αυτά τα πολύπλοκα καθήκοντα. Μια έρευνα δείχνει πως στην ανταλλαγή δεδομένων δεν περιλαμβάνονται συγκεκριμένα πληροφορίες για την υγεία, ακόμη και με τη συγκατάθεση του ασθενούς, λόγω ανησυχιών για την παραβίαση των κανόνων απορρήτου. Ενώ η προσοχή μπορεί να είναι αξιολάμπαστη, καθιστά δύσκολο για τους παρόχους και τους ασθενείς να επιτύχουν τα επιδιωκόμενα οφέλη της ανταλλαγής πληροφοριών για την υγεία.
- **Ανεπαρκείς δυνατότητες αντιστοίχισης ασθενών:** Η συγκατάθεση, η ανταλλαγή δεδομένων και η ακρίβεια περιπλέκονται περαιτέρω από την αναξιοπιστία των τρεχουσών τεχνολογιών αντιστοίχισης ασθενών. Ακόμα κι αν οι πάροχοι θα μπορούσαν να εξομαλύνουν τις προκλήσεις της αποστολής ευαίσθητων πληροφοριών σε κρατικές γραμμές, δεν μπορούν ακόμα να είναι σίγουροι ότι τα δεδομένα θα αποδοθούν στον κατάλληλο ασθενή από την άλλη άκρη. Τα διαφορετικά συστήματα EHR μπορούν να δημιουργούν και να αποθηκεύουν δημογραφικές πληροφορίες χρησιμοποιώντας διαφορετικά πρότυπα, καθιστώντας αδύνατο για τους παρόχους να ταξινομήσουν σωστά τα EHR χωρίς να προσθέτουν ένα άλλο επίπεδο τεχνολογίας που δημιουργήθηκε ειδικά για την επίλυση του προβλήματος. Τα συστήματα υγείας μπορούν επίσης να χρησιμοποιούν διάφορα κριτήρια ως βάση για την αντιστοίχιση των ασθενών. Ενώ μερικοί βασίζονται σε διευθύνσεις και ονόματα, άλλοι εκχωρούν στους ασθενείς έναν εσωτερικό αριθμό υποθέσεων που λειτουργεί μόνο στο δικό τους σύστημα. Σύμφωνα με μία μελέτη [85], μια πρόσφατη προσπάθεια για επίτευξη διαλειτουργικότητας με έναν άλλο οργανισμό καθυστέρησε επειδή ο άλλος οργανισμός βασιζόταν στον αριθμό κοινωνικής ασφάλισης ενός ασθενούς για αντιστοίχιση ασθενών. Οι προσπάθειες για τη βελτίωση της

αντιστοίχισης των ασθενών με την ανάπτυξη ενός εθνικού αναγνωριστικού ασθενούς έχουν παραμείνει αδρανείς για χρόνια, αφήνοντας ομάδες ενδιαφερομένων όπως να βρουν τεχνολογίες αντιμετώπισης προβλημάτων.

- **Οικονομικές ανησυχίες και Κόστος επενδύσεων:** Ίσως ακόμη πιο δύσκολο από τα τεχνικά ζητήματα που ανακύπτουν στην αναμόρφωση της υποδομής πληροφορικής για την υγεία, είναι η εξεύρεση χρημάτων για την υποστήριξη αυτών των έργων. Οι οργανισμοί υγειονομικής περίθαλψης που έχουν ήδη επενδύσει χιλιάδες ή εκατομμύρια δολάρια σε συστήματα EHR που δεν πληρούν τις απαιτήσεις διαλειτουργικότητας των δεδομένων για την υγεία, είναι προφανώς απρόθυμοι να επενδύσουν επιπλέον χρήματα. Οι ενδιαφερόμενοι και εκπρόσωποι δηλώνουν ότι το κόστος που συνδέεται με την επίτευξη της διαλειτουργικότητας μπορεί να είναι απαγορευτικό για τους παρόχους, εν μέρει λόγω των δομών τιμολόγησης που αναπτύσσονται από τους πωλητές EHR. Παράλληλα, πολλά συστήματα EHR απαιτούν πολλαπλές προσαρμοσμένες διεπαφές - οι οποίες είναι ειδικά σχεδιασμένες συνδέσεις με άλλα συστήματα πληροφορικής για την υγεία - προκειμένου να διευκολυνθεί η διαλειτουργικότητα με άλλους παρόχους και οργανισμούς. Τα έξοδα ωστόσο που σχετίζονται με αυτές τις προσαρμοσμένες διεπαφές πληρώνονται συνήθως από τους αγοραστές EHR. Οι πάροχοι ενδέχεται επίσης να χρειαστεί να πληρώσουν νομικά τέλη στους δικηγόρους που διαπραγματεύονται τις συμβάσεις ανταλλαγής δεδομένων μεταξύ των συστημάτων υγείας. Οι οργανισμοί διαλειτουργικότητας προσπαθούν να μειώσουν την ανάγκη για επενδύσεις στην προσαρμογή του EHR απλοποιώντας τις βασικές τεχνολογίες που συνδέουν τα συστήματα πληροφορικής υγείας. Οι κρατικοί ή περιφερειακοί οργανισμοί ανταλλαγής πληροφοριών για την υγεία συχνά αναλαμβάνουν μεγάλο μέρος της βαριάς τεχνικής ανύψωσης για τα μέλη τους με αντάλλαγμα ετήσιας αμοιβής, και μπορεί να συμβάλουν στην κάλυψη μέρους του κόστους γύρω από την επανεξέταση του τροχού για κάθε νέο μέλος της ομάδας. Άλλες ομάδες υποστηρίζουν την τεχνολογία API ως έναν φθηνό και αποτελεσματικό τρόπο για να παρακάμψουν το πρόβλημα προσαρμογής. Τα API επιτρέπουν στους παρόχους να επωφεληθούν από τη νοοτροπία αυτόματης αναπαραγωγής (plug-and-play) που υπήρξε τόσο επιτυχημένη στον κόσμο της κινητής τεχνολογίας, και αποτελούν μια πολλά υποσχόμενη οδό για τη βελτίωση της διαλειτουργικότητας των δεδομένων υγείας.

- **Ανάγκη διακυβέρνησης πληροφόρησης, εμπιστοσύνης και συνεργασίας:** Καμία από αυτές τις τεχνικές εξελίξεις δεν προσφέρει οφέλη για τους ασθενείς ή τους παρόχους, αν οι οργανισμοί εξακολουθούν να διστάζουν να εμπιστεύονται ο ένας τον άλλο. Οι βιομηχανικές ομάδες πρέπει να διαβεβαιώσουν τους οργανισμούς υγειονομικής περίθαλψης ότι οι επενδύσεις αυτές αξίζουν, μέσω της καθιέρωση σαφών, ολοκληρωμένων και αμοιβαία επωφελών πλαισίων για την ανταλλαγή πληροφοριών για την υγεία. Αυτές οι πρακτικές διακυβέρνησης μπορούν να περιλαμβάνουν οργανωτικές πολιτικές που σχετίζονται με την προστασία της ιδιωτικής ζωής, την ασφάλεια των πληροφοριών, τη χρήση δεδομένων, τα τεχνικά πρότυπα και άλλα θέματα που επηρεάζουν την ανταλλαγή πληροφοριών μεταξύ των οργανωτικών ορίων. Η οικοδόμηση εμπιστοσύνης σε μια κοινότητα φροντίδας απαιτεί από τους παρόχους να υιοθετήσουν ένα όραμα που βασίζεται στην αξία και επικεντρώνεται στον ασθενή, για το ρόλο τους στο οικοσύστημα της υγειονομικής περίθαλψης. Τα μεγάλα δεδομένα μπορεί να είναι ένα στρατηγικό πλεονέκτημα για μεμονωμένους οργανισμούς, αλλά γίνονται πραγματικά ισχυροί μόνο όταν οι ασθενείς έχουν πρόσβαση σε όλες τις πληροφορίες υγείας χωρίς περιορισμούς.

Καθώς η διαχείριση της υγείας του πληθυσμού και η εξατομικευμένη περίθαλψη καθίστανται πιο αναπόσπαστες από τις μεταβαλλόμενες δομές αποζημίωσης, οι πάροχοι αρχίζουν να αναγνωρίζουν την ανάγκη για τη διαλειτουργικότητα των δεδομένων για την υγεία και την ανταλλαγή πληροφοριών. Όπως και η εξέλιξη των τεχνικών προτύπων, αυτή η εξέλιξη είναι αργή και δύσκολη. Όμως, τα στοιχεία για τη διαλειτουργικότητα συνεχίζουν να αυξάνονται και οι πάροχοι μετατοπίζουν την προοπτική τους αντίστοιχα. Το επιχείρημα για τη διαλειτουργικότητα των δεδομένων για την υγεία καθίσταται όλο και πιο επιτακτικό καθώς οι ιδιωτικές βιομηχανίες και οι ομοσπονδιακοί οργανισμοί συνεχίζουν τις εργασίες τους για να φέρουν τα πρότυπα δεδομένων, τη διαχείριση πληροφοριών και την ανταλλαγή πληροφοριών για την υγεία σε παρόχους που αποδέχονται ότι η συνεργασία αποτελεί το κλειδί για την επιτυχία στο μέλλον. Το βασικό εμπόδιο δεν είναι τα συστήματα αυτά καθ' αυτά, αλλά η δυσκολία ανταλλαγής δεδομένων μεταξύ συστημάτων ή η ασυμβατότητα των δεδομένων. Οι αριθμοί είναι απλοί αλλά αποθαρρυντικοί: εάν κάθε σύνδεση συστήματος σε σύστημα απαιτεί ξεχωριστή, μη τυπική διασύνδεση, ο αριθμός των διεπαφών που απαιτούνται για τη σύνδεση n συστημάτων είναι περίπου $n^2/2$. Έτσι, η πλήρης συνδεσιμότητα 20 συστημάτων πληροφοριών απαιτεί περίπου 200 διεπαφές, επομένως για 40 πληροφοριακά συστήματα, ο αριθμός αυξάνεται σε περίπου 800.

2.2.2 Κατηγορίες Διαλειτουργικότητας στον τομέα της Υγείας

Η εκτεταμένη κατανομή και ανταλλαγή πληροφοριών απαιτεί την επίτευξη τουλάχιστον δύο επιπέδων διαλειτουργικότητας, αναφερόμενοι στη Συντακτική Διαλειτουργικότητα (ικανότητα δύο ή περισσότερων συστημάτων να ανταλλάσσουν πληροφορίες μέσω καθορισμένων δομών μηνυμάτων) και στη Σημασιολογική Διαλειτουργικότητα (ικανότητα των πληροφοριών που μοιράζονται από τα συστήματα να γίνονται κατανοητές στο επίπεδο των εννοιών του ορισμού του αντικειμενικού τομέα). Η Σημασιολογική Διαλειτουργικότητα είναι απαραίτητη για την αυτόματη επεξεργασία πληροφοριών, η οποία θα επιτρέψει την εφαρμογή προηγμένων κλινικών εφαρμογών. Στην πραγματικότητα, η υγειονομική περίθαλψη ασχολείται κυρίως με τη διαχείριση πληροφοριών και γνώσεων και από την άποψη της διάδοσης των κλινικών πληροφοριών, ο στόχος εδώ δεν είναι μόνο η ανταλλαγή δεδομένων και πληροφοριών, αλλά η έξυπνη επαναχρησιμοποίηση και επεξεργασία αυτών των δεδομένων. Σαν αποτέλεσμα αυτών, η διαλειτουργικότητα στον τομέα της υγείας μπορεί να επικεντρωθεί στους δύο ακόλουθους κύριους άξονες:

Άξονας 1: Διαλειτουργικότητα ανταλλαγής μηνυμάτων υγείας

Η χρήση κατάλληλων διεπαφών (interface engines), βοηθά στην ανταλλαγή πληροφοριών μεταξύ ετερογενών πληροφοριακών συστημάτων. Οι διεπαφές αυτές ουσιαστικά συλλέγουν τα δεδομένα από την κύρια εφαρμογή, τα κωδικοποιούν σε μηνύματα και τα μεταδίδουν πάνω από το χρησιμοποιούμενο δίκτυο σε μια άλλη εφαρμογή. Στην πλευρά του λήπτη, το ληφθέν μήνυμα αποκωδικοποιείται και επεξεργάζεται κατάλληλα και τα δεδομένα που περιέχει προωθούνται στην κύρια εφαρμογή του λήπτη προκειμένου να αποθηκευτούν και να επεξεργαστούν ανάλογα με τις εκάστοτε ανάγκες. Σε έναν τέτοιο σχεδιασμό, ο αριθμός των διεπαφών που χρησιμοποιούνται έχει να κάνει με το πλήθος των εφαρμογών που πρόκειται να ανταλλάξουν μηνύματα. Για να αντιμετωπιστεί αυτό το πρόβλημα, προτιμάται να χρησιμοποιείται η κατάλληλη κωδικοποίηση των μηνυμάτων, ώστε κάθε εφαρμογή που γνωρίζει την κωδικοποίηση αυτή να μπορεί να ανταλλάσσει μηνύματα χρησιμοποιώντας την ίδια διεπαφή. Ένα από τα πιο διαδεδομένα πρότυπα κωδικοποίησης μηνυμάτων στην υγεία είναι το HL7. Παρόλα αυτά όμως, αν κάποιο σύστημα είναι συμβατό με την τρέχουσα έκδοση του HL7, δεν σημαίνει απαραίτητα ότι μπορεί να επικοινωνήσει εύκολα με συστήματα που χρησιμοποιούν παλαιότερες εκδόσεις του HL7, μιας και υπάρχουν σημαντικά κενά στην αντιστοιχία των ορισμών των εκδόσεων.

Άξονας 2: Διαλειτουργικότητα EHR

Το EHR ενός ασθενή, μπορεί να οριστεί ως η ψηφιακή αποθήκευση πληροφοριών υγείας σχετικά με έναν ασθενή με σκοπό τη χρήση όλων των συλλεχθέντων πληροφοριών για τη συνεχιζόμενη φροντίδα της υγείας, την εκπαίδευση, την έρευνα και τη διατήρηση σε κάθε βήμα της εμπιστευτικότητας. Επειδή οι πληροφορίες του EHR αποθηκεύονται σε πλήθος οργανισμών (διαφορετικά νοσοκομεία, κέντρα υγείας, ιατρεία, διαγνωστικά κέντρα) δεν υπονοείται η διαλειτουργικότητά τους. Η συνέχιση της φροντίδας όμως απαιτεί οι εμπλεκόμενοι στην παροχή υπηρεσιών υγείας (ιατροί, νοσηλευτές, κ.λπ.) να έχουν πρόσβαση στο σύνολο των κλινικών δεδομένων και του ιστορικού των κινήσεων του ασθενή. Για το σκοπό αυτό προωθούνται διάφορες προσπάθειες προτυποποίησης για την εξασφάλιση της επιθυμητής διαλειτουργικότητας των EHR, όπως η CEN/TC 251 EHRcom [86], openEHR [87] και HL7 CDA. Ακόμη και σήμερα όμως δεν έχει επιτευχθεί η ανταλλαγή δομημένων ψηφιακών πληροφοριών των EHR μεταξύ των υπάρχοντων συστημάτων. Άλλωστε η ύπαρξη πολλών και διαφορετικών προτύπων για τον ίδιο σκοπό, συνεπάγεται πως η χρήση ενός από αυτά δεν λύνει το πρόβλημα της διαλειτουργικότητας.

Όπως έχει αναφερθεί στο υπο-Κεφάλαιο 2.1.2, η διαλειτουργικότητα δεδομένων μπορεί να χωριστεί σε τέσσερις διαφορετικές κατηγορίες. Ομοίως και για τον τομέα της υγείας, το HIMSS έχει ορίσει τα ίδια τέσσερα επίπεδα διαλειτουργικότητας, τα οποία αναφέρονται συνοπτικά ακολούθως [77].

2.2.2.1 Νομική Διαλειτουργικότητα

Η Νομική Διαλειτουργικότητα επιτρέπει σε ένα σύστημα πληροφοριών να ανταλλάσσει δεδομένα με άλλο σύστημα πληροφοριών. Ουσιαστικά, αναπτύσσει τα δομικά στοιχεία ανταλλαγής πληροφοριών μεταξύ διαφορετικών συστημάτων, καθορίζοντας τις απαιτήσεις διασυνδετικότητας που απαιτούνται για ένα σύστημα ή μια εφαρμογή για να μοιράζονται δεδομένα και να λαμβάνουν δεδομένα από άλλο. Το σύστημα που λαμβάνει αυτές τις πληροφορίες δεν χρειάζεται να ερμηνεύει τα δεδομένα, τα οποία είναι άμεσα διαθέσιμα για χρήση.

2.2.2.2 Οργανωτική Διαλειτουργικότητα

Η Οργανωτική Διαλειτουργικότητα περιλαμβάνει τα τεχνικά στοιχεία καθώς και σαφή πολιτικά, κοινωνικά και οργανωτικά στοιχεία. Αυτά τα στοιχεία διευκολύνουν την ασφαλή, απρόσκοπτη και έγκαιρη επικοινωνία και χρήση δεδομένων εντός και μεταξύ των οργανισμών και των ατόμων. Η συμπερίληψη

αυτών των μη τεχνικών παραμέτρων καθιστά δυνατή τη διαλειτουργικότητα που ενσωματώνεται στις διαδικασίες και στις ροές εργασιών τελικού χρήστη, με τρόπο που να υποστηρίζει την αποτελεσματικότητα, τις σχέσεις και τη γενική υγεία και ευεξία, μέσω της συνεργασίας με κοινά δεδομένα τόσο εντός όσο και εντός οργανωτικών ορίων.

2.2.2.3 Σημαιολογική Διαλειτουργικότητα

Η Σημαιολογική Διαλειτουργικότητα είναι το υψηλότερο επίπεδο σύνδεσης. Δύο ή περισσότερα διαφορετικά συστήματα ή τμήματα συστημάτων μπορούν να ανταλλάξουν και να χρησιμοποιήσουν εύκολα πληροφορίες. Εδώ, η ίδια η δομή της ανταλλαγής δεδομένων και ο τρόπος κωδικοποίησης των ίδιων των δεδομένων επιτρέπει στους παρόχους να μοιράζονται δεδομένα ασθενών ακόμη και όταν χρησιμοποιούν τελείως διαφορετικές λύσεις λογισμικού από διάφορους προμηθευτές. Ουσιαστικά εκμεταλλεύεται τόσο τη διάρθρωση της ανταλλαγής δεδομένων όσο και την κωδικοποίηση των δεδομένων, συμπεριλαμβανομένου του τυποποιημένου διαθέσιμου στο κοινό λεξιλογίου, έτσι ώστε τα συστήματα διαχείρισης των πληροφοριών που λαμβάνουν τα εκάστοτε δεδομένα, να μπορούν να τα ερμηνεύουν. Αυτού του είδους διαλειτουργικότητα, έχει συνέπειες όχι μόνο για τη συνεργασία μεταξύ παρόχων υγειονομικής περίθαλψης αλλά και για ερευνητές και επιστήμονες που χρειάζονται μεγάλα ποσά δεδομένων για τη διεξαγωγή μελετών σχετικά με αναδυόμενες ασθένειες και άλλα προβλήματα δημόσιας υγείας.

2.2.2.4 Τεχνική (Συντακτική) Διαλειτουργικότητα

Η Τεχνική (Συντακτική) Διαλειτουργικότητα καθορίζει τη μορφή της ανταλλαγής δεδομένων. Αυτό έχει να κάνει με τα πρότυπα που διέπουν τη μορφή των μηνυμάτων που αποστέλλονται από το ένα σύστημα στο άλλο, έτσι ώστε ο επιχειρησιακός ή κλινικός σκοπός των πληροφοριών να είναι προφανής και να μεταφέρεται χωρίς αλλοίωση. Ουσιαστικά καθορίζει τη σύνταξη των δεδομένων που ανταλλάσσονται. Στην περίπτωση αυτή, αναφερόμαστε στις πληροφορίες αναφορικά με το επίπεδο των πεδίων δεδομένων, όπως σε μια βάση δεδομένων των αρχείων των ασθενών.

2.2.3 Πρότυπα Διαλειτουργικότητας στον τομέα της Υγείας

Οι βιομηχανοποιημένες κοινωνίες μας εξαρτώνται σε μεγάλο βαθμό από τα πρότυπα. Η συμμόρφωση με τα πρότυπα διευκολύνει την ανταλλαγή και τη δυνατότητα υποκατάστασης, δημιουργεί ανεξαρτησία από τους κατασκευαστές, διευκολύνει τον ανταγωνισμό και δημιουργεί διασυνοριακή διαλειτουργικότητα. Η τυποποίηση των δεδομένων υγείας καθιστά ευκολότερη τη σύγκριση, την κατηγοριοποίηση και, κατά συνέπεια, την υγειονομική περίθαλψη. Επιπλέον, η συμμόρφωση προς τα πρότυπα αυξάνει την εμπιστοσύνη στην ασφαλή λειτουργία των συστημάτων υπό προκαθορισμένες συνθήκες. Η προτυποποίηση απαιτείται εξίσου για τα δεδομένα γενικά και για τα κλινικά δεδομένα ειδικότερα, για τα οποία η ασφάλεια, η εναλλάξιμη λειτουργία και η διαλειτουργικότητα είναι ένας ανώτερος στόχος, ιδίως όσον αφορά το αναδυόμενο πεδίο της επιστήμης των δεδομένων.

Τα κλινικά δεδομένα υποδηλώνουν τους ασθενείς, τις καταγγελίες τους, σημεία, ασθένειες, χειρουργικές επεμβάσεις, φάρμακα, εργαστηριακές τιμές κλπ. Καταγράφονται σε πληροφοριακά συστήματα διαφόρων ειδών (ηλεκτρονικά αρχεία υγείας, αρχεία ασθενειών, τεκμηρίωση κλινικών δοκιμών, βάσεις δεδομένων θνησιμότητας) που είναι συχνά ετερογενή, ελλιπή και μερικές φορές εσφαλμένα. Τα κλινικά δεδομένα διαμορφώνονται σύμφωνα με τις ειδικές ανάγκες για τις οποίες συλλέγονται, όπως η αναφορά, η επικοινωνία και η χρέωση. Όπου η επικοινωνία μεταξύ των επαγγελματιών στον τομέα της υγείας είναι υψίστης σημασίας, οι κακώς δομημένες αφηγήσεις τείνουν να επικρατούν έναντι των δομημένων και κωδικοποιημένων δεδομένων. Δεδομένου ότι το κείμενο πρέπει απλώς να είναι κατανοητό από τους ανθρώπους, η χρήση ενός κοινού λεξιλογίου και ενός συνόλου χαρακτήρων είναι επαρκή και η ανοχή όσον αφορά τις αποκλίσεις γραμματικής και ορθογραφίας και τα σφάλματα δεν αποτελούν σημαντικά ζητήματα. Το ελεύθερο κείμενο είναι σημασιολογικά διαλειτουργικό μόνο εάν και τα δύο μέρη χρησιμοποιούν τις λέξεις με την ίδια ακριβώς έννοια και το ίδιο περιεχόμενο. Ωστόσο, η περαιτέρω επεξεργασία δεδομένων, π.χ. τα λεγόμενα σενάρια δευτερογενούς χρήσης για κλινικά δεδομένα όπως η υποστήριξη της λήψης αποφάσεων μέσω υπολογιστή, η αναδρομική και η πρόβλεψη των δεδομένων, τείνουν να παρεμποδίζονται από τοπικά λεξικά δεδομένων και να λείπουν σχετικές περιγραφές. Ανεξάρτητα από το εάν απευθύνονται σε σενάρια πρωτογενούς ή δευτερογενούς χρήσης για κλινικά δεδομένα, τα τελευταία πρέπει να ακολουθούν κοινά πρότυπα. Τα πρότυπα αυτά πρέπει να περιγράφουν:

- Την προέλευση των δεδομένων, αναφορικά με τους δημιουργούς τους, τους χρόνους δημιουργίας και τις σχετικές διαδικασίες.
- Τα πρότυπα πληροφοριών στα οποία ενσωματώνονται τα δεδομένα.
- Τα Λεξιλόγια/ορολογίες/οντολογίες που χρησιμοποιούνται για τη σύνδεση της έννοιας με τα δεδομένα.
- Τις σημασιολογικές περιγραφές ή τις αναπαραστατικές μονάδες (κωδικοί, ετικέτες) σε αυτά τα λεξιλόγια.
- Τους τυπικούς ή κειμενικούς ορισμούς αυτών των παραστατικών μονάδων.
- Τις επίσημες γλώσσες που χρησιμοποιούνται για τα παραπάνω.

Φαίνεται πως ένα από τα σημαντικότερα θέματα και προβλήματα στο χώρο της υγείας είναι η έλλειψη μιας μοναδικής ή αλλιώς ταυτόχρονης πολύπλευρης κωδικοποίησης και ταξινόμησης. Η παραγωγή δεδομένων στον τομέα της υγείας είναι υπέρογκη. Η πηγή (διοικητικά, οικονομικά, ιατρικά, νοσηλευτικά, κ.ο.κ.) και η μορφή τους (ηχητικά, βίντεο, εικόνες, κείμενα, βιοσήματα) ποικίλει. Ακόμη και σήμερα η έλλειψη ενός τυποποιημένου λεξικού, ορολογίας, προτύπων και πρωτοκόλλων στο χώρο των επιστημών της υγείας, αποτελεί σημαντικό σκόπελο στην περαιτέρω προώθηση αυτοματοποιημένων συστημάτων και διαδικασιών. Η χρήση των υπολογιστών στη διαχείριση των δεδομένων υγείας κάνει ακόμη πιο αισθητό το συγκεκριμένο πρόβλημα, μιας και απαιτεί ομοιομορφία στη συστηματοποίηση και κωδικοποίηση των δεδομένων και των ορισμών τους προκειμένου να επιτευχθεί η ορθή επεξεργασία τους από τα πληροφοριακά συστήματα. Μέσω της ορθής κωδικοποίησης, επιτρέπεται η τυποποιημένη και συστηματική αποθήκευση δεδομένων ώστε να διατηρούνται τα ιατρικά ιστορικά, να υποβοηθούνται οι διαγνώσεις και να παρέχονται εν τέλει καλύτερες και ποιοτικότερες υπηρεσίες υγείας. Ταυτόχρονα, όπως και σε κάθε άλλη διοικητική διαδικασία, μέσω της κωδικοποίησης, μπορούμε να συγκρίνουμε τα δεδομένα και να καταλήξουμε σε διοικητικά και στατιστικά συμπεράσματα, χάρασοντας έγκαιρα την κατάλληλη στρατηγική. Ως κωδικοποίηση στο χώρο της υγείας [88], ορίζουμε τη διαδικασία οργάνωσης δεδομένων υγείας σε κατηγορίες, στις οποίες δίνονται κωδικοί, (που είναι συνήθως αριθμητικοί ή αλφαριθμητικοί) με σκοπό τη συντόμευση, αποθήκευση και εύκολη ανάκτηση των δεδομένων αυτών. Εναλλακτικά μπορούμε να την ορίσουμε ως το μετασχηματισμό των λεκτικών περιγραφών των ασθενειών, των τραυματισμών, ή των διαδικασιών, σε

αλφαριθμητικούς προσδιορισμούς. Ένα σύστημα κωδικοποίησης οφείλει σε έναν μόνο κωδικό να συλλέγει όλες τις ορολογίες που αναφέρονται, π.χ. σε μία ασθένεια και τα λογικά συνώνυμά της. Η κωδικοποίηση προσφέρει επομένως μια «κοινή γλώσσα» αναπαράστασης και ερμηνείας των δεδομένων και των εννοιών, το οποίο είναι βασικό προαπαιτούμενο για την επίτευξη της σημασιολογικής διαλειτουργικότητας. Με αυτό τον τρόπο μειώνονται η ασάφεια και τα σφάλματα και αναβαθμίζεται η ποιότητα των δεδομένων, ενώ επιτυγχάνεται η τυποποίηση και η συγκρισιμότητα των δεδομένων, η ταχύτερη επεξεργασία και αναζήτηση των δεδομένων μέσω υπολογιστών και η παραγωγή λογικών συσχετισμών μεταξύ τους. Η κωδικοποίηση λοιπόν αποτελεί την απόδοση ενός κωδικού σε μια οντότητα. Αν και υφίστανται πολλών ειδών κωδικοποιήσεις, τα δύο κυριότερα είδη στη σημασιολογία [88] είναι: η ταξινόμηση (classification), η οποία έχει ως κύριο χαρακτηριστικό την απόδοση κωδικών σε οντότητες ώστε οι οντότητες να ορίζονται κατά το δυνατόν αμφιμονοσήμαντα, και η ονοματολογία (nomenclature), η οποία είναι η ταξινομημένη συλλογή συνδεδεμένων όρων. Άλλα είδη κωδικοποίησης αποτελούν ο θησαυρός λέξεων (thesaurus), η ταξινομία (taxonomy), και η επίσημη ορολογία (formal terminology). Επομένως, γίνεται άμεσα αντιληπτή η ανάγκη για διεθνή συνεργασία στην προσπάθεια ενιαίας κωδικοποίησης, ταξινόμησης και προτύπων στο χώρο της υγείας. Τέτοιες ενέργειες έχουν ήδη γίνει αρκετές (π.χ. International Classification of Diseases 10 (ICD-10) [89], Digital Imaging and Communications in Medicine (DICOM) [90], HL7, Systematized Nomenclature of Human and Veterinary Medicine (SNOMED) [91], Logical Observation Identifiers Names and Codes (LOINC) [92]). Η Παγκόσμια Οργάνωση Υγείας (World Health Organization – WHO) [93] έχει εκδώσει το ICD-10 ως ένα παράδειγμα ταξινόμησης, ενώ η κωδικοποίηση SNOMED αποτελεί παράδειγμα μιας ονοματολογίας. Ο τρόπος κωδικοποίησης δε θα πρέπει απλά να επικεντρώνεται στον περιγραφικό σκοπό των δεδομένων, αλλά να στοχεύει στην αξιοποίησή τους σε κλινικό, οικονομικό και διοικητικό επίπεδο. Είναι σημαντικό επομένως να βρεθεί μια κοινή ορολογία τόσο σε επίπεδο ονοματολογίας όσο και σε επίπεδο κωδικοποίησης η οποία θα μπορεί να αποδίδει με ποιότητα και αξιοπιστία την παραγόμενη, και συλλεχθείσα πληροφορία στο χώρο της υγείας. Σύμφωνα με τον Krippendorff [94], η αξιοπιστία παράγεται μέσω της σταθερότητας, της ικανότητας αναπαραγωγής και ακρίβειας, ενώ η ποιότητα με τη σωστή απόδοση των όρων, έτσι ώστε να επιτρέπεται η ανάκτηση των δεδομένων με συνεπή τρόπο. Όλα τα συστήματα πληροφορικής επεξεργάζονται και αποθηκεύουν δεδομένα, αλλά και επικοινωνούν μεταξύ τους. Ο τρόπος επεξεργασίας, αποθήκευσης και μεταφοράς των δεδομένων όμως δεν είναι ο ίδιος για όλα τα συστήματα. Για αυτό το λόγο, θα πρέπει να στηρίζονται σε ένα πρότυπο.

Στο χώρο της υγείας μπορούμε να διακρίνουμε μια σειρά από βασικές κατηγορίες προτύπων:

- Πρότυπα επικοινωνίας
- Πρότυπα αναπαράστασης κλινικών δεδομένων
- Πρότυπα αναγνώρισης
- Πρότυπα ασφάλειας ιατροτεχνολογικών προϊόντων
- Πρότυπα αρχιτεκτονικής ηλεκτρονικού φακέλου υγείας
- Πρότυπα ασφάλειας των δεδομένων και εξασφάλισης του ιατρικού απορρήτου

Με την εισαγωγή των υπολογιστών στη διαχείριση των δεδομένων υγείας, είναι αναγκαία η ομοιομορφία στη συστηματοποίηση/κωδικοποίηση των δεδομένων αυτών και των ορισμών τους. Χωρίς μια προκαθορισμένη ορολογία είναι αδύνατον να γίνει μια αυτοματοποιημένη σύνθεση των δεδομένων υγείας. Για παράδειγμα, σε ένα συγκεκριμένο ασθενή μπορεί να καταγραφεί από έναν ιατρό ότι έχει «βραχύτητα αναπνοής» και αργότερα κάποιος άλλος ιατρός να καταχωρήσει για τον ίδιο ασθενή ότι έχει «δύσπνοια». Ένα πρόγραμμα υπολογιστή θα αποτύχει στο να υποδείξει ότι ο συγκεκριμένος ασθενής έχει το ίδιο πρόβλημα και στις δύο αυτές περιπτώσεις. Σε παγκόσμιο επίπεδο, υπάρχει μια τεράστια κινητικότητα στο θέμα της ανάπτυξης προτύπων στο χώρο της υγείας. Για το λόγο αυτό, υφίσταται πληθώρα προτύπων σε όλα τα επίπεδα. Οι ομάδες που τα αναπτύσσουν προσπαθούν να χρησιμοποιήσουν ως βάση τα καλύτερα χαρακτηριστικά από τα υπάρχοντα πρότυπα, παρά να ξεκινούν από το μηδέν. Παρόλα αυτά, ακόμη δεν υπάρχει κάποιο κοινά αποδεκτό πρότυπο σε διεθνές επίπεδο που να εξασφαλίζει τα επιθυμητά οφέλη. Στην Ευρώπη, η TC 251) του Ευρωπαϊκού Οργανισμού Τυποποίησης CEN είναι υπεύθυνη για την τυποποίηση μηνυμάτων πληροφορικής στην υγεία.

Μέχρι σήμερα, η υιοθέτηση προτύπων για τα κλινικά δεδομένα είναι χαμηλή. Η διόρθωση αυτής της καθυστέρησης θα είναι ζωτικής σημασίας για την εξάπλωση του δυναμικού των κλινικών δεδομένων για ποικίλα σενάρια επαναχρησιμοποίησης. Αυτό απαιτεί σημαντικές προσπάθειες από όλους τους εμπλεκόμενους φορείς, τους δημιουργούς και τους διατηρητές των προτύπων, καθώς και τους χρήστες τους. Στον Πίνακα 2-1, διαφαίνονται ορισμένα από τα πιο σημαντικά πρότυπα στον χώρο της υγείας, εκ των οποίων, ορισμένα από αυτά αναλύονται στα κεφάλαια που ακολουθούν.

Πίνακας 2-1 – Πρότυπα στον τομέα της υγείας

Οργανισμός Προτύπων	Πρότυπο	Σκοπός
Federative Committee on Anatomical Terminology (FCAT) [95]	Terminologia Anatomica (TA)	Ορολογίες ανατομίας στα αγγλικά και στα λατινικά
	V2	Πρωτόκολλο μηνυμάτων
	V3 (RIM)	Πληροφορίες οντολογιών
Health Level Seven (HL7)	CDA Level 1-3	Μοντέλο πληροφοριών για κλινικά έγγραφα
	FHIR	Μοντέλο πληροφοριών και εγγράφων
Integrating the Healthcare Enterprise (IHE) [96]	Several Integration Profiles	Κλινικές ροές εργασίας με αναφορές σε κλινικά δεδομένα
	TS22220:2011	Ταυτοποίηση θεμάτων φροντίδας
	21090:2011	Εναρμονισμένοι τύποι δεδομένων για ανταλλαγή πληροφοριών
International Organization for Standardization (ISO)	13606	Υψηλού επιπέδου περιγραφή κλινικών μοντέλων πληροφοριών
	23940 (ContSys)	Διαδικασίες υγείας για τη συνέχιση της φροντίδας
	14155	Κλινικές έρευνες
	IDMP	Ιατρικά προϊόντα
National Electrical Manufacturers Association (NEMA) [97]	DICOM	Ιατρικές εικόνες και σχετικά δεδομένα
	openEHR foundation	Προσδιορισμός κλινικών μοντέλων πληροφοριών
Regenstrief Institute [98]	LOINC	Κωδικοποίηση για κλινικές παρατηρήσεις
	UCUM	Προτυποποιημένη αναπαράσταση μονάδων μέτρησης σύμφωνα με το System International (SI)

PCHAlliance (Personal Connected Health Alliance) [99]	Continua Design Guidelines	Συλλογή δεδομένων από προσωπικές ιατρικές συσκευές
SNOMED International γνωστό ως International Health Terminology Standards Development Organization		Κωδικοποίηση/ Οντολογίες για την αναπαράσταση Ηλεκτρονικών Φακέλων Υγείας
World Health Organization (WHO)	ICD-10/ ICD- 11	Κατηγοριοποίηση ασθενειών
	ICF	Κατηγοριοποίηση λειτουργικότητα, δυσλειτουργιών και υγείας
	ICHI	Κατηγοριοποίηση ιατρικών διαδικασιών
	INN	Γενική ονομασία για υποκατάστατα φαρμάκων
	ATC	Κατηγοριοποίηση συστατικών φαρμάκων
World Organization of Family Doctors (WONCA) [100]	ICPC	Κατηγοριοποίηση πρωτεύουσας φροντίδας

2.2.3.1 DICOM (Digital Imaging and Communications in Medicine)

Το πρότυπο επικοινωνίας
ιατρικής ψηφιακής εικόνας
DICOM προτάθηκε από το
Αμερικανικό Κολλέγιο



Ραδιολογίας (American College of Radiology - ACR) [101] και την Εθνική Ένωση Κατασκευαστών Ηλεκτρολόγων (National Electrical Manufacturers Association - NEMA) το 1985. Το πρότυπο χρησιμοποιείται συνήθως για τη μεταφορά, αποθήκευση, διανομή και εξέταση ιατρικών εικόνων. Το πρότυπο DICOM, επιχειρεί να κωδικοποιήσει τόσο τα στοιχεία τα οποία συνοδεύουν τις ιατρικές εικόνες, όσο και τις ίδιες τις εικόνες, ώστε να επιτευχθεί η αυτόματη

σύνδεση των συσκευών παραγωγής εικόνας με το νοσοκομειακό πληροφοριακό σύστημα, ανεξαρτήτως συσκευής ή λειτουργικού συστήματος. Το πρότυπο επικοινωνίας ιατρικής ψηφιακής εικόνας DICOM αποτελεί σήμερα ένα παγκόσμιο πρότυπο που χρησιμοποιείται στα περισσότερα σύγχρονα νοσηλευτικά ιδρύματα. Η δομή του σχεδιάστηκε έτσι ώστε να εξασφαλίζει τη διαλειτουργικότητα των συστημάτων με σκοπό την παραγωγή, αποθήκευση, απεικόνιση, επεξεργασία, αποστολή, ανάκτηση, αναζήτηση και εκτύπωση ιατρικών εικόνων και των παραγόμενων αναφορών, καθώς και των σχετικών ροών εργασίας. Είναι σημαντικό να αναφερθεί πως ο τρόπος δόμησης και κατασκευής των αρχείων DICOM επιτρέπει σε έναν εξυπηρετητή αρχείων να αποθηκεύει και να αναζητά εικόνες χωρίς να γνωρίζει τον τρόπο με τον οποίο κωδικοποιούνται τα δεδομένα της εικόνας. Με την αγορά εξοπλισμού όπου συμμορφώνεται με το συγκεκριμένο πρότυπο, εξασφαλίζεται η επιθυμητή διαλειτουργικότητα για την επικοινωνία, παραγωγή, διαχείριση και διανομή των ιατρικών εικόνων. Όλα τα δεδομένα (π.χ. ασθενείς, μηχανήματα, μελέτες, κ.λπ.) θεωρούνται από το DICOM ως αντικείμενα (objects) τα οποία έχουν συγκεκριμένες ιδιότητες (attributes). Για το λόγο αυτό θεωρούμε ότι το DICOM έχει αντικειμενοστραφή σχεδίαση. Τα αντικείμενα αυτά και οι ιδιότητές τους ακολουθούν το DICOM Information Object Definitions (IODs). Το κάθε IOD μπορεί να έχει ένα σύνολο από χαρακτηριστικά που το περιγράφουν. Το DICOM περιέχει άνω των 2.000 χαρακτηριστικών (καλύπτοντας ουσιαστικά τα βασικά χαρακτηριστικά), που βρίσκονται στο λεξικό δεδομένων του DICOM (DICOM Data Dictionary), ώστε να υπάρχει συνοχή μεταξύ της ονοματολογίας, της μορφοποίησης και της επεξεργασίας των διαφόρων αυτών χαρακτηριστικών. Για παράδειγμα, το ονοματεπώνυμο του ασθενή, το φύλο, η ημερομηνία γέννησης, κ.ά., εμπεριέχονται όλα στο DICOM Data Dictionary. Τα χαρακτηριστικά αυτά μορφοποιούνται σύμφωνα με 27 Value Representation τύπους που περιλαμβάνουν μεταξύ άλλων την ημερομηνία, τον χρόνο, το όνομα, αναγνωριστικά, κ.ο.κ. Έχοντας πλέον τα δεδομένα σε μορφή DICOM, μπορούν να διοχετευτούν μέσω του δικτύου, μεταξύ διαφορετικών μηχανημάτων και λογισμικού που υποστηρίζουν DICOM. Αυτά στο DICOM ονομάζονται Οντότητες Εφαρμογής (Application Entities - AEs) και μπορούν και παρέχουν υπηρεσίες μεταξύ τους. Ενδεικτικά παραδείγματα AEs είναι ο αξονικός τομογράφος, ο μαγνητικός τομογράφος, ο υπερηχογράφος, ο υπολογιστής ή το κινητό με δυνατότητα προβολής DICOM εικόνων, ο αποθηκευτικός χώρος με λογισμικό επεξεργασίας εικόνων DICOM, κ.ά. Επειδή η κάθε υπηρεσία μεταξύ των AEs συνοδεύεται από την ανταλλαγή συγκεκριμένων δεδομένων, συνηθίζεται να συνδυάζουμε τον τύπο των υπηρεσιών με τα δεδομένα IODs που επεξεργάζονται. Αυτές οι συσχετίσεις ονομάζονται στο DICOM ως Ζεύγη Υπηρεσιών - Αντικειμένων (Service-Object-

Pairs - SOPs) και ομαδοποιούνται σε κλάσεις που λέγονται SOP Classes. Οι αιτούντες την υπηρεσία ονομάζονται στο DICOM ως Χρήστες Υπηρεσιών Κλάσης (Service Class Users - SCUs) και οι παρέχοντες την υπηρεσία Πάροχοι Υπηρεσιών Κλάσης (Service Class Providers - SCPs). Όπως είναι φυσικό, αυτές οι σχέσεις είναι σχετικές και μπορεί να αλλάξουν ανά πάσα στιγμή, ανάλογα με το ποιο σύστημα αιτείται μια υπηρεσία και ποιο την εξυπηρετεί. Η επίτευξη μιας τέτοιας ανταλλαγής δεδομένων μεταξύ ενός SCU και ενός SCP ονομάζεται συσχέτιση (association). Έτσι, κάθε μεταφορά ξεκινά με την εγκαθίδρυση μιας συσχέτισης (Association Establishment) και η πληροφορία που ανταλλάσσουν μεταξύ τους οι δύο συσκευές ονομάζεται Πλαίσια Παρουσίασης (Presentation Contexts). Στην περίπτωση επομένως όπου ταιριάζουν τα πλαίσια (contexts) των δύο συσκευών, επιτυγχάνεται η μεταξύ τους σύνδεση και μπορεί πλέον να ξεκινήσει η επεξεργασία SCU-SCP. Λόγω της ύπαρξης διαφορετικών συσκευών και εφαρμογών DICOM που προέρχονται από διαφορετικούς κατασκευαστές, η κάθε μονάδα συνοδεύεται από τη δική της Δήλωση Συμμόρφωσης DICOM (DICOM Conformance Statement). Μέσω αυτού γνωρίζει κανείς ποιες υπηρεσίες SOPs υποστηρίζονται και σε ποιο βαθμό (SCP, SCU ή και οι δύο). Εν κατακλείδι, το DICOM έχει παίξει καταλυτικό ρόλο στην σύγχρονη ιατρική, με ποικίλους τρόπους:

- Υποστηρίζει εξαιρετικής ποιότητας εικόνες με δυνατότητα 65.536 (16 bits) αποχρώσεων του γκρι για μονόχρωμες απεικονίσεις. Έτσι μπορεί και ξεχωρίζει ακόμη και τις απειροελάχιστες διαφορές σε χρωματικό τόνο σε ιατρικές εικόνες.
- Όλοι οι σύγχρονοι εξοπλισμοί σύλληψης ιατρικών ψηφιακών εικόνων υποστηρίζουν πλέον το συγκεκριμένο πρότυπο, παράγοντας DICOM εικόνες και επικοινωνώντας μέσα από δίκτυα DICOM.
- Μέσω του DICOM δεν αποθηκεύονται μόνον οι εικόνες με κατάλληλο τρόπο, αλλά καταχωρούνται και άλλα πολύτιμα χαρακτηριστικά και πληροφορίες τους, όπως: η θέση του ασθενή σε τρεις διαστάσεις, τα μεγέθη και οι προσανατολισμοί των εικόνων, το πάχος της τομής που απεικονίζεται, η δοσολογία της ακτινοβολίας και το μέγεθος της έκθεσης, η χρήση φίλτρων επεξεργασίας εικόνων, κ.ά. Οι εικόνες εμπλουτίζονται με όλες αυτές τις πληροφορίες δίνοντας την δυνατότητα για περαιτέρω επεξεργασία και εύκολη ερμηνεία. Για παράδειγμα, είναι δυνατή η ανασύνθεση μιας τριοδιάστατης εικόνας από τις τομές της αρχικής εικόνας σε δύο διαστάσεις.

- Τα DICOM μηνύματα και αρχεία χρησιμοποιούν ένα πλήθος από ιδιότητες (που καθορίζονται στο DICOM Data Dictionary) ώστε να μεταφέρουν πολύτιμα δεδομένα, όπως το ονοματεπώνυμο του ασθενή, το βάθος χρώματος της εικόνας, τη διάγνωση του ασθενή, κ.ο.κ.
- Το DICOM επιτρέπει να καθοριστούν οι δυνατότητες λειτουργίας ενός ιατρικού μηχανήματος με μεγάλη ακρίβεια και ανεξάρτητα από τον κατασκευαστή.

2.2.3.2 SNOMED (Systematized Nomenclature of Medicine - Clinical Terms)

Το SNOMED CT, θεωρείται ως μια από τις πιο περιεκτικές, πολύγλωσσες κλινικές ορολογίες υγειονομικής περιθαλψής στον κόσμο. Η υιοθέτηση μιας τυποποιημένης κλινικής ορολογίας για χρήση σε συστήματα πληροφοριών υγείας σε παγκόσμιο επίπεδο μπορεί να έχει σημαντική συμβολή προς τη βελτίωση της ποιότητας και της ασφάλειας των παρεχόμενων υπηρεσιών. Το SNOMED CT αποτελεί μια κλινική ορολογία η οποία παρέχει κλινικό περιεχόμενο και κοινό τρόπο έκφρασης για κλινική τεκμηρίωση και αναφορές. Μπορεί να χρησιμοποιηθεί στην κωδικοποίηση, ανάκτηση και ανάλυση των κλινικών δεδομένων. Το SNOMED CT προέκυψε από τη συγχώνευση του SNOMED Reference Terminology (SNOMED RT) [102] που αναπτύχθηκε από το Κολέγιο Αμερικανών Παθολόγων (College of American Pathologists - CAP) [103] και τους Κλινικούς Όρους Έκδοση 3 (Clinical Terms Version 3 - CTV3) [104] που αναπτύχθηκαν από τον Εθνικό Οργανισμό Υγείας (National Health Service - NHS) του Ηνωμένου Βασιλείου. Η ορολογία αποτελείται από έννοιες, όρους και σχέσεις με σκοπό την ακριβή αναπαράσταση της κλινικής πληροφορίας στο χώρο της υγείας και χρησιμοποιείται ευρέως σε ηλεκτρονικούς φακέλους υγείας. Οι περισσότερες εφαρμογές στο χώρο της υγείας εστιάζονται στη συλλογή κλινικών δεδομένων, τη συσχέτισή τους με κλινικές βάσεις γνώσης, την ανάκτηση πληροφοριών, καθώς και στο συνδυασμό δεδομένων και στην ανταλλαγή τους. Λόγω της χαοτικής μορφής και δομής του χώρου της υγείας, οι πληροφορίες και τα δεδομένα μπορεί να προκύψουν από διαφορετικούς χώρους, σε διαφορετικές στιγμές και με διαφορετικό τρόπο. Το SNOMED CT προσφέρει την προτυποποίηση της κλινικής πληροφορίας. Οι εφαρμογές μπορούν να χρησιμοποιήσουν τις έννοιες, τις ιεραρχίες και τις συσχετίσεις ως κοινό έδαφος για την ανάλυση των δεδομένων. Έτσι μπορούν να προκύψουν εύκολα τα αποτελέσματα μελετών και ερευνών, να αποτιμηθεί η ποιότητα και



το κόστος των παρεχόμενων υπηρεσιών, κ.ο.κ. Η χρήση προτυποποιημένης ορολογίας ωφελεί τόσο τους επαγγελματίες υγείας (ιατρούς, νοσηλευτές, κ.λπ.) όσο και τους ασθενείς, τους διαχειριστές και τους τεχνικούς που αναπτύσσουν τις κλινικές εφαρμογές. Μέσω του SNOMED CT υποβοηθείται η παροχή αξιόπιστης, πλήρους και κατάλληλα δομημένης πληροφορίας στους επαγγελματίες υγείας (π.χ. ιατρικό ιστορικό, ασθένειες, εργαστηριακά αποτελέσματα, θεραπευτικές αγωγές, κ.λπ.), η οποία με τη σειρά της βελτιώνει σημαντικά τα αποτελέσματα για τον ίδιο τον ασθενή. Η κατάλληλη κλινική ορολογία επιτρέπει στους παρόχους υπηρεσιών υγείας να αναγνωρίζουν εύκολα τους ασθενείς, βάσει της σωστής κωδικοποίησης των πληροφοριών στο φάκελό τους και συνεπώς διευκολύνουν τις αποφάσεις για την ορθότερη δυνατή θεραπευτική αγωγή. Το ζήτημα της διαλειτουργικότητας και της ηλεκτρονικής ανταλλαγής αρχείων υγείας διευκολύνεται σημαντικά από το SNOMED CT. Το SNOMED-CT σχεδιάστηκε και υλοποιήθηκε έτσι ώστε να μην υπάρχουν μεγάλες επικαλύψεις των πολλαπλών αξόνων του. Έτσι, η χρήση του δίνει τη δυνατότητα δημιουργίας, αλλά και κατανόησης από τον υπολογιστή απλών φράσεων με ιατρικό περιεχόμενο. Επειδή η κωδικοποίηση SNOMED-CT έχει μεγάλο αριθμό από όρους, θεωρείται δύσχρηστη για άμεση χρήση από τον άνθρωπο, ενώ έχει αξία σε αυτοματοποιημένα συστήματα κωδικοποίησης.

2.2.3.3 ICD-10 (International Classification of Diseases)

Το ICD, είναι ένα σύστημα που ασχολείται με την ομαδοποίηση/ ταξινόμηση των ασθενειών, η οποία στηρίζεται στην αιτιολογία, και τη συμπτωματολογία. Ουσιαστικά καταχωρεί τις παθολογικές



οντότητες σύμφωνα με προκαθορισμένα κριτήρια. Μεταξύ των λόγων χρήσης του ICD-10 (10η έκδοση του ICD) περιλαμβάνονται η ανάλυση της γενικής κατάστασης της υγείας των ομάδων πληθυσμών και του ελέγχου της συχνότητας και της εξάπλωσης των ασθενειών και άλλων προβλημάτων υγείας σε σχέση με άλλες μεταβλητές, όπως τα χαρακτηριστικά και οι περιπτώσεις των ατόμων που έχουν επηρεαστεί, οι αποζημιώσεις, η κατανομή των πόρων, η ποιότητα και οι οδηγίες. Χρησιμοποιείται για να ταξινομήσει τις ασθένειες και άλλα προβλήματα υγείας που καταγράφονται σε πολλούς τύπους αρχείων ζωτικής σημασίας, συμπεριλαμβανομένων των πιστοποιητικών θανάτου και των αρχείων υγείας. Εκτός από τη διευκόλυνση της αποθήκευσης και της ανάκτησης των διαγνωστικών πληροφοριών για κλινικούς, επιδημιολογικούς και ποιοτικούς λόγους, αυτά τα αρχεία παρέχουν επίσης τη βάση για τη σύνταξη των εθνικών στατιστικών θνησιμότητας και νοσηρότητας. Στην

αρχική του έκδοση περιλάβανε τις εξής ομάδες: επιδημικά νοσήματα (λοιμώδη), ιδιοσυστατικά ή γενικά νοσήματα, τοπικά νοσήματα ανάλογα με την ανατολική εντόπιση, διαταραχές ανάπτυξης και τραυματισμοί. Η τελευταία αναθεώρηση της διεθνούς ταξινόμησης των ασθενειών είναι το σύστημα ICD-10, που δημοσιεύθηκε το 1992. Η ταξινόμηση κατά ICD, αποτελεί έναν απλό κωδικοποιημένο κατάλογο κατηγοριών τριών χαρακτήρων, καθεμιά από τις οποίες μπορεί να διαιρεθεί περαιτέρω μέχρι και σε δέκα υποκατηγορίες τεσσάρων χαρακτήρων. Στη θέση του καθαρά αριθμητικού συστήματος κωδικοποίησης των προηγούμενων αναθεωρήσεων, η δέκατη αναθεώρηση (ICD-10) χρησιμοποιεί ένα αλφαριθμητικό κωδικό ο οποίος έχει ένα γράμμα στην πρώτη θέση και έναν αριθμό στη δεύτερη, τρίτη και τέταρτη θέση. Ο τέταρτος χαρακτήρας ακολουθεί μετά από τελεία, σαν να αποτελεί δεκαδικό ψηφίο. Επομένως, μπορούν να υπάρχουν κωδικοί αριθμοί που κυμαίνονται από A00.0 μέχρι Z99.9. Το γράμμα U δε χρησιμοποιείται. Το σύστημα αυτό λαμβάνει υπόψη νέες εμπειρίες και γνώσεις, καλύπτει ορισμένα κενά σε σχέση με τις προηγούμενες εκδόσεις και παρουσιάζει καινούργια δομή, χρησιμοποιώντας αλφαριθμητικό κώδικα (π.χ. η ασθένεια «ερυθηματώδης λύκος» έχει κωδικό L23, ενώ ο «υπό οξύς δερματικός ερυθηματώδης λύκος» τον κωδικό L23.1). Ειδικότερα, η ταξινόμηση διαιρείται σε 21. Ο πρώτος χαρακτήρας του κωδικού της ICD είναι ένα γράμμα, και κάθε γράμμα σχετίζεται με ένα συγκεκριμένο κεφάλαιο, εκτός από το γράμμα D που χρησιμοποιείται τόσο στο Κεφάλαιο II - Νεοπλάσματα, όσο και στο Κεφάλαιο III - Νόσοι του αίματος και των αιμοποιητικών οργάνων και διαταραχών που εμπλέκουν τον ανοσολογικό μηχανισμό, και το γράμμα H που χρησιμοποιείται, τόσο στο Κεφάλαιο VII - Νοσήματα του οφθαλμού και των εξαρτημάτων του, όσο και στο Κεφάλαιο VIII - Παθήσεις του ωτός και της μαστοειδούς απόφυσης. Τέσσερα κεφάλαια (Κεφάλαια I, II, XIX και XX) χρησιμοποιούν περισσότερα από ένα γράμμα στην πρώτη θέση των κωδικών τους. Κάθε κεφάλαιο περιέχει αριθμό κατηγοριών τριών χαρακτήρων, ικανό να καλύψει τα περιεχόμενά του. Είναι σημαντικό να τονισθεί, πως δεν χρησιμοποιούνται όλοι οι διαθέσιμοι κωδικοί και συνεπώς υπάρχει περιθώριο για μελλοντικές αναθεωρήσεις και επεκτάσεις. Τα Κεφάλαια I έως XVII σχετίζονται με νοσήματα και άλλες νοσολογικές καταστάσεις, και το Κεφάλαιο XIX με τραυματισμούς, δηλητηριάσεις και ορισμένες άλλες συνέπειες εξωτερικών αιτιών. Τα υπόλοιπα κεφάλαια περιλαμβάνουν θέματα σχετικά με τα διαγνωστικά δεδομένα. Το Κεφάλαιο XVIII καλύπτει τα συμπτώματα, σημεία και παθολογικά, κλινικά και εργαστηριακά ευρήματα, που δεν είναι ταξινομημένα κάπου αλλού. Το Κεφάλαιο XX, τις εξωτερικές αιτίες νοσηρότητας και θνησιμότητας. Το κεφάλαιο αυτό χρησιμοποιήθηκε παραδοσιακά για την ταξινόμηση των αιτιών των τραυματισμών και

δηλητηριάσεων, αλλά, από την ένατη αναθεώρηση (ICD-9) και μετά, καλύπτει και όλες τις καταγεγραμμένες εξωτερικές αιτίες νόσων και άλλων νοσολογικών καταστάσεων. Το Κεφάλαιο XXI - Παράγοντες που επηρεάζουν την κατάσταση υγείας και την επαφή με τις υγειονομικές υπηρεσίες, προορίζεται για την ταξινόμηση των δεδομένων που επεξηγούν την αιτία επαφής με τις υπηρεσίες υγείας ενός ατόμου το οποίο δεν ασθενεί στην παρούσα φάση, ή τις συνθήκες υπό τις οποίες γίνεται η περίθαλψη του ασθενούς στο συγκεκριμένο χρονικό διάστημα ή τις συνθήκες που έχουν κάποια άλλη επίδραση στην περίθαλψη αυτού του ατόμου. Μέσω του ICD-10 επιτρέπεται η συστηματική ανάλυση των καταγραφών, η ερμηνεία και η σύγκριση των δεδομένων θνησιμότητας και νοσηρότητας που έχουν συγκεντρωθεί από διάφορες χώρες ή περιοχές και σε διαφορετικές χρονικές περιόδους.

2.2.3.4 LOINC (Logical Observation Identifiers Names and Codes)

Το LOINC διατηρείται από το Regenstrief Institute του Indiana University των ΗΠΑ. Ξεκίνησε το 1994 από το συγκεκριμένο



ινοστιτούτο, ως ανάγκη για την ψηφιακή μετακίνηση των κλινικών δεδομένων από τα εργαστήρια που παράγουν τα δεδομένα για τα νοσοκομεία, τους ιατρούς και τους λοιπούς χρήστες των κλινικών πληροφοριών. Σκοπός του είναι η διευκόλυνση στην ανταλλαγή πληροφοριών και στην εξαγωγή συμπερασμάτων για κλινική, διοικητική χρήση και έρευνα. Το LOINC κωδικοποιεί τόσο εργαστηριακά, όσο και κλινικά αποτελέσματα. Το LOINC καλύπτει όλες τις παρατηρήσεις σε κλινικά εργαστήρια, συμπεριλαμβανομένων των ακόλουθων τομέων ειδικότητας: χημεία (συμπεριλαμβανομένης της παρακολούθησης των θεραπευτικών φαρμάκων και της τοξικολογίας), αιματολογία, μικροβιολογία (συμπεριλαμβανομένης της παρασιτολογίας και της ιολογίας), ορολογικών εξετάσεων, κυτταρολογία, τράπεζα αίματος, χειρουργική παθολογία και γονιμότητα. Καλύπτει επίσης και τους περισσότερους όρους από την κτηνιατρική. Συμπεριλαμβάνει επίσης και μετρήσεις που απαιτούνται συνήθως για την ερμηνεία των ελέγχων, όπως για παράδειγμα η δόση για τις συγκεντρώσεις των φαρμάκων που χρησιμοποιούνται στη φαρμακοκινητική, οι μονάδες αίματος που χορηγήθηκαν για την τράπεζα αίματος, κ.λπ. Το κλινικό τμήμα καλύπτει οτιδήποτε μη εργαστηριακό. Συνεπώς, υπό την ίδια φιλοσοφία, σχετίζεται με ότι μπορεί να ελεγχθεί, μετρηθεί ή παρατηρηθεί σε έναν ασθενή χωρίς τη λήψη κάποιου δείγματος. Συμπεριλαμβάνει κωδικούς για την παρακολούθηση π.χ. των ζωτικών σημείων, του ηλεκτροκαρδιογραφήματος, υπέρηχου καρδιάς, ουρολογικών απεικονίσεων,

ακτινολογικών εξετάσεων/μελετών, κ.ο.κ. Ο σκοπός του LOINC είναι να δημιουργήσει μοναδικά/καθολικά αναγνωριστικά (ονόματα και κωδικούς) που θα χρησιμοποιούνται στα πλαίσια ήδη υπάρχοντων μηνυμάτων (HL7, CEN/TC 251, DICOM). Οι κωδικοί LOINC εμφανίζονται μέσα στα μηνύματα HL7 με το χαρακτηριστικό «LN». Όταν χρησιμοποιούνται μέσα σε μηνύματα επιτρέπουν την ανταλλαγή κλινικών εργαστηριακών δεδομένων μεταξύ ετερογενών υπολογιστικών συστημάτων. Για το λόγο αυτό απαιτείται το κάθε αναγνωριστικό να έχει ένα πλήρως καθορισμένο όνομα το οποίο θα δημιουργείται με συγκεκριμένο τρόπο. Έτσι οι χρήστες μπορούν να δημιουργούν ονόματα για τους ελέγχους/τις παρατηρήσεις τους, τα οποία μπορούν να συνδεθούν με τα μοναδικά/καθολικά αναγνωριστικά ελέγχου εκμεταλλευόμενοι ημι-αυτόματες μεθόδους. Σε κάθε όνομα αντιστοιχεί ένας κωδικός (LOINC code) που χρησιμοποιείται για να αναγνωριστούν τα αποτελέσματα των ελέγχων/παρατηρήσεων στις ηλεκτρονικές αναφορές. Το πλήρες όνομα μιας κλινικής παρατήρησης αποτελείται από έξι ή επτά μέρη: το όνομα του μέρους που μετράτε (π.χ. γλυκόζη, προπρανολόλη), το χαρακτηριστικό που παρατηρείται (π.χ. η συγκέντρωση της ουσίας, η μάζα, ο όγκος), η χρονική στιγμή της μέτρησης (π.χ. στιγμιαία ή όχι), το είδος του δείγματος (π.χ. ούρα, ορός), η κλίμακα μέτρησης (π.χ. ποιοτική ή ποσοτική), καθώς και η μέθοδος της μέτρησης (π.χ. ραδιοανοσοανάλυση). Αυτά περιγράφονται, σύμφωνα με τα ανωτέρω, με την ακόλουθη σύνταξη: :::: Στην ανωτέρω σύνταξη ο χαρακτήρας «:», είναι μέρος του ονόματος και χρησιμοποιείται για να διαχωρίσει τα κύρια μέρη του ονόματος. Το πρώτο μέρος μπορεί να διαχωριστεί επιπλέον σε έως και τρία υπο-τμήματα τα οποία διαχωρίζονται από τον χαρακτήρα «^». Το πρώτο υπο-τμήμα μπορεί επίσης να περιέχει πολλά επίπεδα αυξημένης ταξινόμησης τα οποία διαχωρίζονται από μια τελεία «.»». Το τρίτο και το τέταρτο μέρος του ονόματος (time aspect και system/sample) μπορούν επίσης να έχουν ένα υπο-τμήμα.

2.2.3.5 HL7 v2/HL7 v3 (Health Level Seven Version 2/Version 3)

HL7 v2

Το HL7 v2 επικεντρώθηκε στην ανταλλαγή που βασίζεται σε μεταφορά μηνυμάτων και απαιτεί σημαντική προσαρμογή του ιστοτόπου, γεγονός που οδήγησε σε σημασιολογικές ασυνέπειες μεταξύ των υλοποιήσεων. Επίσης, επειδή το HL7 αναπτύχθηκε για πρώτη φορά το 1987 για να ταιριάζει με την τεχνολογία εκείνη την εποχή, είναι σε μια ξεπερασμένη μορφή διαφορετικών πεδίων που χωρίζονται από σύμβολα «|» και «^». Ο οργανισμός HL7 αναφέρει πως για το HL7 v2, τα μηνύματα εξελίχθηκαν χρησιμοποιώντας μια



προσέγγιση «από τη βάση προς την κορυφή» η οποία αντιμετώπισε μεμονωμένες ανάγκες μέσω μιας εξελισσόμενης μεθοδολογίας. Ωστόσο, δεν μπορούν να προσδιοριστούν αυτόματα σημασιολογικά συμπεράσματα των πληροφοριών. Τα μηνύματα HL7 v2 χρησιμοποιούν μια σύνταξη κωδικοποίησης μη αναγνώσιμη από τον άνθρωπο, που βασίζεται σε ειδικά σύμβολα. Κάθε τμήμα αποτελείται από πεδία που χωρίζονται από οριοθέτες. Οι προεπιλεγμένοι οριοθέτες είναι στην μορφή «|» για τον διαχωριστή πεδίου, στην μορφή «^» για τον διαχωριστή συνιστωσών και στην μορφή «&» για τον διαχωριστή υποσυστημάτων. Το σύμβολο «~» είναι ο προεπιλεγμένος διαχωριστής επανάληψης. Κάθε τμήμα αρχίζει με μια συμβολοσειρά 3 χαρακτήρων που προσδιορίζει τον τύπο τμήματος. Κάθε τμήμα του μηνύματος περιέχει μια συγκεκριμένη κατηγορία πληροφοριών. Κάθε μήνυμα έχει ένα πεδίο που προσδιορίζει τον τύπο του μηνύματος ως το πρώτο του τμήμα (MSH). Ο τύπος μηνύματος καθορίζει τους τύπους αναμενόμενων τομέων στο μήνυμα. Οι τύποι τομέων που χρησιμοποιούνται σε συγκεκριμένο τύπο μηνύματος καθορίζονται από τη γραμματική σημειογραφία που χρησιμοποιείται στα πρότυπα HL7. Τα μηνύματα HL7 v2 περιέχουν μια κεφαλίδα μηνύματος, που αποτελείται όχι μόνο από το πρώτο τμήμα (MSH) αλλά αποτελεσματικά από την ομάδα των τομέων «MSH [SFT] [EVN] [MSA] {[ERR]}».

HL7 v3

Για την αντιμετώπιση των δυσκολιών του HL7 v2, αναπτύχθηκε η έκδοση 3 του προτύπου HL7. Το πρότυπο αυτό προσφέρει ένα πλαίσιο για την έκφραση κλινικών δηλώσεων κατά σημασιολογικά συνεπή τρόπο. Τα μηνύματα HL7 v3 έχουν δύο στρώματα: ένα εξωτερικό μήνυμα μετάδοσης (το ισοδύναμο του MSH [SFT] [MSA] {[ERR]}, όπου το ERR περιορίζεται στα σφάλματα που σχετίζονται με την αποδοχή) και ένα Trigger Event Control Act Wrapper ισοδύναμο του EVN {[ERR]}, όπου το ERR περιορίζεται σε σφάλματα που σχετίζονται με την Αίτηση Αναγνώρισης. Το πρότυπο αυτό ενσωματώνει ένα νέο πρότυπο για την αναπαράσταση πληροφοριών και την ανταλλαγή μηνυμάτων σε σύγκριση με το HL7 v2. Αυτή η νεότερη έκδοση προσανατολίζεται στη σημασιολογική διαλειτουργικότητα μεταξύ συστημάτων και είναι συνεπής με τις τεχνολογικές εξελίξεις, το σκοπό και τις συστάσεις που προκύπτουν από τον Σημασιολογικό Ιστό. Ως εκ τούτου, η εφαρμογή του HL7 v3 μπορεί να θεωρηθεί ως πλατφόρμα για την εφαρμογή σύγχρονων προτύπων κατανεμημένων συστημάτων, όχι μόνο στην ηλεκτρονική υγεία, αλλά και ως πλατφόρμα για να πειραματιστεί με αναδυόμενα πρότυπα κατανεμημένων συστημάτων γενικότερα. Έχει ως στόχο να υποστηρίξει όλες τις ροές εργασίας της υγειονομικής περίθαλψης. Η

ανάπτυξή του ξεκίνησε γύρω στο 1995, καταλήγοντας σε μια αρχική τυποποιημένη έκδοση το 2005. Το πρότυπο v3, σε αντίθεση με την έκδοση v2, βασίζεται σε μια τυπική μεθοδολογία και σε αντικειμενοστραφείς αρχές. Τα μηνύματα είναι ακόμη ανθρώπινα αναγνώσιμα όπως στην έκδοση v2, αλλά έχουν μορφή XML. Η προσέγγιση μοντελοποίησης HL7 v3 βασίζεται σε ένα τυποποιημένο μοντέλο αναφοράς πληροφοριών που αντιπροσωπεύει την κύρια επιχειρηματική λογική οποιουδήποτε περιβάλλοντος υγειονομικής περίθαλψης, από το οποίο μπορούν να οριστούν συγκεκριμένα μηνύματα και έγγραφα.

Τα βασικά χαρακτηριστικά του HL7 v3 είναι:

- Τα βασικά στοιχεία της επαναχρησιμοποίησης αναφορικά με τους α) τύπους των δεδομένων και β) τους τύπους στοιχείων μηνυμάτων.
- Η σύνταξη που χρησιμοποιείται είναι XML, η οποία από τη φύση της, οδηγεί σε πιο ανθρώπινα, και ευανάγνωστα μηνύματα.
- Όλα αντιμετωπίζονται ως ένα μοντέλο - χρησιμοποιούνται μέθοδοι μηχανικής οδήγησης (Model Driven Engineering - MDE) για την εξαγωγή XML Schemas.
- Υποστηρίζουν την ανταλλαγή υπηρεσιών «μηνυμάτων» καθώς και μιας προσέγγισης ηλεκτρονικού εγγράφου (CDA).
- Το πρότυπο δημιουργήθηκε για να υποστηρίξει όλες τις ροές εργασίας της υγειονομικής περίθαλψης, συμπεριλαμβανομένης (αλλά όχι μόνο) της ανταλλαγής δεδομένων μεταξύ οργανισμών. Έχει την ικανότητα να υποστηρίξει την ανταλλαγή ιατρικών δεδομένων μεγάλης πολυπλοκότητας (π.χ. για υποστήριξη αποφάσεων, EHRs, κλινική έρευνα).

Διαφορές HL7 v2 και HL7 v3

Σε κάποιο βαθμό, το HL7 v3 είναι το HL7 v2 της επόμενης γενιάς: βασίζεται σε επαναχρησιμοποιούμενες δομές, σε υποκείμενο ορισμό τύπου δεδομένων και χρησιμοποιεί ένα βιομηχανικό πρότυπο ως μέθοδο κωδικοποίησης σύνταξης. Το HL7 v3 (σε σύγκριση με το HL7 v2) έχει ωστόσο την ικανότητα να εκφράζει σύνθετα ιατρικά δεδομένα και είναι κατάλληλο για την ανταλλαγή δεδομένων μεταξύ οργανισμών. Οι στόχοι του HL7 v2 ήταν πολύ πιο μέτριοι από τους στόχους του HL7 v3, κάτι που το καθιστά επίσης ακατάλληλο για χρήση όταν κάποιος έχει την υποχρέωση να ανταλλάσσει πολύπλοκα ιατρικά δεδομένα.

Δεδομένου ότι το HL7 v3 σχεδιάστηκε για να συλλάβει επίσης πολύπλοκα κλινικά δεδομένα που δεν μπορούν να εκφραστούν σωστά σε HL7 v2, η δομή μηνύματος v3 επιτρέπει επίσης σχέσεις μεταξύ δραστηριοτήτων. Για παράδειγμα, υποθέτοντας ότι θέλουμε να μεταφέρουμε πως «Η Eve Everywoman υποβλήθηκε σε επείγουσα χειρουργική επέμβαση και της χορηγήθηκαν τα φάρμακα X και Y αφού διαγνώσθηκε με ρήξη των πνευμόνων που καθορίστηκε χρησιμοποιώντας υπερηχογράφημα», κάποιος απλά δεν μπορεί να εκφράσει τις διάφορες σχέσεις (π.χ. χρονική, αιτιώδης) μεταξύ των διαφόρων ιατρικών δραστηριοτήτων του HL7 v2. Το HL7 v3 έχει τη δυνατότητα να εκφράσει όλες αυτές τις σχέσεις, δημιουργώντας ένα γράφημα που αποτελείται από τμήματα. Επιπλέον, το HL7 v3 καθοδηγείται πλήρως από μοντέλα και όλα τα μοντέλα παράγονται (με αντικειμενοστραφή τρόπο, από ένα σχετικά αφηρημένο μοντέλο πληροφοριών αναφοράς (Reference Information Model - RIM). Αυτό το μοντέλο χρησιμεύει για την προώθηση της συνοχής μεταξύ των διαφόρων μηνυμάτων. Στους όρους HL7 v2 είναι σαν να έχει οριστεί ένα αφηρημένο/γενικό τμήμα γενικής παρατήρησης, με προκαθορισμένες ορολογίες. Έτσι, η ικανότητα αυτή οδηγεί σε μεγαλύτερη συνέπεια και χρήση κατάλληλα περιορισμένων μοντέλων για συγκεκριμένες περιπτώσεις χρήσης.

2.2.3.6 HL7 FHIR (Health Level Seven Fast Healthcare Interoperability)

Το FHIR είναι ένα πρότυπο που περιγράφει τις μορφές και τα στοιχεία δεδομένων (πόροι) και ένα API για την ανταλλαγή ηλεκτρονικών αρχείων υγείας. Το πρότυπο δημιουργήθηκε από την οργάνωση προτύπων υγειονομικής



περίθαλψης HL7. Το FHIR βασίζεται σε προηγούμενα πρότυπα δεδομένων μορφής από το HL7, όπως η v2 και η v3. Είναι όμως πιο εύκολη η υλοποίησή του, επειδή χρησιμοποιεί μια σύγχρονη τεχνολογία API βασισμένη στον ιστό, συμπεριλαμβανομένου ενός πρωτότυπου πρωτοκόλλου RESTful [105] που βασίζεται σε HTTP, Hypertext Markup Language (HTML) [106] και Cascading Style Sheets (CSS) [107] για ενσωμάτωση της διεπαφής χρήστη, μία επιλογή από JavaScript Object Notation (JSON) [108] ή XML για την αναπαράσταση δεδομένων και το Atom [109] για τα αποτελέσματα. Ένας από τους στόχους του είναι να διευκολύνει τη διαλειτουργικότητα των συστημάτων ιατρικής περίθαλψης παλαιού τύπου, να διευκολύνει την παροχή πληροφοριών για την υγειονομική περίθαλψη στους παρόχους υγειονομικής περίθαλψης και τα άτομα σε μια ευρεία ποικιλία συσκευών από υπολογιστές έως κινητά τηλέφωνα, και να επιτρέπει την εφαρμογή τρίτων προγραμματιστών να παρέχουν

ιατρικές εφαρμογές οι οποίες μπορούν εύκολα να ενσωματωθούν σε υπάρχοντα συστήματα. Επειδή το FHIR εφαρμόζεται πάνω από το πρωτόκολλο HL7 και το πρωτόκολλο HTTPS (HTTP Secure) [110], τα μηνύματα μπορούν να αναλυθούν μέσω πλατφόρμων ανάλυσης δεδομένων για συλλογή δεδομένων σε πραγματικό χρόνο. Σε αυτή την έννοια, οι οργανισμοί υγειονομικής περίθαλψης είναι σε θέση να συλλέγουν δεδομένα σε πραγματικό χρόνο από συγκεκριμένα τμήματα σε μηνύματα FHIR καθώς αυτά τα μηνύματα περνούν από το δίκτυο. Αυτά τα δεδομένα μπορούν να μεταδοθούν σε ένα χώρο αποθήκευσης δεδομένων όπου μπορούν να συσχετιστούν με άλλα δεδομένα πληροφορικής. Οι πιθανές περιπτώσεις χρήσης περιλαμβάνουν επιδημιολογική παρακολούθηση, παρακολούθηση συνταγής φαρμάκων, προειδοποιήσεις αλληλεπίδρασης με ανεπιθύμητα φάρμακα και χρόνους αναμονής για το δωμάτιο έκτακτης ανάγκης.

Το FHIR τεχνικά αποτελείται από ένα καλό υπόβαθρο, το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- Είναι ένας πιο λεπτομερής τρόπος ανταλλαγής δεδομένων χωρίς την άκαμπτη ροή εργασίας του παραδοσιακού HL7.
- Δεν επιβαρύνεται το SOAP αφού χρησιμοποιείται μια απλή προσέγγιση RESTful style.
- Δίνεται μεγάλη έμφαση στην εφαρμογή συμμόρφωσης και αναφοράς ως μέρος της βασικής διαδικασίας.
- Επικεντρώνεται στο να χτυπήσει το 80% των κοινών περιπτώσεων χρήσης αντί του 20% των εξαιρέσεων.
- Το FHIR συνδυάζει τα καλύτερα χαρακτηριστικά των HL7 v2, HL7 v3 και CDA, αξιοποιώντας τις πιο σύγχρονες τεχνολογίες υπηρεσιών ιστού.
- Ο σχεδιασμός του FHIR βασίζεται σε υπηρεσίες RESTful web. Αυτό έρχεται σε αντίθεση με την πλειονότητα των προφίλ EHR που βασίζονται σε υπηρεσίες ιστού SOAP. Με τις υπηρεσίες RESTful web, ενσωματώνονται οι βασικές λειτουργίες HTTP, όπως η δημιουργία, η ανάγνωση, η ενημέρωση και η διαγραφή.

- Το FHIR βασίζεται σε αρθρωτά στοιχεία που ονομάζονται «πόροι» και οι πόροι αυτοί μπορούν να συνδυαστούν μεταξύ τους για την επίλυση κλινικών και διοικητικών προβλημάτων με πρακτικό τρόπο. Οι πόροι μπορούν να επεκταθούν και να προσαρμοστούν ώστε να παρέχουν μια πιο εύχρηστη λύση στη ζήτηση της υγειονομικής περίθαλψης για επιλογή και προσαρμογή. Τα συστήματα μπορούν εύκολα να διαβάσουν τις επεκτάσεις χρησιμοποιώντας το ίδιο πλαίσιο με άλλους πόρους.

Το FHIR, όπως αναφέρθηκε, αντιπροσωπεύει τα κλινικά δεδομένα ως πόρους, όπου κάθε πόρος είναι μια συνεκτική έκφραση νοήματος που δηλώνεται με όρους καλά καθορισμένων πεδίων και τύπων δεδομένων. Οι ορισμοί των κλινικών πόρων του FHIR είναι συγκεκριμένες και διαισθητικές έννοιες, όπως η συνταγογράφηση φαρμάκων, η ανεπιθύμητη αντίδραση, η διαδικασία και η κατάσταση. Αυτοί οι πόροι αποτελούν ένα γράφημα κλινικών δεδομένων με ρητές αναφορές μεταξύ πόρων. Για παράδειγμα, ένας πόρος MedicationPrescription αναφέρει ρητά τον συνταγογράφο του (FHIR Practitioner), τον ασθενή του (FHIR patient) και το φάρμακο που έχει συνταγογραφηθεί (FHIR Medication). Το FHIR δεν περιλαμβάνει λεπτομερή μοντέλα για κάθε πτυχή ενός κλινικού αρχείου, αλλά παρέχει έναν ενσωματωμένο μηχανισμό επεκτασιμότητας για τον εμπλουτισμό των υφιστάμενων ορισμών πόρων. Οι ορισμοί των πόρων FHIR δεν υπόσχονται άμεσα σημασιολογικά συνεπή δεδομένα εκτός πλαισίου. Αντίθετα, για να εξυπηρετήσουν ξεχωριστά περιβάλλοντα (π.χ. EHRs, ροές εργασίας υποβολής αναφορών δημόσιας υγείας, φορητές συσκευές), οι πόροι FHIR ενδέχεται να χρησιμοποιούν ειδικά φορτία δεδομένων με διαφορετικές ορολογίες. Η σημασιολογική συνοχή βασίζεται σε ένα επίπεδο αφαίρεσης που ονομάζεται προφίλ FHIR που περιορίζει και επεκτείνει τους ορισμούς πόρων σε συγκεκριμένα περιβάλλοντα. Η δημιουργία προφίλ μπορεί να επιτρέψει τη διαλειτουργικότητα για την υποκατάσταση εφαρμογών, καλύπτοντας τις απαιτήσεις, καθορίζοντας τα συστήματα κωδικοποίησης και επιβάλλοντας συμβάσεις σχετικά με την εκπροσώπηση των δεδομένων στους πόρους. Το API FHIR είναι μια σύγχρονη διεπαφή HTTP προσανατολισμένη στις πηγές για την αναζήτηση, δημιουργία, ανάγνωση, ενημέρωση και διαγραφή των πόρων FHIR που αντιπροσωπεύουν δεδομένα κλινικών, διοικητικών (π.χ. ορισμοί περιορισμού και δομές ομαδοποίησης) ερωτημάτων σε επίπεδο πληθυσμού.

Οφέλη του FHIR

Γενικότερα, μια κεντρική πρόκληση για τα πρότυπα υγειονομικής περίθαλψης είναι ο τρόπος αντιμετώπισης της μεταβλητότητας που προκαλείται από διάφορες διαδικασίες υγειονομικής περίθαλψης. Με την πάροδο του χρόνου, προστίθενται περισσότερα πεδία και προτεραιότητες στις προδιαγραφές,

προσθέτοντας σταδιακά το κόστος και την πολυπλοκότητα στις εφαρμοζόμενες υλοποιήσεις. Η εναλλακτική λύση βασίζεται σε προσαρμοσμένες επεκτάσεις, αλλά αυτές δημιουργούν και πολλά προβλήματα εφαρμογής. Το FHIR επιλύει αυτή την πρόκληση καθορίζοντας ένα απλό πλαίσιο για την επέκταση και προσαρμογή των υφιστάμενων πόρων. Όλα τα συστήματα, ανεξάρτητα από τον τρόπο με τον οποίο αναπτύσσονται, μπορούν εύκολα να διαβάσουν αυτές τις επεκτάσεις και οι ορισμοί επέκτασης μπορούν να ανακτηθούν χρησιμοποιώντας το ίδιο πλαίσιο με την ανάκτηση άλλων πόρων.

Μεταξύ άλλων, το FHIR προσφέρει πολλές βελτιώσεις σε σχέση με τα υπάρχοντα πρότυπα:

- Μια ισχυρή εστίαση στην εφαρμογή - γρήγορη και εύκολη υλοποίηση
- Πολλαπλές βιβλιοθήκες υλοποίησης, και πολλά παραδείγματα διαθέσιμα για την εκκίνηση ανάπτυξης
- Είναι δωρεάν για χρήση χωρίς περιορισμούς
- Οι πόροι βάσης εκτός πλαισίου διαλειτουργικότητας μπορούν να χρησιμοποιηθούν ως έχουν, αλλά μπορούν επίσης να προσαρμοστούν για τοπικές απαιτήσεις
- Η αναπτυξιακή πορεία εξέλιξης από τα πρότυπα HL7 v2 και CDA μπορεί να συνυπάρχουν και να αλληλοσυμπληρώνονται
- Διαθέτει μία ισχυρή βάση στα πρότυπα XML, JSON, HTTP, OAuth [111] κ.λπ.
- Διαθέτει υποστήριξη για RESTful αρχιτεκτονικές και επίσης απρόσκοπτη ανταλλαγή πληροφοριών με τη χρήση μηνυμάτων ή εγγράφων
- Διαθέτει συνοπτικές και εύκολα κατανοητές προδιαγραφές
- Αποτελεί μια μορφή μηνυμάτων αναγνώσιμων από τον άνθρωπο, για εύκολη χρήση από προγραμματιστές
- Κάνει χρήση οντολογιών με αυστηρή και επίσημη αντιστοίχιση για ορθότητα

2.2.4 Τεχνικές Διαλειτουργικότητας στον τομέα της Υγείας

Τα τελευταία χρόνια, πολλά έργα έχουν επιδιώξει τη διαλειτουργικότητα των πληροφοριακών συστημάτων αναφορικά με τα δεδομένα υγείας. Οι διαφορετικές προσεγγίσεις έχουν προτείνει λύσεις βασισμένες σε συγκεκριμένα πρότυπα και τεχνολογίες προκειμένου να ικανοποιηθούν οι ανάγκες ενός συγκεκριμένου σεναρίου, αλλά μέχρι στιγμής δεν έχουν δοθεί καθολικά πλαίσια διαλειτουργικότητας, όπως περιγράφονται στη συνέχεια. Η πρώτη ομάδα προσεγγίσεων χρησιμοποιεί τεχνολογίες XML για την επίτευξη διαλειτουργικότητας. Ένα παράδειγμα είναι η επίτευξη διαλειτουργικότητας μεταξύ ενός προωθητικού συστήματος και του HL7 CDA που διεξάγεται στην Ταιβάν [112]. Σε αυτή τη λύση, και τα δύο μοντέλα αναπαράστασης αναπαρίστανται σε XML και οι αντιστοιχίσεις τους ορίζονται χρησιμοποιώντας τους κανόνες eXtensible Stylesheet Language Transformations (XSLT) [113]. Η δεύτερη ομάδα προσεγγίσεων χρησιμοποιεί την XML και την Web Ontology Language (OWL) [114], αν και με διαφορετικούς σκοπούς. Στο [115], επιδιώχθηκε η διαλειτουργικότητα μεταξύ των διαφόρων συστημάτων υγειονομικής περίθαλψης που συμμορφώνονται με διαφορετικά πρότυπα, επισημαίνοντας τα μηνύματα της υπηρεσίας Web μέσω των αρχέτυπων που ορίζονται στο OWL. Σε μια τέτοια προσέγγιση, οι αντιστοιχίσεις μεταξύ αυτών των αρχέτυπων ορίζονται χειροκίνητα χρησιμοποιώντας ένα γραφικό εργαλείο. Στη συνέχεια, αυτές οι αντιστοιχίσεις εφαρμόζονται σε περιπτώσεις κλινικών δεδομένων XML. Η ίδια ερευνητική ομάδα παρουσίασε στο [116] μια προσέγγιση βασισμένη σε αρχέτυπα, οντολογίες και σημασιολογικές τεχνικές για τη διαλειτουργικότητα μεταξύ των συστημάτων HL7 CDA και ISO 13606 [117]. Αυτή η έρευνα μοιράζεται με τα προηγούμενα αποτελέσματα όπου τα αρχέτυπα αντιπροσωπεύονται σε OWL, αλλά προσεγγίζουν το πρόβλημα χρησιμοποιώντας το HL7 RIM ως βασικό και κοινό μοντέλο πληροφοριών. Κατά συνέπεια, αυτό το μοντέλο πληροφοριών υποστηρίζει τον ημιαυτόματο ορισμό των αντιστοιχιών μεταξύ των προτύπων. Οι αντιστοιχίσεις ορίζονται μεταξύ αρχέτυπων συγκρίνοντας κατά πόσο οι δομές που χρησιμοποιούνται σε αμφότερα τα πρότυπα είναι εξειδικεύσεις της ίδιας κατηγορίας του HL7 RIM χρησιμοποιώντας μηχανισμούς συλλογισμού. Οι προκύπτουσες αντιστοιχίσεις εφαρμόζονται σε κλινικά δεδομένα XML, τα οποία μετατρέπονται σε OWL και στη συνέχεια μετατρέπονται ξανά σε XML. Μια άλλη έρευνα διαλειτουργικότητας είναι αυτή μεταξύ του openEHR και του συστήματος COSMIC Cambio [118]. Σε αυτήν την έρευνα προσδιορίστηκαν τα δύο μοντέλα πληροφοριών. Η λύση αυτή, όπως αυτή που διεξάγεται στην Ταιβάν, είναι προσανατολισμένη στην επίλυση του συγκεκριμένου προβλήματος διαλειτουργικότητας των δύο αυτών συστημάτων και ως εκ

τούτου είναι δύσκολο να προσαρμοστεί στη χρήση της με άλλα κλινικά μοντέλα ή πρότυπα. Δεδομένης της σημασίας της διαλειτουργικότητας, οι κύριοι οργανισμοί κλινικής τυποποίησης, ISO, HL7 και openEHR, καταβάλλουν προσπάθειες για την εναρμόνιση των προδιαγραφών τους. Αυτό αποδεικνύεται από το τρίτο μέρος του προτύπου ISO 13606 που ονομάζεται «Αρχέτυπα αναφοράς και Κατάλογοι όρων», το οποίο παρέχει έναν ενημερωτικό οδηγό για να αντιπροσωπεύει τις κλινικές πληροφορίες που κωδικοποιούνται σύμφωνα με το HL7 και το openEHR χρησιμοποιώντας τις δομές ISO 13606. Επιπλέον, η προδιαγραφή αναφέρει ότι, προκειμένου να επιτευχθεί πλήρης σημασιολογική διαλειτουργικότητα συστημάτων που βασίζονται σε αυτούς τους κανόνες, είναι απαραίτητη η εναρμόνιση του λεξιλογίου και των τύπων δεδομένων που χρησιμοποιούνται. Προκειμένου να εναρμονιστούν οι τύποι δεδομένων, αναπτύχθηκε το πρότυπο ISO 21090 [119]. Επιπροσθέτως, τα Αναλυτικά Κλινικά Μοντέλα (Detailed Clinical Models – DCM) [120] έχουν πρόσφατα οριστεί ως μια πρωτοβουλία για τον ορισμό των κλινικών πληροφοριών ανεξάρτητα από ένα συγκεκριμένο κλινικό πρότυπο αλλά με σκοπό να προσφέρεται η δυνατότητα μετασχηματισμού σε άλλα πρότυπα. Ωστόσο, παρά την ανάπτυξη των ιατρικών προτύπων, η διαλειτουργικότητα και η ενσωμάτωση δεδομένων εξακολουθούν να είναι ένα ανοικτό πρόβλημα. Το [121] αποτελεί παράδειγμα ερευνητικών προσπαθειών που αφορούν την ενσωμάτωση των ιατρικών προτύπων προκειμένου να διασφαλιστεί η διαλειτουργικότητα μεταξύ των προτύπων HL7 και IEEE 1451 κατά την παρακολούθηση των δεδομένων ασθενών. Οι συγγραφείς στο [122] πρότειναν μια οντολογία για να περιγράψουν τα ζωτικά σημεία του ασθενούς και να επιτρέψουν τη σημασιολογική διαλειτουργικότητα κατά την παρακολούθηση των δεδομένων των ασθενών. Στην ίδια κατεύθυνση, το [123] πρότεινε μια διαδραστική υγειονομική περίθαλψη που βασίζεται σε οντολογίες με αισθητήρες που μπορούν να φορεθούν, για να αποκτηθούν πληροφορίες περιεχομένου σε πραγματικό χρόνο χρησιμοποιώντας οντολογικές μεθόδους, ενσωματώνοντας εξωτερικά δεδομένα για την πρόληψη ασθενειών. Επιπλέον, στο [124] παρουσιάζεται ένα πλαίσιο για την ημιαυτόματη ενσωμάτωση και σύνδεση των ετερογενών συνόλων δεδομένων δημιουργώντας παραστάσεις OWL αυτών των συνόλων δεδομένων που μπορούν να χρησιμοποιηθούν για επερωτήσεις. Επιπλέον, το [125] προτείνει μια Αρχιτεκτονική Τεχνολογίας Συστημάτων Πληροφορικής (Cloud Health Information Systems Technology Architecture - CHISTAR) που μπορεί να επιτύχει διαφορετικούς τύπους διαλειτουργικότητας με μια μεθοδολογία γενικού σχεδιασμού που χρησιμοποιεί ένα μοντέλο αναφοράς που ορίζει το συνηθισμένο σύνολο δομών δεδομένων και ένα μοντέλο αρχέτυπου που καθορίζει τη λογική κλινικών δεδομένων. Θα πρέπει να σημειωθεί το έργο του [126], όπου το μοντέλο NIST

Big Data εμπλουτίστηκε με σημασιολογία για να κατανοήσει έξυπνα τα συλλεχθέντα δεδομένα και να παράγει πολύτιμες πληροφορίες με τη συσχέτιση διάσπαρτων ιατρικών δεδομένων που προέρχονται από πολλαπλές φορητές συσκευές.

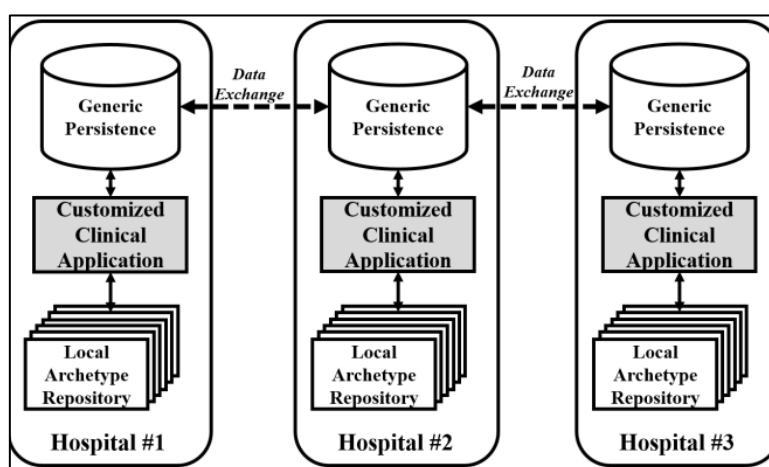
Συνοπτικά, κάθε συγκεκριμένη λύση βοηθά στην παροχή πρόσβασης σε κλινικές πληροφορίες ασθενών για ορισμένους κλινικούς οργανισμούς. Ένα από τα μειονεκτήματα αυτών των προσεγγίσεων είναι ότι δεν προτείνουν πλαίσια και μεθόδους εύκολο να εφαρμοστούν σε διαφορετικά πρότυπα. Από τη μια πλευρά, οι τεχνολογίες XML παρέχουν ένα περιορισμένο σημασιολογικό μοντέλο το οποίο δεν φαίνεται να είναι η καλύτερη λύση για την επίτευξη του στόχου της διαλειτουργικότητας, αφού για παράδειγμα η ενσωμάτωση των κλινικών δεδομένων με ορολογίες θα ήταν δύσκολη. Ειδικότερα, η χρήση του XSLT για τον καθορισμό των κανόνων αντιστοίχισης δεν αποτελεί πλέον την καλύτερη διαθέσιμη επιλογή, δεδομένης της διαθεσιμότητας των γλωσσών αντιστοίχισης και μετασχηματισμού που δημιουργούνται από την κοινότητα και οι οποίες έχουν καλύτερες ιδιότητες από πλευράς συντήρησης, επαναχρησιμοποίησης και υποστήριξης στην ανάπτυξη λογισμικού διαδικασίες. Από την άλλη πλευρά, η χρήση OWL και οντολογιών για την υποστήριξη τέτοιων διαδικασιών διαλειτουργικότητας είναι μια καλή πρακτική που συνιστάται, έτσι ώστε οι παγκόσμιες λύσεις να χρησιμοποιούν αυτή την τεχνολογία.

2.2.4.1 Κομβικές Τεχνικές Διαλειτουργικότητας στον τομέα της Υγείας

Σύμφωνα με τις προαναφερθείσες τεχνικές και προσεγγίσεις, πολλαπλές έρευνες προς την επίτευξη διαλειτουργικότητας δεδομένων σε διαφορετικούς τομείς έχουν πραγματοποιηθεί. Ορισμένες από τις πιο κομβικές έρευνες, οι οποίες επηρέασαν την προτεινόμενη Προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων, περιγράφονται στη συνέχεια.

Ποιότητα Δεδομένων για Σημασιολογικά Διαλειτουργικά EHRs (Data Quality for Semantic Interoperable Electronic Health Records) [127]

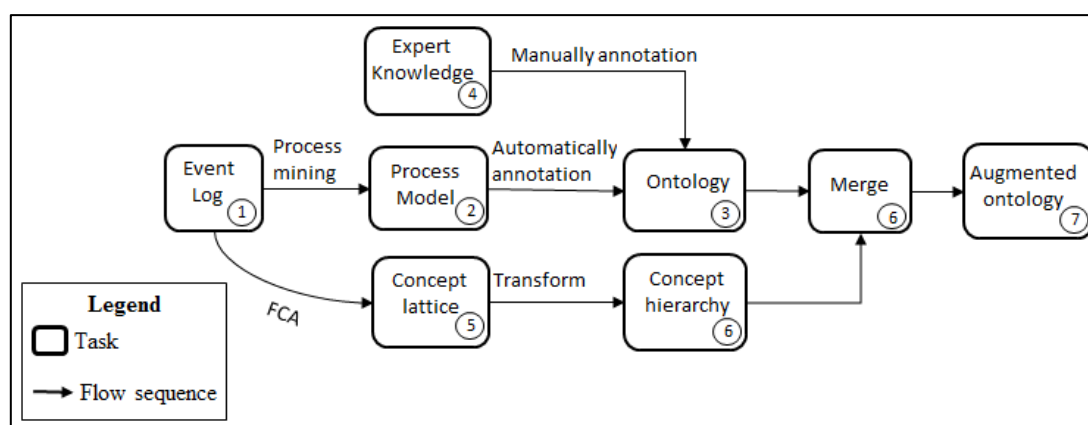
Η συγκεκριμένη έρευνα προτείνει να χρησιμοποιηθεί η εφαρμογή που βασίζεται σε αρχέτυπα (τα οποία καθορίζονται μέσω της Archetype Description Language (ADL) [128]) που ορίζεται από το πρότυπο openEHR, για να διατηρούνται τα δεδομένα σε μια γενική μορφή (Εικόνα 2-5). Κάθε νοσοκομείο μπορεί να κατεβάσει οποιοδήποτε αριθμό αρχέτυπων βάσει των τοπικών του απαιτήσεων και να τα αποθηκεύσει σε ένα τοπικό αποθετήριο αρχέτυπων. Το αποθετήριο αρχέτυπων διαφόρων νοσοκομείων μπορεί να έχει ξεχωριστό σύνολο αρχέτυπων, ακριβώς το ίδιο σύνολο αρχέτυπων ή επικαλυπτόμενο σύνολο αρχέτυπων ανάλογα με τις τοπικές τους απαιτήσεις. Με βάση τον τοπικό χώρο αποθήκευσης αρχέτυπων, κατασκευάζεται μια προσαρμοσμένη κλινική εφαρμογή για το αντίστοιχο νοσοκομείο. Έτσι, τα δεδομένα που συλλέγονται από την τοπική εφαρμογή αποθηκεύονται σε μια γενική μορφή που προτείνεται στην παρούσα έρευνα. Το γενικό σχήμα επιτρέπει τη μετακίνηση δεδομένων από διάφορους οργανισμούς χωρίς να πραγματοποιούνται αλλαγές στο επίπεδο σχήματος. Επιπλέον, το προτεινόμενο γενικό σχήμα μπορεί να επεκταθεί για να ενσωματώσει έναν μηχανισμό ασφαλείας που με τη σειρά του βοηθά στη βελτίωση της ποιότητας των δεδομένων.



Εικόνα 2-5 - Ποιότητα δεδομένων για σημασιολογικά διαλειτουργικά EHRs

Ενίσχυση Σημασιολογικής Διαλειτουργικότητας στην Υγειονομική Περιθαλψη με τη χρήση Εξόρυξης Σημασιολογικών Διαδικασιών (Enhancing Semantic Interoperability in Healthcare using Semantic Process Mining) [129]

Η βασική ιδέα της έρευνας αυτής είναι η εξόρυξη σημασιολογικών διαδικασιών σε μορφή οντολογιών, όπου η ενέργεια αυτή θα επιτρέψει στη μηχανή συμπερασμάτων να αντλήσει νέες γνώσεις. Ο συνδυασμός της σημασιολογίας και των διαδικασιών μπορεί να βοηθήσει στην ανταλλαγή πληροφοριών μεταξύ των εφαρμογών με τον πλέον σωστό και πλήρη τρόπο ή / και στην αναδιάρθρωση των επιχειρηματικών διαδικασιών παρέχοντας ένα εργαλείο για την εξέταση της αντιστοιχίας των οντολογιών διεργασιών. Η προτεινόμενη μέθοδος για τον σημασιολογικό εμπλουτισμό του αρχείου καταγραφής συμβάντων στις οντολογίες χρησιμοποιώντας την Ανάλυση Τυπικής Αντίληψης, παρουσιάζεται στην Εικόνα 2-6. Το «Βήμα 1» σχετίζεται με τη σύλληψη του αρχείου καταγραφής συμβάντων, το οποίο πρέπει να περιέχει τις πληροφορίες σχετικά με τις εκτελέσεις της διαδικασίας. Οι τεχνικές εξόρυξης διεργασιών θα χρησιμοποιηθούν για την απόκτηση του μοντέλου διαδικασίας στο «Βήμα 2». Το μοντέλο της διαδικασίας παρέχει γνώσεις σχετικά με τον τρόπο με τον οποίο συνδέονται οι δραστηριότητες, με το ποιο εκτελούσαν τις δραστηριότητες, το κοινωνικό δίκτυο, το χρόνο εκτέλεσης και άλλα.

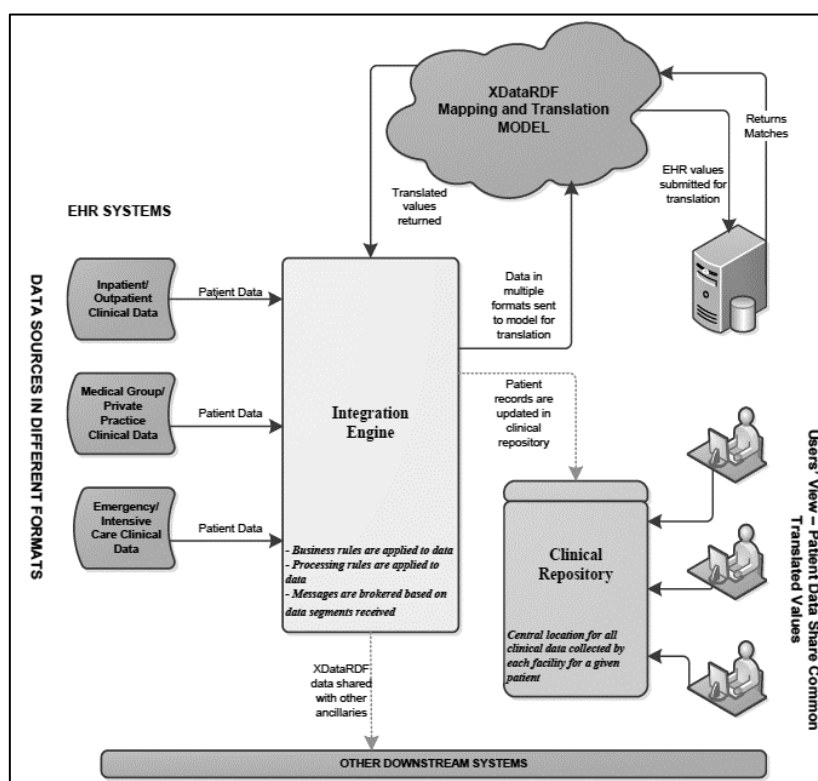


Εικόνα 2-6 - Ενίσχυση της σημασιολογικής διαλειτουργικότητας στην υγειονομική περιθαλψη με τη χρήση εξόρυξης σημασιολογικών διαδικασιών

Προς ένα Εννοιολογικό Πλαίσιο για Συνεχή Χρήση: Ένα Τεχνικό Σχέδιο για την Επίτευξη Σημασιολογικής Διαλειτουργικότητας στα Ηλεκτρονικά Συστήματα Καταγραφής Υγείας (Towards a Conceptual Framework for Persistent Use: A Technical Plan to Achieve Semantic Interoperability within Electronic Health Record Systems) [130]

Η προτεινόμενη λύση ενσωματώνει το RDF ως θεμέλιο για την επίτευξη σημασιολογικής διαλειτουργικότητας. Το προτεινόμενο μοντέλο, σε αντίθεση

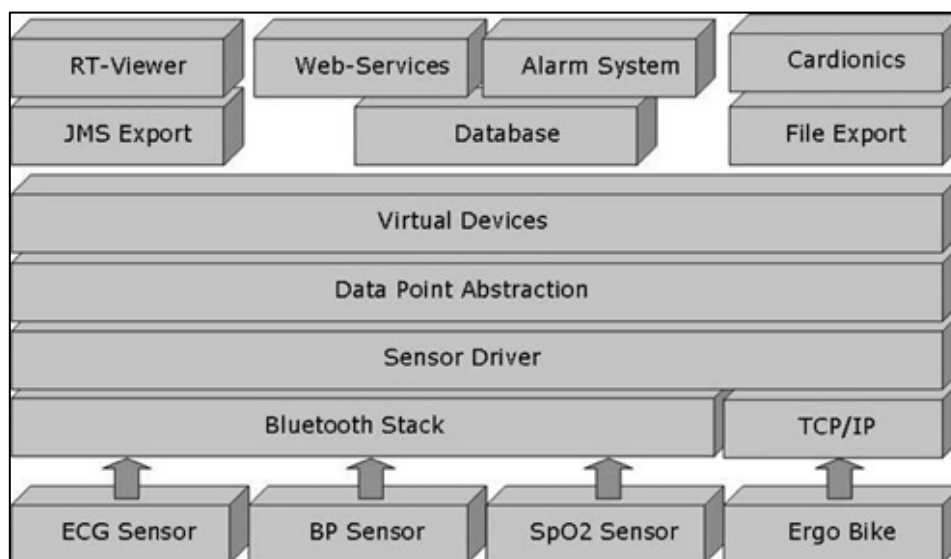
με τις προηγούμενες λύσεις, παρέχει ένα ολοκληρωμένο πακέτο για τα συστήματα υγείας για την επίτευξη πραγματικής διαλειτουργικότητας. Η εφαρμογή του Resource Description Framework (RDF) [131] για την επίτευξη της διαλειτουργικότητας, επιτρέποντας τον εύκολο συνδυασμό και την αλληλεπίδραση πολλαπλών μοντέλων δεδομένων και λεξιλογίων μέσα σε ένα ενιαίο περιβάλλον υγείας, μειώνοντας έτσι τις πιθανότητες ασάφειας των δεδομένων. Η ακρίβεια των δεδομένων και η συνέχεια των απεικονίσεων παρέχουν τα δομικά στοιχεία της σημασιολογικής διαλειτουργικότητας. Αυτοί οι παράγοντες είναι εμφανείς στον πυρήνα του προτύπου RDF. Η ακρίβεια δεν αναφέρεται μόνο στα πρωτογενή δεδομένα αλλά περιλαμβάνει επίσης τη συμμόρφωση με τους νόμους που ισχύουν για την επίτευξη σημασιολογικής διαλειτουργικότητας των δεδομένων της υγειονομικής περίθαλψης. Η συνέχιση της αντιστοίχισης των δεδομένων αναφέρεται στην ικανότητα ενσωμάτωσης οποιωνδήποτε αλλαγών που συμβαίνουν σε ένα πρότυπο με την πάροδο του χρόνου, ως αποτέλεσμα των ενημερώσεων στα πρότυπα ή στην ομοσπονδιακή εντολή και αντανακλώντας αυτές τις αλλαγές στη αντιστοίχιση. Συνολικά, η χρήση του RDF στο προτεινόμενο μοντέλο μετάφρασης θα εξασφαλίσει ότι η εγκυρότητα των αντιστοιχημένων δεδομένων θα ικανοποιεί το επίπεδο ακρίβειας που απαιτείται για τη μετατροπή των διαφόρων προτύπων υγειονομικής περίθαλψης στο περιβάλλον αυτό, προωθώντας έτσι τη σημασιολογική διαλειτουργικότητα (Εικόνα 2-7).



Εικόνα 2-7 - Ένα τεχνικό σχέδιο για την επίτευξη της σημασιολογικής διαλειτουργικότητας στα ηλεκτρονικά συστήματα καταγραφής της υγείας

SAPHIRE - έξυπνη παρακολούθηση της υγειονομικής περιθαλψης με βάση μια πλατφόρμα σημασιολογικής διαλειτουργικότητας (SAPHIRE - intelligent healthcare monitoring based on semantic interoperability platform) [132]

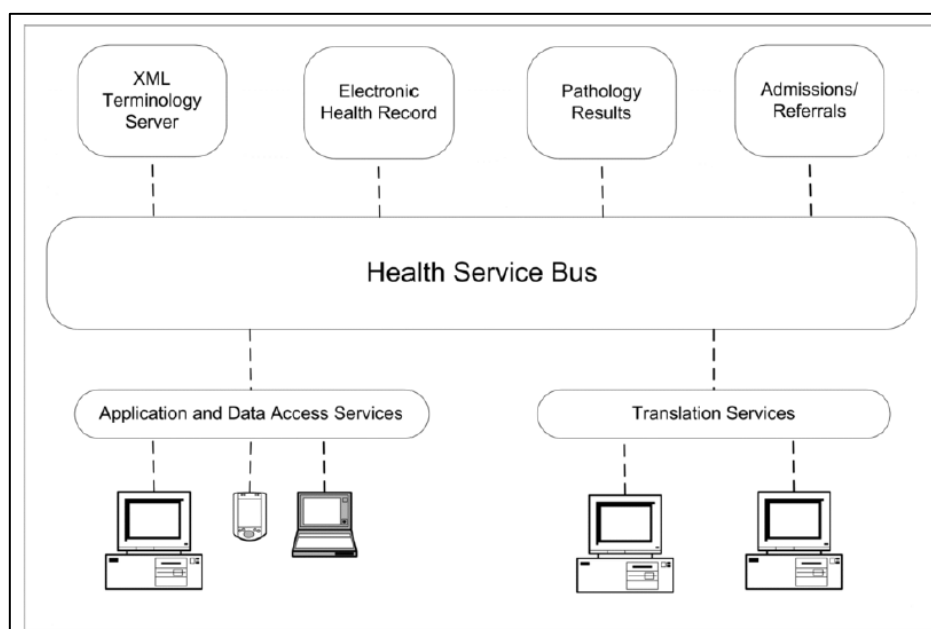
Σε αυτή την αρχιτεκτονική (Εικόνα 2-8), τα δεδομένα υγείας αποθηκεύονται σε τοπικά αποθετήρια EHR. Ωστόσο, είναι επίσης εγγεγραμμένα σε ένα μητρώο μαζί με ένα σύνολο μεταδεδομένων, έτσι ώστε τα δεδομένα να μπορούν στη συνέχεια να εντοπιστούν και να έχουν πρόσβαση όποτε αποθηκεύονται. Μέσω αυτής της αρχιτεκτονικής, επιλύεται η ανακάλυψη και η πρόσβαση στα σχετικά δεδομένα των EHRs. Ωστόσο, για να μπορέσει ο ενδιαφερόμενος να εκμεταλλευτεί τα δεδομένα που είναι αποθηκευμένα, πρέπει να εκπροσωπούνται σε μορφή επεξεργάσιμη από μηχανή. Στο SAPHIRE, χρησιμοποιούνται τα έγγραφα σε μορφή HL7 CDA όπου και οι καταχωρήσεις και οι ενότητες σχολιάζονται μέσω σχημάτων κωδικοποίησης, αντιστοιχες με αυτές του HL7 v3.



Εικόνα 2-8 - SAPHIRE - έξυπνη παρακολούθηση της υγειονομικής περιθαλψης με βάση μια πλατφόρμα σημασιολογικής διαλειτουργικότητας

Ένα Πλαίσιο για Σημασιολογική Διαλειτουργικότητα στην Υγειονομική Περιθαλψη: Αρχιτεκτονική με βάση τα Πρότυπα Υγείας (A Framework for Semantic Interoperability in Healthcare: A Service Oriented Architecture based on Health Informatics Standards) [133]

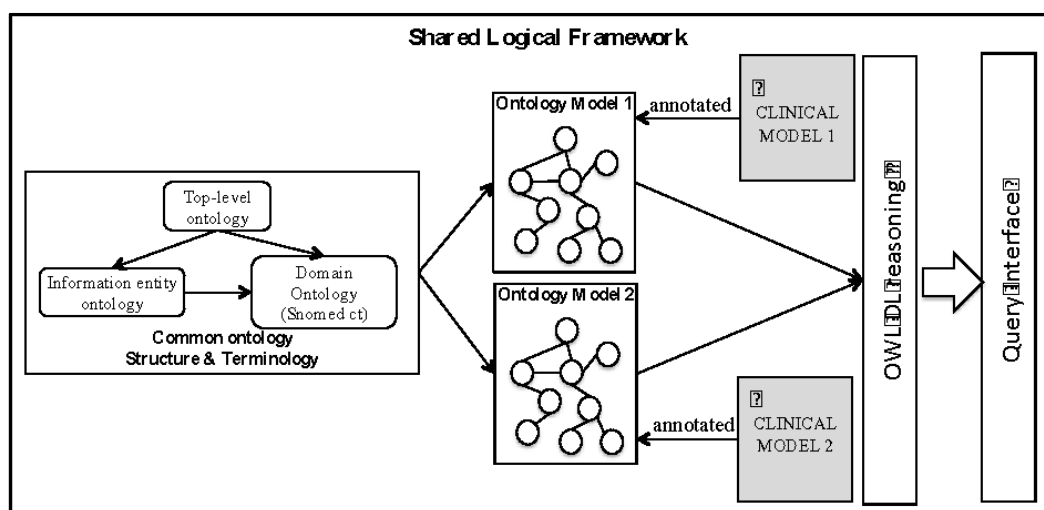
Η υποδομή αυτή βασίστηκε στην ιδεολογία του Enterprise Service Bus (ESB) [134], που ονομάζεται Health Service Bus (HSB). Ένα διάγραμμα επισκόπησης της αρχιτεκτονικής HSB παρουσιάζεται στην Εικόνα 2-9. Το HSB παρέχει επικοινωνία μεταξύ διαφορετικών συστημάτων υγείας, στα οποία μπορούν όλοι να συνδεθούν με αυτό, ανεξάρτητα από τον τύπο λογισμικού ή υλικού που χρησιμοποιείται. Το HSB επίσης παρέχει υπηρεσίες όπως είναι ο διακομιστής ορολογίας και υπηρεσίες μετάφρασης για τα παλαιότερα συστήματα. Το HSB βασίζεται στην τεχνολογία Jini [135], μια αρχιτεκτονική προσανατολισμένη στις υπηρεσίες για την κατασκευή ασφαλών, κλιμακούμενων κατακευματισμένων συστημάτων. Ειδικότερα, έχει δημιουργηθεί μια υπηρεσία Jini για την αναζήτηση της SNOMED CT XML βάσης δεδομένων και αποτελεί μία από τις βασικές υπηρεσίες του HSB. Οι εφαρμογές που συνδέονται με το HSB επικοινωνούν μέσω JavaSpaces, χρησιμοποιώντας το HL7 v3 στο πραγματικό περιεχόμενο των μηνυμάτων. Η Εικόνα 2-9 δείχνει τον απώτερο στόχο του HSB: το στόχο των υπολογιστών και των κινητών υπολογιστικών συσκευών που έχουν πρόσβαση στο HSB προκειμένου να επικοινωνούν και να έχουν πρόσβαση σε υπηρεσίες όπως διακομιστές ορολογίας και βάσεις δεδομένων ασθενών.



Εικόνα 2-9 - Ένα πλαίσιο για τη σημασιολογική διαλειτουργικότητα στην υγειονομική περίθαλψη: Αρχιτεκτονική με βάση τα πρότυπα υγείας

Πως οι οντολογίες ενισχύουν τη σημασιολογική διαλειτουργικότητα στο χώρο της υγείας (How Ontologies Can Improve Semantic Interoperability in Health Care) [136]

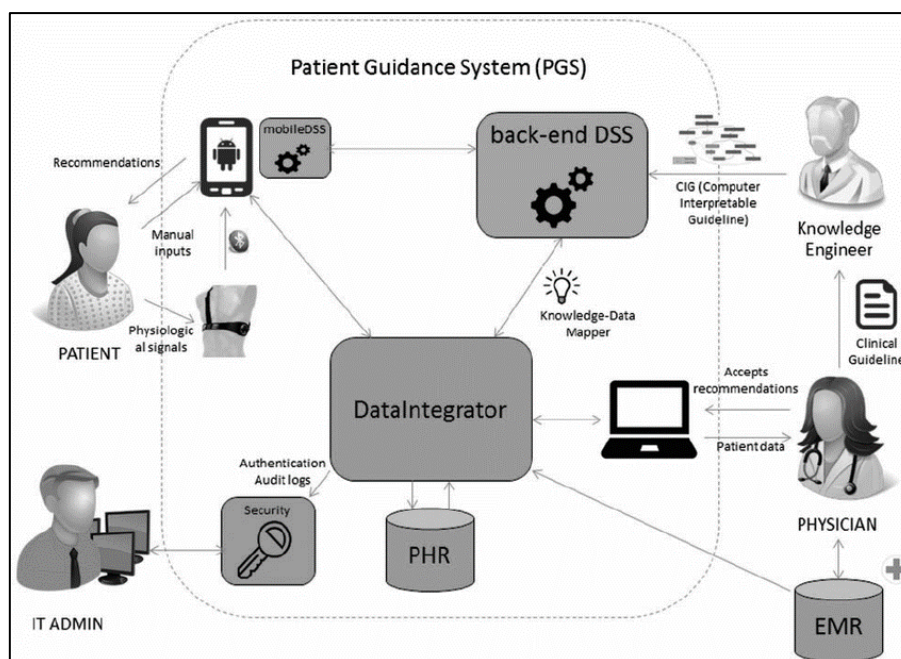
Η συγκεκριμένη έρευνα, έχει ως στόχο αφενός να αναλύσει το ακριβές νόημα της «δέσμευσης» μεταξύ των παραστατικών αντικειμένων και των οντοτήτων σε ιατρική οντολογία. Δεύτερον, πρέπει να δημιουργήσει μοτίβα για τις διαφορετικές κατανομές περιεχομένου μεταξύ μοντέλων πληροφοριών και οντολογιών. Αυτά τα πρότυπα πρέπει να επιτρέπουν τον σχηματισμό σημασιολογικά ισοδύναμων εκφράσεων, που πρέπει να εξακριβώνονται με συλλογιστική μηχανής. Όπως φαίνεται στην Εικόνα 2-10, οι οντότητες πληροφοριών των κλινικών μοντέλων επισημαίνονται με εκφράσεις OWL Description Logic (OWL-DL) [137] σύμφωνα με ορισμένα προκαθορισμένα πρότυπα, με βάση την προτεινόμενη οντολογική υποδομή. Αυτά τα πρότυπα προστίθενται σε κλινικά μοντέλα κατά τη διάρκεια της δημιουργίας τους, προκειμένου να συμπληρωθούν με τα δεδομένα των ασθενών με την κατάλληλη υποστήριξη λογισμικού. Ως αποτέλεσμα, κάθε μοντέλο έχει ένα σύνολο σχολιασμών, οι οποίοι επεξεργάζονται περαιτέρω. Τέλος, ένα σύστημα ερωτήματος επιτρέπει την εκτέλεση ομογενών ερωτημάτων για την ανάκτηση δεδομένων ασθενούς από ετερογενώς εκπροσωπούμενα σύνολα δεδομένων.



Εικόνα 2-10 - Οι οντολογίες ενισχύουν τη σημασιολογική διαλειτουργικότητα

Επίλυση της πρόκλησης διαλειτουργικότητας ενός καταναμημένου πολύπλοκου συστήματος καθοδήγησης ασθενών: Ένας ενσωματιστής δεδομένων που βασίζεται στο πρότυπο Virtual Medical Record του HL7 (Solving the interoperability challenge of a distributed complex patient guidance system: a data integrator based on HL7's Virtual Medical Record standard) [138]

Η έρευνα αυτή περιγράφει έναν Data Integrator (DI) [139], βασισμένο στο μοντέλο πληροφόρησης που παρέχεται από το πρότυπο Virtual Medical Record (VMR) [140] του HL7, το οποίο μπορεί να χρησιμοποιηθεί ως πύλη μεταξύ των πηγών δεδομένων και των στοιχείων από το σύστημα καθοδήγησης ασθενούς (Patient Guidance System - PGS), διευκολύνοντας τη διαλειτουργικότητά τους. Ο DI ενσωματώνει την αποθήκευση δεδομένων, κρύβοντας την πολυπλοκότητα του από τα υπόλοιπα στοιχεία, ενώ συγχρόνως παρέχει ένα API κατάλληλο για τις ανάγκες υλοποίησης των στοιχείων που προσθέτουν νέα δεδομένα στους καταναλωτές δεδομένων (ένας ιατρός που χρησιμοποιεί μια δικτυακή πύλη για να διαβάσει αυτές τις συστάσεις). Συγκεκριμένα, ο DI παρέχει διασυνδέσεις προσανατολισμένες προς τη διαφανή σύνδεση με ένα PHR όπου μπορούν να αποθηκευτούν επιλεγμένα δεδομένα. Ο DI έχει ήδη χρησιμοποιηθεί ως μέρος ενός πραγματικού PGS χρησιμοποιώντας το πρότυπο HL7 vMR και τους μηχανισμούς επέκτασης του, επιδεικνύοντας ότι η σύνθετη σημασιολογία πολλών αλληλοεπιδρώντων συστατικών του PGS μπορεί να χρησιμοποιηθεί (Εικόνα 2-11).

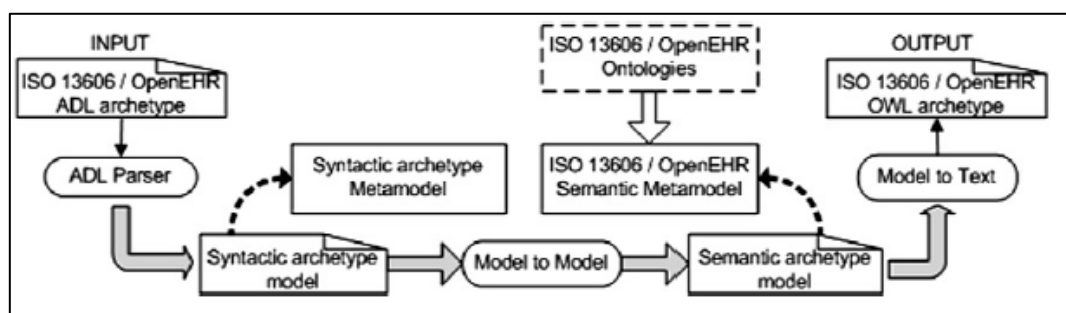


Εικόνα 2-11 - Επίλυση της πρόκλησης διαλειτουργικότητας ενός καταναμημένου πολύπλοκου συστήματος καθοδήγησης ασθενών

Μια προσέγγιση για τη σημασιολογική διαλειτουργικότητα των αρχέτυπων ISO EN 13606 και OpenEHR (An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes) [141]

Στην έρευνα αυτή, δύο κύριες οντολογίες χτίστηκαν για κάθε πρότυπο: (1) η (ISO EN 13606 / OpenEHR)-SP οντολογία, η οποία αντιπροσωπεύει τις κλινικές δομές δεδομένων και τους τύπους δεδομένων που ορίζονται στο μοντέλο αναφοράς κάθε προτύπου και (2) η (ISO EN 13606 / OpenEHR) -AR οντολογία, η οποία καθορίζει το μοντέλο αρχέτυπων και επαναχρησιμοποιεί το SP. Διατίθενται σε δημόσιο χώρο αποθήκευσης και αποτελούν τη βασική τεχνολογική υποδομή. Η διαδικασία μετασχηματισμού διαιρείται στις ακόλουθες τρεις φάσεις (Εικόνα 2-12):

- Το ADL εισόδου εκφράζεται ως συντακτικό μοντέλο, ενώ μετατρέπεται σε μοντέλο που συμμορφώνεται με το πρότυπο αρχέτυπου αντικειμένου.
- Αυτό το συντακτικό μοντέλο μετασχηματίζεται σε ένα σημασιολογικό μοντέλο χρησιμοποιώντας ένα μοντέλο μετασχηματισμού.
- Το σημασιολογικό μοντέλο μετατρέπεται σε OWL σύμφωνα με τις οντολογίες EHR χρησιμοποιώντας ένα μοντέλο μετασχηματισμού κειμένου.



Εικόνα 2-12 - Μια προσέγγιση για τη σημασιολογική διαλειτουργικότητα των αρχέτυπων ISO EN 13606 και OpenEHR

ΚΕΦΑΛΑΙΟ 3

Βασικές Τεχνολογίες

Περιεχόμενα

3.1. Οντολογίες

3.1.1. Ανάγκη Χρήσης Οντολογιών

3.1.2. Δημιουργία Οντολογιών

3.1.3. Χρήση Οντολογιών στον τομέα της Υγείας

3.2. Επεξεργασία Φυσικής Γλώσσας

3.2.1. Τεχνικές Συντακτικής Ομοιότητας

3.2.2. Τεχνικές Σημασιολογικής Ομοιότητας

Περίληψη

Στο **ΚΕΦΑΛΑΙΟ 3**, αναφέρονται οι βασικές τεχνολογίες που χρησιμοποιήθηκαν προς την υλοποίηση της παρούσας προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων, αναφορικά με την ύπαρξη των οντολογιών, της ανάγκης τους και της χρήσης τους στο χώρο της υγείας. Παράλληλα, αναλύονται και τεχνικές οι οποίες στοχεύουν στη δημιουργία οντολογιών. Ακολούθως, γίνεται αναφορά στον τομέα της επεξεργασίας φυσικής γλώσσας, αναφέροντας τη σημασία των τεχνικών συντακτικής και σημασιολογικής ομοιότητας μεταξύ λεκτικών και κειμένων, ένας κλάδος ο οποίος επίσης επηρέασε την τελική προσέγγιση για την επίτευξη των τελικών οντολογικών αντιστοιχίσεων. Εν τέλει, επισημαίνονται και παρουσιάζονται τεχνικές εύρεσης τόσο συντακτικής όσο και σημασιολογικής ομοιότητας μεταξύ λεκτικών και κειμένων.

3.1 Οντολογίες

3.1.1 Ανάγκη Χρήσης Οντολογιών

Οι άνθρωποι, οι οργανισμοί και τα συστήματα λογισμικού πρέπει να επικοινωνούν μεταξύ τους. Ωστόσο, λόγω των διαφορετικών αναγκών και των συνθηκών, μπορεί να υπάρχουν πολύ διαφορετικές απόψεις και παραδοχές σχετικά το ίδιο αντικείμενο. Κάθε ένα από τα προαναφερθέντα χρησιμοποιεί διαφορετική φρασεολογία, έχοντας διαφορετικές, επικαλυπτόμενες ή/και εσφαλμένες ιδέες, δομές και μεθόδους. Η επακόλουθη έλλειψη κοινής κατανόησης οδηγεί σε κακή επικοινωνία μέσα και μεταξύ αυτών των ανθρώπων και των συστημάτων. Στο πλαίσιο της δημιουργίας ενός συστήματος πληροφορικής, αυτή η έλλειψη κοινής αντίληψης οδηγεί σε δυσκολίες στον προσδιορισμό των απαιτήσεων και συνεπώς στον καθορισμό μιας προδιαγραφής του συστήματος, περιορίζοντας έτσι τη διαλειτουργικότητα και την πιθανή επαναχρησιμοποίηση του συστήματος ή το διαμοιρασμό του. Ο τρόπος αντιμετώπισης αυτών των προβλημάτων είναι η μείωση ή η εξάλειψη της εννοιολογικής και ορολογικής σύγχυσης και η επίτευξη κοινής κατανόησης. Μια τέτοια κατανόηση μπορεί να λειτουργήσει ως ενοποιητικό πλαίσιο για τις διάφορες απόψεις και να χρησιμεύσει ως βάση για την επικοινωνία μεταξύ των ατόμων με διαφορετικές ανάγκες και οπτικές γωνίες. Παράλληλα μπορεί να οδηγήσει στη διαλειτουργικότητα μεταξύ των συστημάτων η οποία επιτυγχάνεται μεταφράζοντας μεταξύ διαφορετικών μεθόδων μοντέλων, γλωσσών και προγραμματιστικών εργαλείων.

Η λύση στο πρόβλημα αυτό είναι να υπάρξει μία διαμοιρασμένη κατανόηση η οποία μπορεί να βοηθήσει στη διαδικασία ταυτοποίησης απαιτήσεων και καθορισμού των προδιαγραφών για ένα σύστημα πληροφορικής. Αυτό ισχύει ιδιαίτερα όταν οι απαιτήσεις αφορούν διαφορετικές ομάδες που χρησιμοποιούν διαφορετική ορολογία στον ίδιο τομέα ή σε πολλαπλούς τομείς. Η Οντολογία [142] είναι ο όρος που χρησιμοποιείται για την αναφορά στην διαμοιρασμένη κατανόηση κάποιου τομέα ενδιαφέροντος που μπορεί να χρησιμοποιηθεί ως ενοποιητικό πλαίσιο για την επίλυση των ανωτέρω προβλημάτων. Μια οντολογία συνεπάγεται αναπόφευκτα ή ενσωματώνει κάποιο είδος κοσμοθεωρίας σε σχέση με ένα συγκεκριμένο τομέα. Η παγκόσμια άποψη συχνά θεωρείται ως ένα σύνολο εννοιών (π.χ. οντοτήτων, χαρακτηριστικών, διαδικασιών), των ορισμών τους και των σχέσεών τους. Μια τέτοια εννοιοποίηση μπορεί να είναι σιωπηρή. Η οντολογία λοιπόν, χρησιμοποιείται μερικές φορές για να αναφερθεί σε αυτή τη σιωπηρή

εννοιοποίηση. Ωστόσο, η πιο συνηθισμένη χρήση είναι ότι η οντολογία είναι μια ρητή αναφορά ή αναπαράσταση μιας εννοιολογικής σκέψης.

Μια οντολογία μπορεί να πάρει μια ποικιλία μορφών, αλλά αναγκαστικά θα περιλαμβάνει ένα λεξιλόγιο όρων και κάποια εξειδίκευση του νόηματός τους. Ο βαθμός διατύπωσης με τον οποίο δημιουργείται ένα λεξιλόγιο και το νόημα ποικίλλει σημαντικά. Υπάρχουν τέσσερις διαφορετικοί τρόποι, ανάλογα με το πως θα καθοριστεί μια οντολογία [143]:

- Εξαιρετικά άτυπος τρόπος, όπου η οντολογία εκφράζεται χαλαρά στη φυσική γλώσσα.
- Ημι-ανεπίσημος τρόπος, όπου η οντολογία εκφράζεται σε περιορισμένη και δομημένη μορφή φυσικής γλώσσας, αυξάνοντας σημαντικά τη σαφήνεια και μειώνοντας την ασάφεια.
- Ημι-τυπικός τρόπος, όπου η οντολογία εκφράζεται σε μια τεχνητή τυπικά καθορισμένη γλώσσα.
- Αυστηρά τυπικός τρόπος, όπου η οντολογία εκφράζεται με σχολαστικά ορισμένους όρους με επίσημη σημασιολογία, θεωρήματα και αποδείξεις τέτοιων ιδιοτήτων.

Ουσιαστικά, οι οντολογίες είναι συμφωνίες για κοινές αντιλήψεις. Οι κοινές ιδέες περιλαμβάνουν τα εννοιολογικά πλαίσια για τη μοντελοποίηση της γνώσης του τομέα, συμπεριλαμβάνοντας ειδικά πρωτόκολλα περιεχομένου για την επικοινωνία μεταξύ διαλειτουργικών πρακτόρων και συμφωνίες για την εκπροσώπηση συγκεκριμένων θεωριών τομέα. Στο πλαίσιο της ανταλλαγής γνώσεων, οι οντολογίες καθορίζονται με τη μορφή ορισμών του λεξιλογίου αναπαραγωγής. Μια πολύ απλή περίπτωση θα ήταν μια ιεραρχία τύπων, προσδιορίζοντας τις τάξεις και τις σχέσεις υποκειμένων τους. Τα σχήματα σχεσιακής βάσης δεδομένων χρησιμεύουν επίσης ως οντολογίες, καθορίζοντας τις σχέσεις που μπορούν να υπάρχουν σε κάποια κοινή βάση δεδομένων και τους περιορισμούς ακεραιότητας που πρέπει να τηρούν.

Μια οντολογία λοιπόν είναι μια επίσημη περιγραφή της γνώσης ως ένα σύνολο εννοιών εντός ενός τομέα και των σχέσεων που υπάρχουν μεταξύ τους. Για να επιτρέψουμε μια τέτοια περιγραφή, πρέπει να προσδιορίσουμε τυπικά στοιχεία όπως τα αντικείμενα, τις κλάσεις, τις ιδιότητες και τις συσχετίσεις καθώς και περιορισμούς, κανόνες και αξιώματα. Ως αποτέλεσμα, οι οντολογίες δεν

εισάγουν μόνο μια αντιπροσωπευτική και επαναχρησιμοποιήσιμη εκπροσώπηση γνώσης, αλλά μπορούν επίσης να προσθέσουν νέες γνώσεις σχετικά με τον τομέα.

Υπάρχουν, βεβαίως, και άλλες μέθοδοι που χρησιμοποιούν επίσημες προδιαγραφές για την εκπροσώπηση της γνώσης, όπως λεξιλόγια, ταξινομίες, θεματικοί χάρτες και λογικά μοντέλα. Ωστόσο, σε αντίθεση με τα ταξινομικά καθεστώτα ή τα σχήματα σχεσιακών βάσεων δεδομένων, οι οντολογίες εκφράζουν σχέσεις και επιτρέπουν στους χρήστες να συνδέουν πολλαπλές έννοιες με άλλες έννοιες, με διάφορους τρόπους.

Ως ένα από τα δομικά στοιχεία των σημασιολογικών τεχνολογιών, οι οντολογίες αποτελούν μέρος της στοίβας των προτύπων του W3C για τον Σημασιολογικό Ιστό [144]. Παρέχουν στους χρήστες την απαραίτητη δομή για τη σύνδεση μιας πληροφορίας με άλλες πληροφορίες στο Διαδίκτυο των Συνδεδεμένων Δεδομένων (Internet of Things - IoT) [145]. Επειδή χρησιμοποιούνται για τον προσδιορισμό κοινών παραστάσεων μοντελοποίησης δεδομένων από κατανεμημένα και ετερογενή συστήματα και βάσεις δεδομένων, οι οντολογίες επιτρέπουν τη διαλειτουργικότητα της βάσης δεδομένων, την αναζήτηση σε βάσεις δεδομένων και την ομαλή διαχείριση της γνώσης.

Μερικά από τα κύρια χαρακτηριστικά των οντολογιών είναι ότι εξασφαλίζουν μια κοινή κατανόηση των πληροφοριών και ότι κάνουν ρητές υποθέσεις. Ως αποτέλεσμα, η διασύνδεση και η διαλειτουργικότητα του μοντέλου το καθιστούν ανεκτίμητο για την αντιμετώπιση των προκλήσεων πρόσβασης και διερεύνησης δεδομένων σε μεγάλους οργανισμούς. Επίσης, βελτιώνοντας τα μεταδεδομένα και την προέλευση, επιτρέποντας έτσι στους οργανισμούς να κατανοήσουν καλύτερα τα δεδομένα τους, οι οντολογίες βελτιώνουν την ποιότητα των δεδομένων. Ένα από τα κύρια χαρακτηριστικά των οντολογιών είναι ότι, έχοντας τις βασικές σχέσεις ανάμεσα σε έννοιες ενσωματωμένες σε αυτές, επιτρέπουν την αυτοματοποιημένη συλλογιστική για τα δεδομένα. Αυτός ο συλλογισμός είναι εύκολος να εφαρμοστεί σε βάσεις δεδομένων σημασιολογικών γραφημάτων που χρησιμοποιούν οντολογίες ως σημασιολογικά σχήματα τους. Επιπλέον, οι οντολογίες λειτουργούν σαν «εγκέφαλος». «Εργάζονται και λογίζονται» με έννοιες και σχέσεις, με τρόπους που προσεγγίζουν τον τρόπο με τον οποίο οι άνθρωποι αντιλαμβάνονται. Εκτός από το σκεπτικό, οι οντολογίες παρέχουν μια πιο συνεκτική και εύκολη πλοήγηση, καθώς οι χρήστες κινούνται στη δομή οντολογίας από τη μία έννοια στην άλλη. Ένα άλλο πολύτιμο χαρακτηριστικό είναι ότι οι οντολογίες είναι

εύκολο να επεκταθούν καθώς οι σχέσεις και η αντιστοίχιση ιδεών είναι εύκολο να προστεθούν στις υπάρχουσες οντολογίες. Ως αποτέλεσμα, το μοντέλο αυτό εξελίσσεται με την ανάπτυξη δεδομένων χωρίς να επηρεάζεται από εξαρτώμενες διαδικασίες και συστήματα, εάν κάτι πάει στραβά ή πρέπει να αλλάξει. Οι οντολογίες παρέχουν επίσης τα μέσα για την αναπαραγωγή οποιωνδήποτε μορφών δεδομένων, συμπεριλαμβανομένων των μη δομημένων, ημιδομημένων ή δομημένων δεδομένων, επιτρέποντας την ομαλότερη ενσωμάτωση των δεδομένων, την ευκολότερη έννοια και την εξόρυξη δεδομένων.

Ενώ οι οντολογίες παρέχουν ένα πλούσιο σύνολο εργαλείων για τη μοντελοποίηση δεδομένων, η χρησιμότητά τους έρχεται με ορισμένους περιορισμούς. Ένας τέτοιος περιορισμός είναι οι διαθέσιμες κατασκευές ιδιότητας. Για παράδειγμα, ενώ παρέχει ισχυρές κατασκευές κατηγορίας, η πιο πρόσφατη έκδοση της γλώσσας (OWL2) [143], έχει ένα κάπως περιορισμένο σύνολο κατασκευών ιδιότητας. Ένας άλλος περιορισμός προέρχεται από τον τρόπο με τον οποίο η OWL χρησιμοποιεί περιορισμούς. Χρησιμοποιούν για τον καθορισμό του τρόπου δομής των δεδομένων και για την πρόληψη της προσθήκης δεδομένων ασυμβίβαστων με αυτούς τους περιορισμούς. Αυτό, ωστόσο, δεν είναι πάντα επωφελές. Συχνά, τα δεδομένα που εισάγονται από μια νέα πηγή θα είναι δομικά ασύμβατα με τους περιορισμούς που ορίζονται χρησιμοποιώντας το OWL. Συνεπώς, αυτά τα νέα δεδομένα θα πρέπει να τροποποιηθούν πριν ενσωματωθούν με αυτά που έχουν ήδη φορτωθεί.

3.1.2 Δημιουργία Οντολογιών

Οι πηγές πληροφοριών μπορούν να είναι δομημένες όπως οι σχεσιακές βάσεις δεδομένων ή ημι-δομημένες όπως οι πηγές δεδομένων XML. Η XML μεταξύ των περισσότερων φορέων τυποποίησης, αποτελεί την πλέον τυποποιημένη μορφή ανταλλαγής δεδομένων. Παρόλο που η XML έχει ξεκινήσει μόνο τα τελευταία χρόνια και εξακολουθεί να αναπτύσσεται ταχέως, υιοθετείται ευρέως στη διαχείριση πληροφοριών. Έχει θεωρηθεί ως η πιο ισχυρή πρόοδος που έχει δει το Διαδίκτυο εδώ και χρόνια. Παρέχει ένα μορφότυπο που είναι αναγνώσιμο από τον άνθρωπο και ερμηνεύεται από μηχανή. Η απλότητα και η ευελιξία της χρήσης ταιριάζει με μια βασική τεχνολογία που δίνει «φορητότητα» πληροφοριών σε πλατφόρμες, εφαρμογές και οργανισμούς, η οποία είναι ζωτικής σημασίας για τις απαιτήσεις ανταλλαγής πληροφοριών, αλλά στερείται υποστήριξης για την έκφραση της σημασιολογίας της γνώσης του τομέα. Σαν αποτέλεσμα, τα τυπικά εννοιολογικά μοντέλα ή οντολογίες είναι απαραίτητα για να εκφράσουν τη σημασιολογία των δεδομένων. Οι πηγές δεδομένων μπορεί να είναι ετερογενείς στη σύνταξη, το σχήμα ή τη σημασιολογία, καθιστώντας έτσι τη διαλειτουργικότητα των δεδομένων δύσκολη υπόθεση [146]. Οι οντολογίες παρέχουν μια υποσχόμενη τεχνολογία για την επίλυση του σημασιολογικού προβλήματος ετερογένειας, διότι επιτρέπουν να εκπροσωπούνται ρητά οι κοινές σημασιολογίας μιας περιοχής. Μια οντολογία μπορεί να οριστεί ως τυπική, ρητή προδιαγραφή μιας κοινής αντίληψης [147], αναφερόμενη σε ένα αφηρημένο μοντέλο φαινομένων στον κόσμο, έχοντας προσδιορίσει τις σχετικές έννοιες αυτών των φαινομένων. Ο όρος «ρητή» σημαίνει ότι ο τύπος των εννοιών που χρησιμοποιούνται και οι περιορισμοί στη χρήση τους καθορίζονται ρητά. Το «τυπικό» αναφέρεται στο γεγονός ότι η οντολογία πρέπει να είναι μηχανικά αναγνώσιμη. Το «κοινό» αντικατοπτρίζει ότι η οντολογία θα πρέπει να συλλάβει συναινετικές γνώσεις αποδεκτές από τις κοινότητες. Μια οντολογία ορίζει τυπικά διαφορετικές έννοιες ενός τομέα και σχέσεις μεταξύ αυτών των εννοιών, οι οποίες μπορούν να εκφραστούν με τη χρήση της OWL.

3.1.2.1 OWL (Web Ontology Language)

Η OWL, έχει γίνει ένα δημοφιλές πρότυπο για την αναπαράσταση και ανταλλαγή δεδομένων. Η OWL υποστηρίζει την αναπαράσταση της γνώσης χρησιμοποιώντας κλάσεις, ιδιότητες και στιγμιότυπα για χρήση σε ένα καταναμημένο περιβάλλον όπως το World Wide Web (WWW) [148]. Η OWL περιλαμβάνει τρεις υπο-γλώσσες που ονομάζονται:

- **OWL Lite [149]:** υποστηρίζει εκείνους τους χρήστες που χρειάζονται κυρίως ιεραρχική ταξινόμηση και απλούς περιορισμούς.
- **OWL-DL:** υποστηρίζει εκείνους τους χρήστες που επιθυμούν τη μέγιστη εκφραστικότητα διατηρώντας παράλληλα την υπολογιστική πληρότητα (όλα τα συμπεράσματα είναι εγγυημένα να υπολογιστούν) και την αποφασιστικότητα (όλοι οι υπολογισμοί θα τελειώσουν σε πεπερασμένο χρόνο).
- **OWL Full [150]:** υποστηρίζει χρήστες που θέλουν μέγιστη εκφραστικότητα και την συντακτική ελευθερία του RDF χωρίς υπολογιστικές εγγυήσεις.

Καθορίζοντας τόσο τις κλάσεις των αντικειμένων όσο και τις ιδιότητες σχέσεων καθώς και την ιεραρχική σειρά τους, η OWL εμπλουτίζει τη μοντελοποίηση οντολογιών σε βάσεις δεδομένων σημασιολογικών γραφημάτων, γνωστά ως τριπλότυπα RDF (RDF triplestores) [151]. Η OWL, που χρησιμοποιείται σε συνδυασμό με έναν αλγόριθμο OWL σε τέτοιου είδους RDF triplestores, επιτρέπει ελέγχους συνέπειας (για να βρει οποιεσδήποτε λογικές ασυνέπειες) και εξασφαλίζει ελέγχους ικανοποίησης (για να διαπιστώσει αν υπάρχουν κλάσεις που δεν μπορούν να έχουν ιδιότητες). Επιπλέον, η OWL έρχεται εξοπλισμένη με μέσα για τον ορισμό της ισοδυναμίας και της διαφοράς μεταξύ κλάσεων, ιδιοτήτων, και στιγμιότυπων. Αυτές οι σχέσεις βοηθούν τους χρήστες να ταιριάζουν με τις έννοιες, ακόμη και αν διάφορες πηγές δεδομένων περιγράφουν αυτές τις έννοιες κάπως διαφορετικά. Εξασφαλίζουν επίσης την αποσαφήνιση μεταξύ διαφορετικών περιπτώσεων που μοιράζονται τα ίδια ονόματα ή περιγραφές. Ως οντολογική γλώσσα, η OWL ασχολείται πρωτίστως με τον ορισμό της ορολογίας που μπορεί να χρησιμοποιηθεί στα έγγραφα RDF, δηλαδή τις κλάσεις και τις ιδιότητες. Οι περισσότερες οντολογικές γλώσσες έχουν κάποιο μηχανισμό για τον καθορισμό μιας ταξινόμησης των τάξεων. Στην OWL αντίστοιχα, μπορεί να καθοριστούν ταξονομίες και για τις κλάσεις και για τις ιδιότητες.

Κλάσεις (Subjects)

Στο RDF, ένα αντικείμενο μιας δήλωσης τύπου `rdf:` αποτελεί μια κλάση, δηλαδή αντιπροσωπεύει ένα σύνολο πόρων. Στην OWL, μπορούμε ρητά να δηλώσουμε ότι αυτός ο πόρος είναι μία κλάση δηλώνοντας ότι είναι του τύπου `rdf:type owl:Class`. Συντακτικά, αυτό σημαίνει ότι χρησιμοποιείται ένα στοιχείο `owl:Class`. Είναι σύνηθες οι RDF και OWL κλάσεις να ονομάζονται με μοναδικά ουσιαστικά. Είναι επίσης σύνηθες για τα ονόματα να ξεκινούν με κεφαλαίο γράμμα και να χρησιμοποιούν μικτά κεφαλαία για ονόματα πολλών

λέξεων. Η χρήση των URI για την ονομασία κλάσεων (και ιδιοτήτων) είναι μια σημαντική πτυχή του Σημασιολογικού Ιστού. Οι ανθρώπινες γλώσσες έχουν συχνά πολύ-σημασιακούς όρους, δηλαδή λέξεις που έχουν πολλαπλές έννοιες. Υποθέτοντας ότι αυτές οι διαφορετικές χρήσεις καθορίστηκαν σε διαφορετικές οντολογίες, τότε θα είχαν διαφορετικά URI. Έτσι, κάθε μία από τις έννοιες ενός όρου θα είχε ένα μοναδικό σύμβολο. Ωστόσο, αυτή η λύση συνθέτει το πρόβλημα της συνωνυμίας, όπου γενικά μπορούν να χρησιμοποιηθούν διαφορετικά σύμβολα με την ίδια έννοια. Εάν το ίδιο όνομα κλάσης χρησιμοποιείται για να σημαίνει το ίδιο πράγμα σε διαφορετικές οντολογίες, τότε τεχνικά αυτές οι κλάσεις θα έχουν διαφορετικά URIs. Ωστόσο, όταν ανακαλυφθούν τέτοια προβλήματα, μπορούν εύκολα να επιλυθούν, αφού μπορούμε να χρησιμοποιήσουμε τα αξιώματα OWL για να συνδέσουμε ρητά αυτές τις τάξεις. Αν θέλουμε να καθορίσουμε πρόσθετες πληροφορίες που περιγράφουν την κλάση, τότε συμπεριλαμβάνουμε ιδιότητες από τα λεξιλόγια RDF Schema (RDFS) [152] και / ή OWL (που αντιπροσωπεύονται από τα υποστοιχεία στη σύνταξη XML). Η ιδιότητα `rdfs:subClassOf` μπορεί να χρησιμοποιηθεί για να συσχετίσει μια κλάση με γενικότερες κλάσεις. Μια κλάση μπορεί επίσης να ειπωθεί ότι έχει ακριβώς τα ίδια μέλη με μια άλλη κλάση χρησιμοποιώντας την ιδιότητα `owl:equivalentClass`. Αυτό χρησιμοποιείται συχνά για συνώνυμες κλάσεις, ιδιαίτερα όταν οι κλάσεις προέρχονται από διαφορετικές οντολογίες.

Ιδιότητες (Properties)

Η OWL μπορεί επίσης να ορίζει δύο τύπους ιδιοτήτων: ιδιότητες αντικειμένων και ιδιότητες τύπων δεδομένων. Οι ιδιότητες αντικειμένων καθορίζουν τις σχέσεις μεταξύ ζευγών πόρων. Από την άλλη πλευρά, οι ιδιότητες τύπων δεδομένων, καθορίζουν τη σχέση μεταξύ ενός πόρου και μιας τιμής τύπου δεδομένων. Υπάρχουν ορισμένοι όροι OWL και RDF που χρησιμοποιούνται για την περιγραφή ιδιοτήτων. Ενώ οι κλάσεις σε RDF και OWL ονοματίζονται με αρχικό κεφαλαίο γράμμα, οι ιδιότητες έχουν συνήθως ένα αρχικό μικρό γράμμα. Ωστόσο, όπως τα ονόματα κλάσεων, τα ονόματα ιδιοτήτων χρησιμοποιούν μικτά κεφαλαία σε πολύπλοκα ονόματα. Όπως και με τις κλάσεις, περιγράφουμε τις ιδιότητες συμπεριλαμβάνοντας τα υπο-στοιχεία. Μια έκφραση που χρησιμοποιεί το `rdfs:subPropertyOf` δηλώνει ότι κάθε ζεύγος αντικειμένων που χρησιμοποιεί την ιδιότητα του αντικειμένου, είναι επίσης ένα έγκυρο ζεύγος αντικειμένων χρησιμοποιώντας την ιδιότητα αντικειμένου. Με αυτό τον τρόπο μπορούν να δημιουργηθούν ταξινομίες ιδιοτήτων. Η `owl:equivalentProperty` δηλώνει ότι οι επεκτάσεις ιδιότητας των δύο ιδιοτήτων είναι οι ίδιες. Δηλαδή, κάθε ζεύγος αντικειμένων για μια ιδιότητα είναι ένα έγκυρο ζεύγος αντικειμένων για το άλλο. Αυτό είναι το αναλογικό

χαρακτηριστικό της `owl:equivalentClass` και χρησιμοποιείται συχνά για να περιγράψει συνώνυμες ιδιότητες. Οι ιδιότητες `rdfs:domain` και `rdfs:range` χρησιμοποιούνται για να καθορίσουν τον τομέα και το εύρος μιας ιδιότητας. Η ιδιότητα `rdfs:domain` προσδιορίζει ότι το αντικείμενο οποιασδήποτε δήλωσης που χρησιμοποιεί την ιδιότητα είναι μέλος της κλάσης που καθορίζει. Ομοίως, το `rdfs:range` προσδιορίζει ότι το αντικείμενο οποιασδήποτε δήλωσης που χρησιμοποιεί την ιδιότητα είναι μέλος της κλάσης ή του τύπου δεδομένων που καθορίζει. Αν και αυτές οι ιδιότητες μπορεί να φαίνονται απλές, μπορούν να οδηγήσουν σε ορισμένες παρανοήσεις και πρέπει να χρησιμοποιηθούν προσεκτικά.

Η OWL ορίζει επίσης έναν αριθμό κατασκευαστών που καθορίζουν τη σημασιολογία των ιδιοτήτων. Ορίζει μια άλλη σχέση μεταξύ ιδιοτήτων χρησιμοποιώντας το `owl:inverseOf`. Για παράδειγμα, αυτό μπορεί να χρησιμοποιηθεί για να αποδειχθεί ότι η ιδιότητα `isParentOf` είναι η `owl:inverseOf` της ιδιότητας `isChildOf`. Έτσι, εάν το A είναι ο γονέας του B, τότε το B είναι απαραίτητως το παιδί του A. Η OWL ορίζει επίσης ένα πλήθος χαρακτηριστικών ιδιοτήτων, όπως η `owl:TransitiveProperty`, `owl:SymmetricProperty`, `owl:FunctionalProperty`, και `owl:InverseFunctionalProperty`. Να σημειωθεί πως, αντίθετα με τους κατασκευαστές που έχουμε δει μέχρι τώρα, αυτές αποτελούν κλάσεις. Αυτό οφείλεται στο γεγονός ότι αυτά τα τέσσερα χαρακτηριστικά χρησιμοποιούνται χρησιμοποιώντας το `rdf:type` της ιδιότητας. Οι κατασκευαστές `owl:TransitiveProperty` και `owl:SymmetricProperty` καθορίζουν ότι η ιδιότητα είναι μια μεταβατική σχέση και μια συμμετρική σχέση αντίστοιχα. Το πρώτο μπορεί να χρησιμοποιηθεί για να περιγράψει την ιδιότητα `isTallerThan`, ενώ το τελευταίο μπορεί να χρησιμοποιηθεί για να περιγράψει την ιδιότητα `isFriendOf`. Η ιδιότητα `owl:FunctionalProperty` δηλώνει ότι κάθε πόρος προσδιορίζει με μοναδικό τρόπο την αξία του για την ιδιότητα. Δηλαδή, κανένας πόρος δεν μπορεί να έχει περισσότερες από μία τιμές για την κάθε ιδιότητα. Από την άλλη πλευρά, ο κατασκευαστής `owl:InverseFunctionalProperty` αναφέρει ότι κάθε αξία ιδιότητας προσδιορίζει με μοναδικό τρόπο το αντικείμενο της ιδιότητας. Σαν αποτέλεσμα, το `owl:InverseFunctionalProperty` καθορίζει μια ιδιότητα που μπορεί να χρησιμοποιηθεί ως πρωτεύον κλειδί για την κλάση αντικειμένου που είναι ο τομέας της ιδιότητας.

Στιγμιότυπα (Instances)

Εκτός από την έκφραση της σημασιολογίας των τάξεων και των ιδιοτήτων, το OWL μπορεί να χρησιμοποιηθεί για την αναφορά σε στιγμιότυπα. Για

παράδειγμα η ιδιότητα `owl:sameAs` χρησιμοποιείται για να δηλώσει ότι δύο στιγμιότυπα είναι ίδια. Αυτό είναι ιδιαίτερα χρήσιμο σε κατανεμημένες περιπτώσεις, όπως στον ιστό, όπου διαφορετικές οντότητες μπορούν να χρησιμοποιούν διαφορετικά αναγνωριστικά για να αναφέρουν τα ίδια πράγματα. Για παράδειγμα, πολλές διευθύνσεις Uniform Resource Locator (URL) [153] ενδέχεται να αναφέρονται στο ίδιο άτομο. Ένα άτομο μπορεί να έχει διαφορετικές διευθύνσεις URL για τις προσωπικές και επαγγελματικές ιστοσελίδες του, ή ακόμα και για πολλαπλές ιστοσελίδες εργασίας, καθώς αλλάζει εργασία με την πάροδο του χρόνου. Είναι επίσης δυνατό να πούμε ότι δύο στιγμιότυπα έχουν διαφορετικά άτομα. Η ιδιότητα `owl:differentFrom` χρησιμοποιείται για να γίνει αυτό. Συχνά συμβαίνει όταν θέλουμε να πούμε ότι ένα σύνολο ατόμων είναι διαχωρισμένα.

3.1.2.2 Κομβικές Τεχνικές Δημιουργίας

Στον τομέα αυτό έχουν αναπτυχθεί διάφορες στρατηγικές για την παραγωγή οντολογιών από ετερογενείς πηγές δεδομένων. Ορισμένες προσεγγίσεις στοχεύουν περισσότερο σε μια γενική αντιστοιχία μεταξύ XML και RDF, είτε έχουν ως στόχο την αντιστοιχία του XML Schema στην OWL. Η OWL και το RDF έχουν την ίδια αξία, αλλά η OWL είναι μια ισχυρότερη γλώσσα με μεγαλύτερη μηχανική ερμηνεία από την RDF. Η OWL έρχεται με ένα μεγαλύτερο λεξιλόγιο και ισχυρότερη σύνταξη από το RDF. Αυτή η ενότητα επικεντρώνεται στις περισσότερο κομβικές στρατηγικές για την παραγωγή οντολογιών OWL.

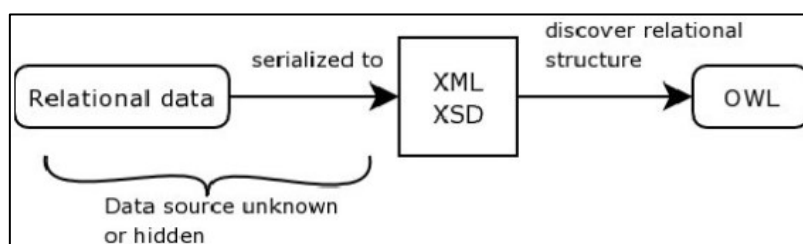
Μετατροπή του XML Schema σε OWL (Lifting XML Schema to OWL) [154]

Ο Ferdinand πρότεινε άμεσες αντιστοιχίες από το XML Schema σε OWL καθώς επίσης περιγράφει αντιστοιχίες από γραφήματα XML σε RDF, με αυτές τις αντιστοιχίες να είναι ανεξάρτητες η μία από την άλλη, δηλαδή οι παραγόμενες περιπτώσεις να μη σέβονται απαραίτητως την οντολογία που δημιουργήθηκε από το XML Schema. Το XML Schema για τη διαδικασία αντιστοιχίας σε OWL βασίζεται σε ένα σύνολο κανόνων ερμηνείας και μετασχηματισμού από το XML Schema στο OWL. Το XML Schema, οι ομάδες μοντέλων και οι ομάδες χαρακτηριστικών έχουν αντιστοιχιστεί σε κατηγορίες OWL. Πιο συγκεκριμένα, τα στοιχεία του `simpleType` και όλων των χαρακτηριστικών αντιστοιχούνται σε ένα `owl:DatatypeProperty`, και τα στοιχεία του `complexType` αντιστοιχίζονται σε ένα `owl:ObjectProperty`. Δύο τύποι κληρονομιάς υποστηρίζονται από το XML Schema: κληρονομιά με περιορισμό και κληρονομιά με επέκταση. Και οι δύο αντιστοιχούνται στον μοναδικό μηχανισμό κληρονομιάς στο OWL: `rdfs:subClassOf`. Οι

αντιστοιχίσεις που προτείνονται σε αυτή την προσέγγιση προορίζονται να εφαρμοστούν για τη μηχανική των εφαρμογών ιστού. Συγκεκριμένα, για την ενίσχυση των παραδοσιακών γλωσσών και εργαλείων XML από τις δυνατότητες του OWL.

Αντιστοιχισή XML σε OWL Οντολογίες (Mapping XML to OWL Ontologies) [155]

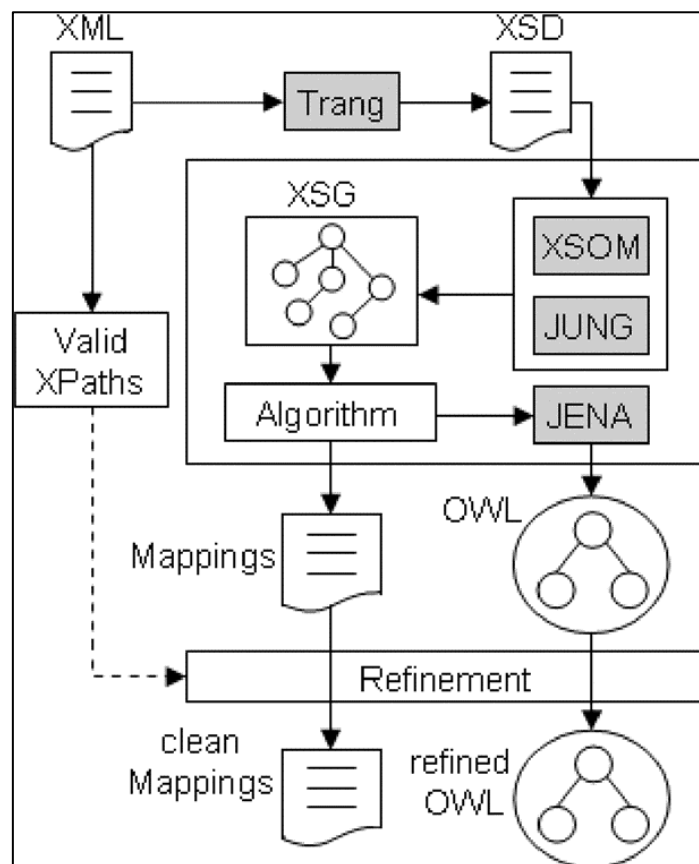
Οι Bohring και Auer έχουν αναπτύξει ένα εργαλείο που ονομάζεται `xml2owl` για να δημιουργήσει μια οντολογία OWL από ένα XML Schema και να μετατρέψει τα δεδομένα XML σε περιπτώσεις OWL συμβατές με τη δημιουργηθείσα οντολογία (Εικόνα 3-1). Από τεχνικής πλευράς, έχουν αναπτύξει τέσσερα στιγμιότυπα XSLT για μετατροπή αρχείων XML σε OWL, χωρίς οποιαδήποτε άλλη παρέμβαση κατά τη διάρκεια του μετασχηματισμού. Τέλος, δημιουργείται αυτόματα το αρχείο OWL. Αυτή η μέθοδος εφαρμόστηκε επίσης στην πλατφόρμα Ontowiki [156]. Οι προσεγγίσεις Bohring και Ferdinand είναι παρόμοιες και βασίζονται σε ένα σύνολο κανόνων μετασχηματισμού από το XML Schema στο OWL. Πρότειναν ότι το XML Schema Definition (XSD) με περιπλοκό περιεχόμενο να μετατρέπεται σε κλάσεις OWL. Όταν ένα στοιχείο περιέχει ένα άλλο στοιχείο, δημιουργείται μια ιδιότητα αντικειμένου OWL μεταξύ των αντίστοιχων κλάσεων OWL, ενώ οι ιδιότητες τύπου δεδομένων OWL εξέρχονται από τα χαρακτηριστικά XML. Οι περιορισμοί αρχιτεκτονικής XML Schema όπως το `minOccurs` ή το `maxOccurs`, αντιστοιχίζονται στους ισοδύναμους περιορισμούς σε OWL, `minCardinality` και `maxCardinality`. Κατά τη δημιουργία οντολογίας από ένα XML Schema, ο χρήστης δεν έχει κανέναν έλεγχο στη διαδικασία. Επομένως, οι οντολογίες που δημιουργούνται δεν είναι σημασιολογικά πλουσιότερες από τα αντιστοιχισμένα XML Schemas. Επιπλέον, δεν αναφέρουν τον τρόπο εφαρμογής τους σε έγγραφα πολλαπλών πηγών. Οι προσεγγίσεις του Ferdinand και του Bohring εισάγουν μια καλή βάση κανόνων για τη δημιουργία οντολογιών OWL από το XML Schema. Ωστόσο, αναφέρονται μόνο στις απλές περιπτώσεις και δεν αναφέρονται σε πολύπλοκες περιπτώσεις που προκύπτουν από την επαναχρησιμοποίηση παγκόσμιων τύπων και στοιχείων. Επίσης, δεν αναφέρουν πώς να καθοριστούν οι αντιστοιχίσεις μεταξύ πολλαπλών πηγών δεδομένων XML και δημιουργίας οντολογίας OWL.



Εικόνα 3-1 - Αντιστοιχισή XML σε OWL Οντολογίες

Δημιουργία Οντολογιών από Πηγές Δεδομένων XML (Building Ontologies from XML Data Sources) [157]

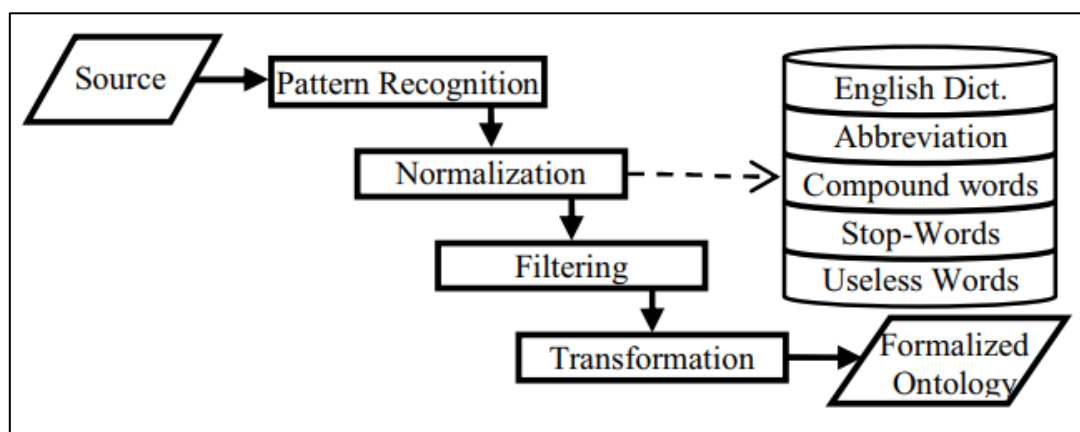
Οι R. Ghawi και N. Cullot πρότειναν ένα εργαλείο, το οποίο ονομάζεται X2OWL (Εικόνα 3-2), το οποίο στοχεύει στην οικοδόμηση μιας οντολογίας OWL από μια πηγή δεδομένων XML. Αυτή η μέθοδος βασίζεται στο XML Schema για να δημιουργήσει αυτόματα τη δομή οντολογίας καθώς και σε ένα σύνολο γεφυρώσεων αντιστοίχισης μεταξύ των οντοτήτων της πηγής δεδομένων XML και της δημιουργούμενης οντολογίας, για τη μετάφραση ερωτημάτων μεταξύ του OWL και του XML. Αυτή η προσέγγιση αντιμετωπίζει απλές και περίπλοκες περιπτώσεις που προκύπτουν από την επαναχρησιμοποίηση παγκόσμιων τύπων και στοιχείων που χρησιμοποιούνται για τη δημιουργία XML Schemas. Τα XML Schemas μπορούν να μοντελοποιηθούν χρησιμοποιώντας διαφορετικά στυλ, μερικά από αυτά χρησιμοποιούν ένα ενιαίο στοιχείο και άλλα χρησιμοποιούν πολλά στοιχεία, μερικά χρησιμοποιούν πολλαπλούς τύπους και άλλα χρησιμοποιούν μόνο τοπικούς τύπους. Ωστόσο, η μέθοδος αντιστοίχισης θα πρέπει να αντιμετωπίσει όλα τα πιθανά σχέδια. Η μέθοδος που παρουσιάζεται περιλαμβάνει επίσης ένα βήμα βελτίωσης που επιτρέπει τον καθαρισμό των γεφυρών αντιστοίχισης και την αναδιάρθρωση της παραγόμενης οντολογίας.



Εικόνα 3-2 - Δημιουργία οντολογιών από πηγές δεδομένων XML

Μετατρέποντας το XML Schema σε OWL μέσω Μοτίβων (Transforming XML Schema to OWL Using Patterns) [158]

Ο I. Bedini και ο N. Benjamin πρότειναν ένα πλαίσιο που ονομάζεται Janus και στοχεύει να δημιουργήσει μια οντολογία από μία μεγάλη πηγή XML Schemas (Εικόνα 3-3). Παρουσιάζουν μια σειρά μοτίβων που επιτρέπουν τον άμεσο αυτόματο μετασχηματισμό από το XML Schema σε OWL επιτρέποντας την ενσωμάτωση πολλών δεδομένων XML στον Σημασιολογικό Ιστό. Επικεντρώνονται σε μια προηγμένη λογική αναπαράσταση των συστατικών του XML Schema και παρουσιάζουν μια εφαρμογή που είναι εφικτή η εξόρυξη πηγών XML Schema για να εξαγάγει επαρκείς γνώσεις ώστε να δημιουργήσει σημαντικές οντολογίες με ιδιαίτερη έκφραση. Η εφαρμογή τους αποδείχθηκε κυρίως ότι είναι πιο ολοκληρωμένη από άλλες και ότι βελτιώνει σημαντικά τον αριθμό των πολύπλοκων μοτίβων μετασχηματισμού σε σχέση με την προσέγγιση Bohring. Αυτή η προσέγγιση δεν αντιμετωπίζει το ζήτημα του τρόπου δημιουργίας της οντολογίας OWL, εάν δεν υπάρχει XML Schema.



Εικόνα 3-3 - Μετατρέποντας το XML Schema σε OWL μέσω μοτίβων

Χρησιμοποιώντας Σχεσιακές Βάσεις Δεδομένων για τη Δημιουργία OWL Οντολογιών από XML Πηγές Δεδομένων (Using Relational Database to Build OWL Ontology from XML Data Sources) [159]

Οι Xu και Li πρότειναν μια προσέγγιση για την κατασκευή οντολογίας OWL από έγγραφο XML με τη βοήθεια ενός μοντέλου σχέσης οντοτήτων. Δηλαδή, προτείνουν μια προσέγγιση αντιστοίχισης XML-to-Relational (XTR) για την αντιστοίχιση ενός εγγράφου XML σε ένα μοντέλο σχέσης οντότητας και στη συνέχεια μια προσέγγιση αντιστοίχισης προς την οντολογία για τη αντιστοίχιση ενός μοντέλου σχέσης οντότητας σε μία οντολογία OWL. Ωστόσο, η οντολογία OWL εκφράζεται με περιορισμένα λεξιλόγια για την περιγραφή σχεσιακής βάσης δεδομένων, επομένως δεν μπορεί να θεωρηθεί ως οντολογία τομέα.

3.1.3 Χρήση Οντολογιών στον τομέα της Υγείας

Τα τρέχοντα συστήματα υγειονομικής περίθαλψης πρέπει να έχουν τη δυνατότητα να επικοινωνούν με σύνθετα και λεπτομερή ιατρικά δεδομένα με ασφαλή, αυτοματοποιημένο και αποτελεσματικό τρόπο. Αυτό μπορεί να επιτευχθεί με την οικοδόμηση οντολογιών ιατρικών πεδίων [160], προκειμένου να αντιπροσωπεύονται συστήματα ιατρικής ορολογίας. Γενικά, μια οντολογία αντιπροσωπεύει μια κοινή, κοινόχρηστη και επαναχρησιμοποιήσιμη άποψη ενός συγκεκριμένου πεδίου εφαρμογής που δίνει νόημα στις δομές πληροφοριών που ανταλλάσσονται μεταξύ διαφορετικών συστημάτων πληροφοριών. Ως αποτέλεσμα, ο οντολογικός στόχος είναι η επίτευξη κοινής γνώσης για τη μετάδοσή του, μεταξύ διαφορετικών ενδιαφερομένων και συστημάτων εφαρμογής. Ο κύριος στόχος είναι η δημιουργία σημασιολογίας με γενικό τρόπο, παρέχοντας τη βάση για συμφωνία σε έναν τομέα. Γενικά, το πλεονέκτημα της ενσωμάτωσης κάθε πηγής πληροφοριών σε μια τοπική οντολογία είναι να επιτρέψει την ανάπτυξη μιας πηγής οντολογίας ανεξάρτητα από άλλες πηγές ή οντολογίες. Η δημιουργία νέων οντολογιών δεν είναι μια προκαθορισμένη διαδικασία, δεδομένου ότι πρέπει να παρθούν πολλές σχεδιαστικές αποφάσεις, ενώ το υπόβαθρο των σχεδιαστών και η εφαρμογή-στόχος, επηρεάζουν τις αποφάσεις τους με διαφορετικούς τρόπους.

Αναφορικά με τη χρήση οντολογιών στον τομέα της υγείας, οι συγγραφείς στο [161] ανέπτυξαν ένα σύστημα πληροφοριών για την υγειονομική περίθαλψη που βασίζεται σε μεθόδους οντολογίας, όπου οι επαγγελματίες της υγειονομικής περίθαλψης κατάφεραν να κατασκευάσουν οντολογίες σχετικά με νέες ασθένειες, διευκολύνοντας την πρόβλεψη των επερχόμενων ασθενειών. Επιπλέον, οι συγγραφείς στο [162] δημιούργησαν ένα μοντέλο πληροφοριών σχετικό με την υγειονομική περίθαλψη που βασίζεται στην οντολογία για να εφαρμόσει ένα πανταχού παρόν περιβάλλον. Στην περίπτωση αυτή, εξελίχθηκαν πληροφορίες για τα συμφραζόμενα και ταξινομήθηκαν για την υλοποίηση των υπηρεσιών υγειονομικής περίθαλψης χρησιμοποιώντας το μοντέλο πληροφοριών πλαισίου που θα μπορούσε να οριστεί μέσω οντολογιών. Ως αποτέλεσμα, οι υπεύθυνοι ανάπτυξης εφαρμογών και υπηρεσιών υγειονομικής περίθαλψης θα μπορούσαν να χρησιμοποιήσουν τις αισθητηριακές πληροφορίες σε διάφορα περιβάλλοντα, δημιουργώντας οντολογίες που εξαρτώνται από τη συσκευή, βασισμένες σε αυτή την κοινή οντολογία. Η έρευνα του [163] θα πρέπει να αναφερθεί, όπου οι συγγραφείς πρότειναν ένα πλαίσιο και ένα σύνολο εργαλείων που θα μπορούσε να προσφέρει ένα ασφαλές ενιαίο σημείο πρόσβασης στην πλήρη εικόνα ενός πελάτη για τις προσωπικές του πληροφορίες για την υγεία, προτείνοντας ένα

οντολογικό πλαίσιο. Αυτό το πλαίσιο ήταν ένα ανεξάρτητο εργαλείο που θα μπορούσε να συγκεντρώνει αυτόματα και να συνδυάζει τις πληροφορίες υγείας ενός πελάτη από τους διάφορους παρόχους στον κύκλο φροντίδας του, και να παρέχει τις πληροφορίες με ασφάλεια χωρίς να ενοχλεί τον πελάτη με πολλαπλά αιτήματα και συμφωνίες κοινής χρήσης. Τέλος, οι ερευνητές στο [164] ανέπτυξαν ένα εργαλείο που βασίζεται σε οντολογίες για τη βελτίωση του τομέα της σημασιολογικής διαλειτουργικότητας στα δεδομένα της υγειονομικής περίθαλψης. Συνοπτικά, οι συγγραφείς εντόπισαν και διευκρίνισαν τις πιθανές πηγές δεδομένων, συνειδητοποίησαν τη σημασιολογική σημασία τους και καθόρισαν σε ποιο βαθμό τα δεδομένα ρουτίνας θα μπορούσαν να χρησιμοποιηθούν ως μέτρο της διαδικασίας ή της έκβασης της απαιτούμενης περίθαλψης, προκειμένου να επικυρωθεί η τελική οντολογία.

Η έννοια της οντολογίας περιλαμβάνει διάφορα δεδομένα / εννοιολογικά μοντέλα (π.χ. ταξινομήσεις, σχήματα βάσεων δεδομένων). Επί του παρόντος, σε ανοικτά ή εξελισσόμενα συστήματα, διαφορετικά μέρη υιοθετούν διαφορετικές οντολογίες. Έτσι, χρησιμοποιώντας μόνο οντολογίες δημιουργείται προβλήματα ετερογένειας σε υψηλότερο επίπεδο. Επομένως, καθώς μπορούν να βρεθούν διαφορετικοί τύποι οντολογιών, πρέπει να κατασκευαστούν τεχνικές αντιστοίχισης οντολογιών, με στόχο τον προσδιορισμό των αντιστοιχιών μεταξύ οντολογικών εννοιών. Η χρήση οντολογιών στους τομείς της διαλειτουργικότητας της υγειονομικής περίθαλψης δεν είναι καινούργια, καθώς έχουν διεξαχθεί αρκετές έρευνες στο πλαίσιο της εξεύρεσης οντολογικών αντιστοιχίσεων μεταξύ των διαφόρων συνιστωσών των οντολογιών στην υγειονομική περίθαλψη. Σε αυτόν τον τομέα, η έρευνα που διεξήχθη στο [115] πρότεινε το πλαίσιο ανταλλαγής μηνυμάτων Artemis για την ανταλλαγή μηνυμάτων μεταξύ HL7 v2 και HL7 v3, όπου η αντιστοίχιση οντολογιών γίνεται χειροκίνητα μετατρέποντας μηνύματα από το HL7 v2 σε OWL και τα μηνύματα OWL σε HL7 v3. Οι συγγραφείς στο [165] επιχείρησαν να ενσωματώσουν το SNOMED-CT και τις οντολογίες των ασθενειών, όπου με βάση το Unified Medical Language System (UMLS) Metathesaurus [166], υπολογίζουν τη σημασιολογική ομοιότητα με χρήση των αλγορίθμων Wu & Palmer [167] και Jiang & Conrath [168] ως μέτρο σημασιολογικής ομοιότητας. Στην έρευνα [169], οι συγγραφείς χρησιμοποίησαν οντολογίες HL7 και openEHR για να σχηματίσουν ένα σύστημα οντολογικής αντιστοίχισης εφαρμόζοντας τα εργαλεία του Falcon [170] και του Agreement Maker [171] στον τομέα της υγειονομικής περίθαλψης. Επιπλέον, οι συγγραφείς στο [172] παρείχαν έναν τρόπο ανίχνευσης και επίλυσης των συγκρούσεων επιπέδου δεδομένων που υπάρχουν στη διαδικασία αντιστοίχισης οντολογιών, επεκτείνοντας την οντολογική σημασιολογική γέφυρα για τη στήριξη της αντιστοίχισης

οντολογιών σε επίπεδο δεδομένων και δημιουργώντας τους απαιτούμενους κανόνες αντιστοίχισης ώστε να συνδυάσει δεδομένα από διαφορετικές οντολογικές πηγές σε κανονική μορφή. Επιπλέον, οι συγγραφείς στο [173] ανέπτυξαν μια εφαρμογή στην οποία δημιούργησαν πολλαπλές οντολογίες που σχετίζονται μεταξύ τους, όπου η αντιστοίχιση οντολογιών πραγματοποιείται για να επιτρέπεται η διαλειτουργικότητα μεταξύ τους. Έχουν γίνει επίσης προσπάθειες συγχώνευσης οντολογιών ιατρικής περίθαλψης με οντολογίες από διάφορους τομείς, όπως στο [174], όπου οι συγγραφείς πρότειναν μια μέθοδο συγχώνευσης τριών πολυτομεακών οντολογιών που σχετίζονται με ασθένειες, τόπους και περιβάλλοντα, στηριζόμενες στην εμπειρογνωμοσύνη και στην ανάλυση στατιστικών δεδομένων. Για το σκοπό αυτό, δεδομένου ότι είναι αναπόφευκτο να υπάρχουν παρόμοιες αλλά όχι οι ίδιες οντολογίες σε ένα συγκεκριμένο τομέα αφού μπορεί να υπάρχουν αρκετοί ορισμοί για μια δεδομένη έννοια, οι συγγραφείς στο [175] παρουσίασαν μια μέθοδο συνδυασμού μέτρων ομοιότητας διαφορετικών κατηγοριών χωρίς να έχουν οντολογικά στιγμιότυπα ή οποιοσδήποτε ανατροφοδοτήσεις από τους χρήστες σχετικά με την αντιστοίχιση των οντολογιών. Για την καλύτερη κατανόηση της τρέχουσας κατάστασης των αυτοματοποιημένων αλγορίθμων στην οντολογική ευθυγράμμιση, έχει σχεδιαστεί μια πρωτοβουλία οντολογικής αντιστοίχισης *Ontology Alignment Evaluation Initiative (OAEI)* [176] για την αξιολόγηση τους.

Επομένως, καθίσταται σαφές ότι η οντολογία που ταιριάζει στην υγειονομική περίθαλψη αποτελεί θέμα έρευνας εδώ και πολλά χρόνια, με μια ευρεία ποικιλία προσεγγίσεων. Οι κύριες διαφορές σε αυτές τις προσεγγίσεις είναι ο βαθμός αυτοματοποίησης στη διαδικασία ολοκλήρωσης, καθώς και ο αριθμός οντολογιών και τομέων που καλύπτουν. Συνεπώς, αυτό που απαιτείται είναι τεχνικές που προσφέρουν υψηλά επίπεδα αυτοματισμού, καλύπτοντας πολλαπλές οντολογίες και υποστηρίζοντας ευθυγράμμιση οντολογίας πολλών τομέων. Παρόλο που αυτές οι έρευνες καλύπτουν την πλειοψηφία των προσεγγίσεων οντολογικής αντιστοίχισης, καμία από αυτές τις τεχνικές δεν ξεπερνά το πρόβλημα της ευθυγράμμισης των οντολογιών που εφαρμόζονται σε διαφορετικές τεχνολογίες, καθώς κάθε μία από αυτές τις έρευνες είναι κατάλληλη για συγκεκριμένα προβλήματα οντολογικής αντιστοίχισης. Έτσι, η ανάγκη για μια καθολική λύση για ταυτοποίηση οντολογιών και προσδιορισμό σημασιολογίας αυξάνεται. Έτσι, είναι εξαιρετικά σημαντικό να εξεταστούν οι αλγόριθμοι αντιστοίχισης οντολογιών που καλύπτουν πολλαπλές πτυχές οντολογίας, καθώς η διαδικασία της αντιστοίχισης της οντολογίας είναι όχι μόνο δύσκολη, αλλά και χρονοβόρα και επιρρεπής στο σφάλμα.

3.2 Επεξεργασία Φυσικής Γλώσσας

3.2.1 Ανάγκη και Χρήση Επεξεργασίας Φυσικής Γλώσσας

Η επεξεργασία φυσικής γλώσσας είναι η τεχνολογία που χρησιμοποιείται για να βοηθήσει τους υπολογιστές να κατανοήσουν τη φυσική γλώσσα του ανθρώπου. Η επεξεργασία φυσικής γλώσσας, συνήθως συντομευμένη ως Natural Language Processing (NLP) [177], είναι ένας κλάδος τεχνητής νοημοσύνης που ασχολείται με την αλληλεπίδραση μεταξύ ηλεκτρονικών υπολογιστών και ανθρώπων που χρησιμοποιούν τη φυσική γλώσσα. Ο απώτερος στόχος του NLP είναι να διαβάζει, να αποκρυπτογραφεί, να κατανοεί και να έχει νόημα για τις ανθρώπινες γλώσσες με τρόπο πολύτιμο. Οι περισσότερες τεχνικές NLP βασίζονται στη μηχανική μάθηση για να αντλήσει νόημα από τις ανθρώπινες γλώσσες. Το NLP είναι η κινητήρια δύναμη πίσω από πολλαπλές κοινές εφαρμογές, όπως οι εφαρμογές μετάφρασης γλώσσας, επεξεργαστές κειμένου όπως το Microsoft Word και Grammarly που απασχολούν NLP για να ελέγχουν την γραμματική ακρίβεια των κειμένων, εφαρμογές διαδραστικής φωνητικής απόκρισης που χρησιμοποιούνται στα κέντρα κλήσεων για την ανταπόκριση στα αιτήματα ορισμένων χρηστών, ή προσωπικές εφαρμογές βοηθού όπως οι OK Google, Siri, Cortana και Alexa.

Το NLP βοηθά τους υπολογιστές να επικοινωνούν με τους ανθρώπους στη γλώσσα τους και να κλιμακώνουν άλλες εργασίες σχετικές με τη γλώσσα. Για παράδειγμα, το NLP δίνει τη δυνατότητα στους υπολογιστές να διαβάζουν κείμενο, να ακούν, να ερμηνεύουν, να μετράνε το συναίσθημα και να καθορίζουν ποια μέρη είναι σημαντικά. Οι σημερινές μηχανές μπορούν να αναλύσουν περισσότερα δεδομένα που βασίζονται στη γλώσσα από ό, τι οι άνθρωποι, χωρίς κόπωση και με συνεπή, αμερόληπτο τρόπο. Λαμβάνοντας υπόψη την εκπληκτική ποσότητα των αδόμητων δεδομένων που παράγονται καθημερινά, από τα ιατρικά αρχεία μέχρι τα κοινωνικά μέσα, η αυτοματοποίηση θα είναι κρίσιμη για την αποτελεσματική ανάλυση των δεδομένων κειμένου και λόγου.

Η ανθρώπινη γλώσσα είναι εκπληκτικά πολύπλοκη και ποικίλη. Εκφράζουμε τον εαυτό μας με άπειρους τρόπους, τόσο προφορικά όσο και γραπτά. Όχι μόνο υπάρχουν εκατοντάδες γλώσσες και διάλεκτοι, αλλά μέσα σε κάθε γλώσσα υπάρχει ένα μοναδικό σύνολο κανόνων γραμματικής, σύνταξης, και όρων. Ενώ η εποπτευόμενη και μη εποπτευόμενη μάθηση (supervised, unsupervised learning) [178] και συγκεκριμένα η βαθιά εκμάθηση (deep learning) [179] χρησιμοποιούνται ευρέως για τη μοντελοποίηση της ανθρώπινης γλώσσας,

υπάρχει επίσης ανάγκη για συντακτική και σημασιολογική κατανόηση και εμπειρογνωμοσύνη στον τομέα που δεν είναι αναγκαστικά παρόν σε αυτές τις προσεγγίσεις μηχανικής μάθησης. Το NLP είναι σημαντικό επειδή βοηθά στην επίλυση της ασάφειας στη γλώσσα και προσθέτει χρήσιμη αριθμητική δομή στα δεδομένα για πολλές εφαρμογές, όπως η αναγνώριση ομιλίας ή η ανάλυση των κειμένων.

Το NLP θεωρείται ένα δύσκολο πρόβλημα στην επιστήμη των υπολογιστών. Είναι η φύση της ανθρώπινης γλώσσας που κάνει το NLP δύσκολο. Οι κανόνες που υπαγορεύουν τη μετάδοση πληροφοριών χρησιμοποιώντας φυσικές γλώσσες δεν είναι εύκολο να κατανοηθούν από τους υπολογιστές μιας και ορισμένοι από αυτούς τους κανόνες μπορεί να είναι γενικοί και αφηρημένοι. Από την άλλη πλευρά, ορισμένοι από αυτούς τους κανόνες μπορεί να είναι χαμηλού επιπέδου. Η πλήρης κατανόηση της ανθρώπινης γλώσσας απαιτεί κατανόηση τόσο των λέξεων όσο και του τρόπου με τον οποίο συνδέονται οι έννοιες για την επίτευξη του σκοπούμενου μηνύματος. Ενώ οι άνθρωποι μπορούν εύκολα να κυριαρχήσουν μια γλώσσα, η ασάφεια και τα ασαφή χαρακτηριστικά των φυσικών γλωσσών είναι αυτό που καθιστά την NLP δύσκολη για την εφαρμογή μηχανών.

Το NLP περιλαμβάνει πολλές διαφορετικές τεχνικές για την ερμηνεία της ανθρώπινης γλώσσας, που κυμαίνονται από τις μεθόδους στατιστικής και μηχανικής μάθησης έως στις τεχνικές βασισμένες σε κανόνες και αλγοριθμικές προσεγγίσεις. Χρειάζεται μια ευρεία ποικιλία προσεγγίσεων, διότι τα δεδομένα που βασίζονται σε κείμενο και φωνή ποικίλλουν ευρέως, όπως και οι πρακτικές εφαρμογές. Οι βασικές εργασίες NLP περιλαμβάνουν ανάλυση, παραμόρφωση, προσθήκη ετικετών μερικής ομιλίας, ανίχνευση γλώσσας και αναγνώριση σημασιολογικών σχέσεων. Σε γενικές γραμμές, τα καθήκοντα του NLP κατανέμουν τη γλώσσα σε μικρότερα, στοιχειώδη κομμάτια, προσπαθούν να κατανοήσουν τις σχέσεις μεταξύ των κομματιών και να διερευνήσουν πώς τα κομμάτια συνεργάζονται για να δημιουργήσουν νόημα. Αυτά τα βασικά καθήκοντα χρησιμοποιούνται συχνά σε ικανότητες NLP υψηλότερου επιπέδου, όπως:

- **Κατηγοριοποίηση περιεχομένου.** Σύνοψη εγγράφων βασισμένη στη γλωσσολογία, συμπεριλαμβανομένης της αναζήτησης και της ευρετηρίασης, των ειδοποιήσεων περιεχομένου και της ανίχνευσης του διπλώματος.

- **Ανεύρεση και μοντελοποίηση θεμάτων.** Ακριβής συλλογή του νοήματος και των θεμάτων σε συλλογές κειμένων και εφαρμογή ανάλυσης κειμένου, όπως βελτιστοποίηση και πρόβλεψη.
- **Συγκεντρωτική εξαγωγή.** Αυτόματη εξαγωγή δομημένων πληροφοριών από πηγές που βασίζονται σε κείμενο.
- **Ανάλυση συναισθημάτων.** Προσδιορισμός της διάθεσης ή των υποκειμενικών απόψεων μέσα σε μεγάλες ποσότητες κειμένου, συμπεριλαμβανομένου του μέσου συναισθήματος και της εξόρυξης γνώμης.
- **Μετατροπή ομιλίας σε κείμενο και μετατροπή κειμένου σε ομιλία.** Αυτόματη μετατροπή φωνητικών εντολών σε γραπτό κείμενο και αντίστροφα.
- **Συνοπτική παρουσίαση εγγράφων.** Δημιουργία αυτόματων συνόψεων μεγάλων κειμένων κειμένου.
- **Μηχανική μετάφραση.** Αυτόματη μετάφραση κειμένου ή ομιλίας από τη μια γλώσσα στην άλλη.

Σε όλες αυτές τις περιπτώσεις, ο πρωταρχικός στόχος είναι να λαμβάνεται η πρωτογενής γλώσσα και να χρησιμοποιούνται κανόνες γλωσσολογίας και αλγόριθμοι για να μεταμορφώσουν ή να εμπλουτίσουν το κείμενο με τέτοιο τρόπο ώστε να αποδίδει μεγαλύτερη αξία.

Το NLP συνεπάγεται την εφαρμογή αλγορίθμων για τον εντοπισμό και την εξαγωγή των κανόνων φυσικής γλώσσας, έτσι ώστε τα δεδομένα των μη δομημένων γλωσσών να μετατρέπονται σε μια μορφή που μπορούν να κατανοήσουν οι υπολογιστές. Όταν παρέχεται το κείμενο, ο υπολογιστής θα χρησιμοποιήσει αλγόριθμους για να εξαγάγει νόημα που σχετίζεται με κάθε πρόταση και να συλλέξει τα απαραίτητα δεδομένα από αυτά. Μερικές φορές, ο υπολογιστής μπορεί να μην καταλάβει καλά την έννοια μιας φράσης, οδηγώντας σε εσφαλμένα αποτελέσματα.

Γενικότερα, το NLP συμβαδίζει με την ανάλυση των κειμένων, η οποία μετρά, ομαδοποιεί και κατηγοριοποιεί τις λέξεις για να εξάγει τη δομή και τη σημασία από μεγάλους όγκους περιεχομένου. Τα αναλυτικά στοιχεία κειμένου χρησιμοποιούνται για την εξερεύνηση του περιεχομένου κειμένου και την εξαγωγή νέων μεταβλητών από ακατέργαστο κείμενο που μπορεί να απεικονιστούν, να φιλτραριστούν ή να χρησιμοποιηθούν ως εισροές στα

μοντέλα πρόβλεψης ή σε άλλες στατιστικές μεθόδους. Το NLP και η ανάλυση κειμένων χρησιμοποιούνται μαζί για πολλές εφαρμογές, όπως:

- **Διερευνητική ανακάλυψη.** Προσδιορισμός προτύπων και ενδείξεων σε μηνύματα ηλεκτρονικού ταχυδρομείου ή γραπτές αναφορές για την ανίχνευση και επίλυση εγκλημάτων.
- **Κατηγοριοποίηση αντικειμένου.** Ταξινόμηση περιεχομένου σε σημαντικά θέματα, ώστε να μπορέσει κάποιος να αναλάβει δράση και να ανακαλύψει τάσεις.
- **Ανάλυση κοινωνικών μέσων.** Παρακολούθηση της ευαισθητοποίησης και του συναισθήματος για συγκεκριμένα θέματα και προσδιορισμός των βασικών παραγόντων επηρεασμού.

Υπάρχουν πολλές κοινές και πρακτικές εφαρμογές του NLP στην καθημερινότητά μας. Ένα υπο-πεδίο NLP που ονομάζεται Κατανόηση Φυσικής Γλώσσας (Natural Language Understanding - NLU) [180] έχει αρχίσει να αυξάνεται στη δημοτικότητα λόγω των δυνατοτήτων του σε γνωστικές εφαρμογές και εφαρμογές τεχνητής νοημοσύνης. Η NLU ξεπερνά τη δομική κατανόηση της γλώσσας για να ερμηνεύσει την πρόθεση, να επιλύσει το αμφιλεγόμενο πλαίσιο και λέξη, και να δημιουργήσει καλά μορφωμένη ανθρώπινη γλώσσα. Οι αλγόριθμοι NLU πρέπει να αντιμετωπίσουν το εξαιρετικά περίπλοκο πρόβλημα της σημασιολογικής ερμηνείας - δηλαδή, να κατανοήσουν την προτεινόμενη έννοια της ομιλούμενης ή γραπτής γλώσσας, με όλες τις λεπτομέρειες, τα συμφραζόμενα και τα συμπεράσματα που εμείς οι άνθρωποι μπορούμε να κατανοήσουμε. Η εξέλιξη του NLP προς την NLU έχει πολλές σημαντικές επιπτώσεις τόσο για τις επιχειρήσεις όσο και για τους καταναλωτές. Ωστόσο, η συντακτική ανάλυση και η σημασιολογική ανάλυση είναι οι κύριες τεχνικές που χρησιμοποιούνται για την ολοκλήρωση εργασιών φυσικής επεξεργασίας γλώσσας.

Συντακτική ανάλυση [181]

Η σύνταξη αναφέρεται στη διάταξη των λέξεων σε μια πρόταση ώστε να δημιουργούν γραμματική έννοια. Στο NLP, η συντακτική ανάλυση χρησιμοποιείται για να εκτιμήσει την ευθυγράμμιση της φυσικής γλώσσας με τους γραμματικούς κανόνες. Υπολογιστικοί αλγόριθμοι χρησιμοποιούνται για να εφαρμόσουν γραμματικούς κανόνες σε μια ομάδα λέξεων και να αντλήσουν νόημα. Ακολουθούν μερικές τεχνικές σύνταξης που μπορούν να χρησιμοποιηθούν:

- **Ενσωμάτωση εννοιών:** Περιλαμβάνει τη μείωση των διάφορων μορφών της λέξης σε μια ενιαία μορφή για εύκολη ανάλυση.
- **Μορφολογική κατάτμηση:** Περιλαμβάνει τη διαίρεση λέξεων σε μεμονωμένες μονάδες.
- **Τμηματοποίηση λέξεων:** Περιλαμβάνει τη διαίρεση ενός μεγάλου κομματιού συνεχών κειμένων σε διαφορετικές μονάδες.
- **Ετικέτα τμήματος ομιλίας:** Περιλαμβάνει τον προσδιορισμό του μέρους του λόγου για κάθε λέξη.
- **Ανάλυση:** Περιλαμβάνει την ανάληψη γραμματικής ανάλυσης για την προτεινόμενη πρόταση.
- **Διακοπή προτάσεων:** Περιλαμβάνει την τοποθέτηση ορίων φράσεων σε ένα μεγάλο κομμάτι κειμένου.
- **Εύρεση λεκτικής ρίζας:** Περιλαμβάνει την περικοπή των λέξεων που φράσσονται στη ριζική μορφή τους.

Σημασιολογική ανάλυση [181]

Η σημασιολογία αναφέρεται στην έννοια που μεταφέρεται από ένα κείμενο. Η σημασιολογική ανάλυση είναι μία από τις δύσκολες πλευρές του NLP που δεν έχει ακόμη επιλυθεί πλήρως. Περιλαμβάνει την εφαρμογή αλγορίθμων για την κατανόηση και ερμηνεία των λέξεων και του τρόπου δομής των προτάσεων. Ακολουθούν μερικές τεχνικές σημασιολογικής ανάλυσης:

- **Αναγνώριση ονομαστικής οντότητας:** Περιλαμβάνει τον προσδιορισμό των τμημάτων ενός κειμένου που μπορούν να αναγνωριστούν και να κατηγοριοποιηθούν σε προκαθορισμένες ομάδες.
- **Αποσαφήνιση αίσθησης της λέξης:** Περιλαμβάνει την απόδοση νοήματος σε μια λέξη που βασίζεται στο πλαίσιο.
- **Δημιουργία φυσικής γλώσσας:** Περιλαμβάνει τη χρήση βάσεων δεδομένων για την εξαγωγή σημασιολογικών προθέσεων και τη μετατροπή τους σε ανθρώπινη γλώσσα.

3.2.2 Τεχνικές Επεξεργασίας Φυσικής Γλώσσας στον τομέα της Υγείας

3.2.2.1 Τεχνικές Συντακτικής Ομοιότητας

Στην παρούσα στιγμή η μέτρηση συντακτικής ομοιότητας ιατρικών πληροφοριών θεωρείται ότι έχει ένα σημαντικό ρόλο στην έρευνα που σχετίζεται με το κείμενο, καθώς χρησιμοποιούνται ευρέως στην πλειοψηφία των καθηκόντων της NLP [182], συμπεριλαμβανομένης της ανάκτησης πληροφοριών, της ταξινόμησης κειμένων, της ομαδοποίησης εγγράφων, της παρακολούθησης θεμάτων, της δημιουργίας ερωτήσεων, της απάντησης σε ερωτήσεις, της βαθμολόγησης απαντήσεων, της μηχανικής μετάφρασης, της συνοπτικής παρουσίασης κειμένου και άλλα. Ο προσδιορισμός της ομοιότητας μεταξύ των συμβολοσειρών είναι ένα σημαντικό μέρος της ομοιότητας των συμβολοσειρών που στη συνέχεια χρησιμοποιείται κυρίως για τον εντοπισμό ομοιότητας κειμένου, παραγράφου και εγγράφου. Στην περίπτωση της ομοιότητας των συμβολοσειρών, δύο σειρές μπορούν να θεωρηθούν παρόμοιες όταν έχουν παρόμοια ή κοινή ακολουθία χαρακτήρων. Προκειμένου να υπολογιστεί η ομοιότητα μεταξύ δύο διαφορετικών λέξεων, έχουν αναπτυχθεί διαφορετικές τεχνικές κατά τη διάρκεια των ετών, προκειμένου να αντιμετωπιστούν οι διαφορετικές ανάγκες και εφαρμογές. Μεταξύ των πιο γνωστών τεχνικών που χρησιμοποιούνται σήμερα είναι η απόσταση Levenshtein [183], ο συντελεστής Sørensen-Dice [184], η ομοιότητα Cosine [185], η απόσταση City-Block [186], η απόσταση Jaro-Winkler [187], η ομοιότητα Jaccard [188], το ευρετήριο Tversky [189], Overlap Coefficient [190], μετρική Tau [191] και απόσταση βάσει γραμματικής ή απόσταση μετρήσεων Term Frequency - Inverse Document Frequency (TF-IDF) [192]. Με περισσότερες λεπτομέρειες, οι συγγραφείς στο [193] καταδεικνύουν μια μέθοδο ταυτοποίησης υβριδικής λογοκλοπής διερευνώντας τη χρήση μιας διαγώνιας γραμμής, η οποία προέρχεται από την απόσταση Levenshtein και τον απλοποιημένο αλγόριθμο Smith-Waterman [194], με σκοπό την εφαρμογή στην ανίχνευση λογοκλοπής. Επιπλέον, οι συγγραφείς στο [195] χρησιμοποιούν την απόσταση City-Block για να υπολογίσουν τις συναρτήσεις συνάθροισης των ασαφών συνόλων και να αναγνωρίσουν την αρχική κατάτμηση ενός συνόλου δεδομένων, προκειμένου να χειριστούν τις ελάχιστες διαφορές μεταξύ δύο προφίλ έκφρασης miRNA [196]. Στο ίδιο πλαίσιο, οι συγγραφείς στο [197] χρησιμοποιούν την ομοιότητα Cosine και Jaccard για την ανάκτηση πληροφοριών έτσι ώστε να αντιμετωπίσουν το πρόβλημα της ομοιότητας εγγράφων από ένα μεγάλο όγκο δεδομένων. Επιπλέον, στο [198] οι συγγραφείς χρησιμοποιούν διανύσματα μεταφοράς δεδομένων TF-IDF με

απόσταση Cosine για γρήγορο και αποτελεσματικό τρόπο ευθυγράμμισης των εγγράφων. Στο ίδιο πλαίσιο, οι συγγραφείς στο [199] προτείνουν την απόσταση Jaro-Winkler και το Naive Bayes Classifier [200] για να εντοπίσουν ασθένειες, λαμβάνοντας ως εισροή το κείμενο που περιέχει τα συμπτώματα της νόσου και χρησιμοποιώντας την Jaro-Winkler απόσταση για να βρουν τα συμπτώματα από την πλευρά του χρήστη.

Στα ακόλουθα υπο-Κεφάλαια, αναλύονται οι τεχνικές συντακτικής ομοιότητας που χρησιμοποιήθηκαν συγκριτικά στην παρούσα προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων.

3.2.2.1.1 Απόσταση Levenshtein

Η απόσταση Levenshtein είναι μια τεχνική μέτρησης ομοιότητας συμβολοσειρών για τον υπολογισμό της διαφοράς μεταξύ δύο ακολουθιών ή συμβολοσειρών. Η απόσταση Levenshtein μεταξύ οποιωνδήποτε δύο συμβολοσειρών υπολογίζεται ως ο ελάχιστος αριθμός επεξεργασιών ενός χαρακτήρα (δηλ. παρεμβολές, διαγραφές ή αντικαταστάσεις) που απαιτούνται για την αλλαγή μιας συμβολοσειράς στην άλλη. Η απόσταση Levenshtein μπορεί επίσης να αναφέρεται ως απόσταση επεξεργασίας, αν και μπορεί επίσης να υποδηλώνει μια μεγαλύτερη οικογένεια μετρήσεων απόστασης. Μαθηματικά, η απόσταση Levenshtein μεταξύ δύο λεκτικών a , b (μήκους $|a|$ και $|b|$ αντίστοιχα) δίνεται από το (1):

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) \\ \min \left\{ \begin{array}{l} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{array} \right. \end{cases} \quad (1)$$

Το $(a_i \neq b_j)$ είναι η συνάρτηση δείκτη ίση με 0 όταν $a_i = b_j$ και ίση με 1, και το $lev_{a,b}(i,j)$ είναι η απόσταση μεταξύ των πρώτων i χαρακτήρων του a και των πρώτων j χαρακτήρων του b . Πρέπει να σημειωθεί ότι το πρώτο στοιχείο στο ελάχιστο αντιστοιχεί στη διαγραφή (από το a έως το b), το δεύτερο στην εισαγωγή και το τρίτο για την αντιστοίχιση ή την αναντιστοίχια, ανάλογα με το αν τα αντίστοιχα σύμβολα είναι τα ίδια.

3.2.2.1.2 Ομοιότητα Cosine

Η ομοιότητα Cosine είναι μια τεχνική μέτρησης ομοιότητας συμβολοσειρών για την ποσοτικοποίηση της ομοιότητας των αλληλουχιών με την αντιμετώπισή τους ως διανύσματα και τον υπολογισμό του συνημίτονου τους. Αυτό παράγει μια τιμή μεταξύ 0 και 1, με 0 που δεν σημαίνει ομοιότητα, και 1 σημαίνει ότι και οι δύο ακολουθίες είναι ακριβώς ίδιες. Σε αυτή την περίπτωση, το πρώτο βήμα είναι να μετατρέψουμε τις συμβολοσειρές που θα συγκριθούν σε διανύσματα, λαμβάνοντας υπόψη τη συχνότητα των μοναδικών γραμμάτων σε μια δεδομένη συμβολοσειρά. Συνοπτικά, τα διανύσματα είναι γεωμετρικά στοιχεία που έχουν μέγεθος (μήκος) και κατεύθυνση. Μπορούν να εκτελεστούν πολλές μαθηματικές πράξεις, αλλά στην περίπτωση της ομοιότητας των Cosine, μόνο το γινόμενο και το μέγεθος των διανυσμάτων πρόκειται να χρησιμοποιηθούν. Το γινόμενο δύο διανυσμάτων a και b μπορεί να υπολογιστεί όπως στο (2):

$$\vec{a} * \vec{b} = a_1b_1 + a_2b_2 + \dots + a_nb_n = \sum_{i=1}^n a_ib_i \quad (2)$$

Το μέγεθος αυτών των δύο διανυσμάτων μπορεί να υπολογιστεί όπως στο (3):

$$\|\vec{a}\| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2} \quad (3)$$

Συνεπώς, για τον μαθηματικό υπολογισμό της ομοιότητας Cosine αυτών των διανυσμάτων, το (4) χρησιμοποιείται όπου το γινόμενο των διανυσμάτων a και b διαιρείται με το γινόμενο του μεγέθους a και μεγέθους b .

$$\text{similarity}(\vec{a}, \vec{b}) = \cos\theta = \frac{\vec{a} * \vec{b}}{\|\vec{a}\| * \|\vec{b}\|} \quad (4)$$

3.2.2.1.3 Ομοιότητα Jaccard

Η ομοιότητα Jaccard είναι μια τεχνική μέτρησης ομοιότητας συμβολοσειρών που συγκρίνει τα μέλη για δύο συμβολοσειρές για να δει ποια μέλη μοιράζονται και ποια είναι διαφορετικά. Είναι ένα μέτρο ομοιότητας για τα δύο λεκτικά σύνολα, με ένα εύρος από 0% έως 100%. Όσο υψηλότερο είναι το ποσοστό, τόσο πιο παρόμοια είναι αυτά τα δύο λεκτικά. Στην περίπτωση της ομοιότητας Jaccard, το (5) χρησιμοποιείται για δύο λεκτικά a και b :

$$J(a, b) = \frac{|a \cap b|}{|a \cup b|} \quad (5)$$

Το $|a \cap b|$ παρέχει τον αριθμό των μελών που μοιράζονται μεταξύ των δύο σειρών συμβολοσειρών και $|a \cup b|$ παρέχει τον συνολικό αριθμό των μελών και στα δύο σύνολα συμβολοσειρών. Η κατανομή του αριθμού των μελών με το συνολικό αριθμό των μελών, παρέχει την Jaccard ομοιότητα τους.

3.2.2.1.4 Ομοιότητα Jaro-Winkler

Η απόσταση Jaro-Winkler είναι μια τεχνική μέτρησης ομοιότητας συμβολοσειρών για την ποσοτικοποίηση της ομοιότητας μεταξύ δύο συμβολοσειρών. Όσο υψηλότερη είναι η απόσταση Jaro-Winkler για δύο λεκτικά, τόσο πιο ταυτόσημα είναι. Η μέτρηση της απόστασης Jaro-Winkler έχει καλύτερη απόδοση για σύντομες συμβολοσειρές, ενώ στην περίπτωση που η βαθμολογία ομοιότητας είναι περίπου 0, σημαίνει ότι δεν υπάρχει ομοιότητα, ενώ το 1 σημαίνει ότι υπάρχει ακριβής αντιστοιχία. Η ομοιότητα Jaro-Winkler δίνεται από το (6):

$$dw = dj + (lp(1 - dj)) \quad (6)$$

Σε αυτή την περίπτωση, πρέπει να υπολογιστεί αρχικά η Jaro απόσταση dj , όπου δοσμένων δύο λεκτικών $s1$ και $s2$, το dj δίνεται από το (7):

$$dj = \begin{cases} 0 \\ \frac{1}{3} \left(\frac{m}{|s1|} + \frac{m}{|s2|} + \frac{m-t}{m} \right) \end{cases} \quad (7)$$

Σε αυτή την περίπτωση, τα $s1$ και $s2$ είναι ο συνολικός αριθμός χαρακτήρων κάθε συμβολοσειράς αντίστοιχα, m είναι ο αριθμός των αντιστοιχων χαρακτήρων και t είναι ο μισός αριθμός μεταθέσεων. Επιπλέον, το l είναι το μήκος του κοινού προθέματος στην αρχή της συμβολοσειράς, μέχρι και τέσσερις χαρακτήρες, ενώ το p είναι ένας σταθερός συντελεστής κλιμάκωσης για το πόσο το τελικό αποτέλεσμα πρέπει να ρυθμιστεί προς τα πάνω για να έχει κοινά προθέματα. Γενικά, το p δεν πρέπει να υπερβαίνει το 0.25, διαφορετικά η απόσταση μπορεί να γίνει μεγαλύτερη από 1. Η τυπική τιμή για αυτή τη σταθερά του Winkler είναι $p = 0,1$.

3.2.2.2 Τεχνικές Σημασιολογικής Ομοιότητας

Αρκετές μέθοδοι για τον προσδιορισμό της σημασιολογικής ομοιότητας μεταξύ όρων έχουν προταθεί στη βιβλιογραφία. Οι μέθοδοι σημασιολογικής ομοιότητας κατατάσσονται σε τέσσερις κύριες κατηγορίες [201]:

- **Μέθοδοι καταμέτρησης άκρων** αναφορικά με την ομοιότητα μεταξύ δύο όρων (εννοιών) ως συνάρτηση του μήκους της διαδρομής που συνδέει τους όρους και τη θέση των όρων στην ταξινόμια.
- **Μέθοδοι περιεχομένου πληροφοριών** αναφορικά με τη μέτρηση της διαφοράς στο περιεχόμενο πληροφοριών μεταξύ δύο όρων ως συνάρτηση της πιθανότητας εμφάνισής τους σε ένα σώμα.
- **Μέθοδοι βασισμένες σε χαρακτηριστικά** αναφορικά με τη μέτρηση ομοιότητας μεταξύ δύο όρων ως συνάρτηση των ιδιοτήτων τους (π.χ. των ορισμών τους ή των γλωσσών στο WordNet [202] ή σημειώσεις περι εύρους στο MeSH [203]) ή με βάση των σχέσεων τους με άλλους παρόμοιους όρους στην ταξινόμια διάρθρωση. Τα κοινά χαρακτηριστικά τείνουν να αυξάνουν την ομοιότητα και (αντίθετα) τα μη κοινά χαρακτηριστικά τείνουν να μειώνουν την ομοιότητα δύο εννοιών.
- Οι **υβριδικές μέθοδοι** συνδυάζουν τις παραπάνω ιδέες όπου η ομοιότητα των όρων υπολογίζεται από τα συνώνυμα που αντιστοιχούν, τους γειτονικούς όρους και τα χαρακτηριστικά ορολογιών. Τα χαρακτηριστικά ορολογιών διακρίνονται περαιτέρω σε μέρη, λειτουργίες και ιδιότητες.

Μια σημαντική παρατήρηση των μεθόδων σημασιολογικής ομοιότητας είναι ότι αποδίδουν υψηλότερη ομοιότητα με τους όρους που είναι στενοί μεταξύ τους (από την άποψη του μήκους της διαδρομής) και χαμηλότεροι στην ιεραρχία (πιο συγκεκριμένοι όροι), και από τους όρους που είναι εξίσου στενοί μαζί αλλά υψηλότερη στην ιεραρχία (γενικότεροι όροι).

Οι μέθοδοι σημασιολογικής ομοιότητας μπορούν επίσης να διακριθούν σε:

- **Μέθοδοι ομοιότητας μοναδικής οντολογίας** που υποθέτουν ότι οι όροι που συγκρίνονται είναι από την ίδια οντολογία (π.χ. WordNet).
- **Μέθοδοι ομοιότητας διασταύρωσης οντολογίας**, οι οποίες συγκρίνουν τους όρους από διαφορετικές οντολογίες (π.χ. MeSH). Οι μέθοδοι

καταμέτρησης περιεχομένου πληροφοριών λειτουργούν αξιοποιώντας πληροφορίες δομής (δηλ. δέση όρων) και περιεχομένου πληροφοριών όρων σε μια ιεραρχία, και είναι καταλληλότερες για τη σύγκριση όρων από την ίδια οντολογία. Επειδή η δομή και το πληροφοριακό περιεχόμενο διαφορετικών οντολογιών δεν είναι άμεσα συγκρίσιμα, οι μέθοδοι ομοιότητας διασταύρωσης οντολογίας συνήθως απαιτούν υβριδικές προσεγγίσεις ή προσεγγίσεις που βασίζονται σε συγκεκριμένα χαρακτηριστικά.

Στα ακόλουθα υπο-Κεφάλαια, αναλύονται οι τεχνικές σημασιολογικής ομοιότητας που χρησιμοποιήθηκαν συγκριτικά στην προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων.

3.2.2.2.1 *Word2Vec*

Το Word2Vec [204] είναι ένα νευρωνικό δίκτυο δύο επιπέδων που επεξεργάζεται το κείμενο. Η είσοδος του είναι ένα σώμα κειμένου και η έξοδος του είναι ένα σύνολο διανυσμάτων. Ενώ το Word2Vec δεν είναι ένα βαθύ νευρωνικό δίκτυο, μετατρέπει το κείμενο σε αριθμητική μορφή, ώστε να μπορεί εύκολα να κατανοηθεί. Το DeepLearning4j [205] υλοποιεί μια κατανεμημένη μορφή Word2Vec για Java και Scala [206, η οποία λειτουργεί με Spark [207].

Οι εφαρμογές του Word2Vec εκτείνονται πέρα από την ανάλυση απλών προτάσεων. Μπορεί να εφαρμοστεί εξίσου καλά με τα γονίδια, τον κώδικα, τις προτιμήσεις, τις λίστες αναπαραγωγής, τα γραφήματα των κοινωνικών μέσων και άλλες λεκτικές ή συμβολικές σειρές στις οποίες μπορεί να γίνει διάκριση. Ο σκοπός και η χρησιμότητα του Word2Vec είναι η ομαδοποίηση των φορέων παρόμοιων λέξεων μαζί στο διανυσματικό χώρο, ανιχνεύοντας μαθηματικά ομοιότητες. Το Word2Vec δημιουργεί διανύσματα που είναι κατανεμημένες αριθμητικές παραστάσεις χαρακτηριστικών λέξεων, χαρακτηριστικά όπως το πλαίσιο μεμονωμένων λέξεων, χωρίς ανθρώπινη παρέμβαση. Λαμβάνοντας υπόψη αρκετά δεδομένα, τη χρήση του και το τρέχον περιβάλλον, το Word2Vec μπορεί να κάνει πολύ ακριβείς εικασίες σχετικά με το νόημα μιας λέξης με βάση τις εμφανίσεις της τρέχουσας λέξης σε άλλες καταστάσεις στο παρελθόν. Αυτές οι εικασίες μπορούν να χρησιμοποιηθούν για να εδραιωθεί η συσχέτιση μιας λέξης με άλλες λέξεις ή έγγραφα συμπλέγματος και να τα ταξινομήσει κατά θέμα. Αυτά τα συμπλέγματα μπορούν να αποτελέσουν τη βάση αναζήτησης, ανάλυσης συναισθημάτων και συστάσεων σε διάφορους τομείς όπως η επιστημονική έρευνα, η νομική ανακάλυψη, το ηλεκτρονικό εμπόριο και η διαχείριση των πελατειακών σχέσεων. Η έξοδος του νευρωνικού

δικτύου Word2Vec είναι ένα λεξιλόγιο στο οποίο κάθε στοιχείο έχει έναν φορέα προσαρτημένο σε αυτόν, ο οποίος μπορεί να τροφοδοτηθεί σε ένα πλαίσιο βαθιάς εκμάθησης ή απλά να ερωτηθεί για να ανιχνεύσει τις σχέσεις μεταξύ των λέξεων.

Τα διανύσματα που χρησιμοποιούμε για να αναπαριστούμε λέξεις ονομάζονται Νευρωνικές Ενσωματώσεις Λέξεων όπου ένα αντικείμενο περιγράφει ένα άλλο, παρόλο που τα δύο αυτά αντικείμενα είναι ριζικά διαφορετικά. Το Word2Vec μετατρέπει τις λέξεις σε διανύσματα, και με αυτόν τον τρόπο κάνει τη φυσική γλώσσα ευανάγνωστη από τον υπολογιστή. Το Word2Vec είναι παρόμοιο με έναν αυτόματο κωδικοποιητή, που κωδικοποιεί κάθε λέξη σε ένα διάνυσμα, αλλά αντί να εκπαιδεύει τις λέξεις εισόδου μέσω της ανασυγκρότησης, όπως κάνει μια περιορισμένη μηχανή Boltzmann [208], το Word2Vec εκπαιδεύει τις λέξεις ενάντια σε άλλες γειτονικές τους λέξεις. Αυτό γίνεται με έναν από τους δύο τρόπους, είτε χρησιμοποιώντας το πλαίσιο για να προβλέψει μια λέξη-στόχο (Continuous Bag of Words - CBOW [209]), είτε χρησιμοποιώντας μια λέξη για να προβλέψει ένα πλαίσιο στόχο, το οποίο ονομάζεται skip-gram [210]. Συνήθως, χρησιμοποιείται η τελευταία μέθοδος επειδή παράγει ακριβέστερα αποτελέσματα σε μεγάλα σύνολα δεδομένων.

3.2.2.2.2 *WordNet*

Το WordNet [202] είναι μια μεγάλη λεξική βάση αγγλικών. Τα ουσιαστικά, τα ρήματα, τα επίθετα και τα επιρρήματα ομαδοποιούνται σε σύνολα γνωστικών συνωνύμων (synsets), έκαστο των οποίων εκφράζει μια ξεχωριστή έννοια. Τα synsets αλληλοσυνδέονται μέσω εννοιολογικά-σημασιολογικών και λεξικών σχέσεων. Το προκύπτον δίκτυο συναφών λέξεων και εννοιών μπορεί να πλοηγηθεί με το πρόγραμμα περιήγησης. Η δομή του WordNet το καθιστά ένα χρήσιμο εργαλείο για την υπολογιστική γλωσσολογία και το NLP. Το WordNet μοιάζει επιφανειακά με έναν θησαυρό, επειδή ομαδοποιεί τις λέξεις με βάση τις έννοιές τους. Ωστόσο, υπάρχουν κάποιες σημαντικές διακρίσεις. Πρώτον, το WordNet συνδέει όχι μόνο τις μορφές λέξεων - τις συμβολοσειρές των γραμμάτων - αλλά τα συγκεκριμένα νοήματα των λέξεων. Ως αποτέλεσμα, οι λέξεις που βρίσκονται σε άμεση γειτνίαση μεταξύ τους στο δίκτυο είναι σημασιολογικά αποσαφηνισμένες. Δεύτερον, το WordNet επισημαίνει τις σημασιολογικές σχέσεις μεταξύ των λέξεων, ενώ οι ομαδοποιήσεις λέξεων σε έναν θησαυρό δεν ακολουθούν κανένα ρητό πρότυπο εκτός από την έννοια της ομοιότητας. Η κύρια σχέση μεταξύ των λέξεων στο WordNet είναι η συνωνυμία. Τα συνώνυμα - λέξεις που υποδηλώνουν την ίδια έννοια και είναι εναλλάξιμες σε πολλά πλαίσια - ομαδοποιούνται σε μη οριζόντια σύνολα.

Επιπλέον, ένα synset περιέχει έναν σύντομο ορισμό (gloss) και, στις περισσότερες περιπτώσεις, μία ή περισσότερες σύντομες προτάσεις που απεικονίζουν τη χρήση των μελών synset. Οι φόρμες λέξεων με αρκετές διακριτές έννοιες αντιπροσωπεύονται σε διαφορετικές συνόψεις. Έτσι, κάθε ζεύγος μορφών-σημαδιών στο WordNet είναι μοναδικό.

Η πιο συχνά κωδικοποιημένη σχέση μεταξύ των synsets είναι η υπερ-δευτερεύουσα σχέση (που ονομάζεται επίσης υπερτονία (hyperonymy), υποτονία (hyponymy) ή είναι-ένα (IS-A) σχέση). Συνδέει γενικότερα synsets, με ολόένα και πιο συγκεκριμένες λέξεις. Έτσι, για παράδειγμα το WordNet δηλώνει ότι τα έπιπλα μιας κατηγορίας περιλαμβάνουν το κρεβάτι, το οποίο με τη σειρά του περιλαμβάνει κουκέτες. Αντίστροφα, οι έννοιες όπως το κρεβάτι και η κουκέτα συνθέτουν τα έπιπλα της κατηγορίας. Όλες οι ιεραρχίες ουσιαστικά ανεβαίνουν στον κόμβο ρίζας (οντότητα). Η σχέση hyponymy είναι μεταβατική: εάν μια πολυθρόνα είναι ένα είδος καρέκλας και αν μια καρέκλα είναι ένα είδος επίπλου, τότε μια πολυθρόνα είναι ένα είδος επίπλου. Το WordNet διακρίνει μεταξύ των τύπων (κοινά ονόματα) και των περιπτώσεων (συγκεκριμένα άτομα, χώρες και γεωγραφικές οντότητες). Έτσι, η πολυθρόνα είναι ένα είδος καρέκλας, ο Μπαράκ Ομπάμα είναι μια περίπτωση ενός προέδρου. Οι περιπτώσεις είναι πάντα κόμβοι στις ιεραρχίες τους.

Τα ρήματα των synsets διευθετούνται σε ιεραρχίες, με τα ρήματα προς το κάτω μέρος των δέντρων να εκφράζουν όλο και πιο συγκεκριμένους τρόπους που χαρακτηρίζουν ένα γεγονός. Ο συγκεκριμένος τρόπος έκφρασης εξαρτάται από το σημασιολογικό πεδίο, όπου για παράδειγμα ο όγκος είναι μια μόνο διάσταση κατά την οποία μπορούν να επεξεργαστούν τα ρήματα. Άλλα είναι η ταχύτητα ή η ένταση του συναισθήματος. Τα ρήματα που περιγράφουν γεγονότα που συνδέονται αναγκαστικά και μονοκατευθυντικά, ενσωματώνονται μεταξύ τους. Τα επίθετα είναι οργανωμένα με όρους αντωνυμίας, όπου τα ζευγάρια των «άμεσων» αντωνυμιών, αντανakλούν την ισχυρή σημασιολογική σύμβαση των μελών τους. Υπάρχουν μόνο λίγα επιρρήματα στο WordNet (δύσκολα, κυρίως, πραγματικά, κλπ.) καθώς η πλειοψηφία των αγγλικών επιρρημάτων προέρχεται άμεσα από επίθετα μέσω μορφολογικής προσκόλλησης.

3.2.2.2.3 *Cortical.io*

Το Cortical.io [211] προτείνει έναν νέο τρόπο για την κατανόηση της φυσικής γλώσσας που ξεπερνά τους περιορισμούς άλλων προσεγγίσεων τεχνητής νοημοσύνης. Το Semantic Folding [212] εμπνέεται από τα πιο πρόσφατα ευρήματα σχετικά με τον τρόπο με τον οποίο ο εγκέφαλος επεξεργάζεται τις πληροφορίες. Αυτή η νέα θεωρία εισάγει μια νέα αναπαράσταση δεδομένων, το Σημασιολογικό Δακτυλικό Αποτύπωμα (Semantic Fingerprint), το οποίο κωδικοποιεί ρητά την έννοια, συμπεριλαμβανομένων όλων των εννοιών και των πλαισίων. Το σύστημα κατανοεί τη συγγένεια δύο στοιχείων μετρώντας την επικάλυψη των δακτυλικών τους αποτυπωμάτων. Ως αποτέλεσμα, είναι πολύ γρήγορο, αξιόπιστο και εύκολο στην εφαρμογή - μια πρωτοποριακή τεχνολογία που εκμεταλλεύεται τη νοημοσύνη του εγκεφάλου για να καταστήσει δυνατή την κατανόηση φυσικών γλωσσών σε δεδομένα μεγάλου κειμένου.

Αρχικά, δημιουργείται ένας σημασιολογικός χώρος που ονομάζεται βάση δεδομένων Retina μέσω της μη επιτηρούμενης μάθησης του υλικού αναφοράς (για παράδειγμα, συμβόλαια, ιατρικά εγχειρίδια, οικονομικά έγγραφα, αιτήματα υποστήριξης). Μια βάση δεδομένων Retina μπορεί να εκπαιδευτεί με διαφορετικές συλλογές κειμένου για να ειδικευτεί σε συγκεκριμένα θέματα ή σε συγκεκριμένο τομέα γλώσσας. Στο επόμενο βήμα, το Cortical.io χρησιμοποιείται για τη μετατροπή του κειμένου σε σημασιολογικά δακτυλικά αποτυπώματα, δηλαδή αριθμητικές παραστάσεις που καταγράφουν το νόημα πίσω από τη φυσική γλώσσα. Σημασιολογικά δακτυλικά αποτυπώματα μπορούν να δημιουργηθούν για γλωσσικά στοιχεία όπως λέξεις, ή προτάσεις και ολόκληρα έγγραφα. Οποιαδήποτε δύο κομμάτια κειμένου μπορούν να συγκριθούν, ανεξάρτητα από το μήκος ή τη γλώσσα. Ενώ τα παραδοσιακά συστήματα NLP βασίζονται σε υπολογισμούς συχνότητας λέξεων, το Cortical.io χρησιμοποιεί μια ουσιαστικά λεπτότερη αναπαράσταση για κάθε λέξη: 16.000 σημασιολογικά χαρακτηριστικά καταγράφονται για κάθε όρο. Τα σημασιολογικά δακτυλικά αποτυπώματα επιτρέπουν την άμεση σύγκριση των σημασιών οποιωνδήποτε δύο κειμένων, που δείχνουν χιλιάδες σημασιολογικές σχέσεις.

Με το Semantic Folding, οι σημασιολογικοί χώροι είναι σταθεροί στις διάφορες γλώσσες, επιτρέποντας την άμεση σύγκριση του κειμένου σε όλες τις γλώσσες χωρίς μηχανική μετάφραση. Τα σημασιολογικά δακτυλικά αποτυπώματα κωδικοποιούνται με τη μορφή των Sparse Distributed Representations (SDRs) [213] ως μια δομή δεδομένων που αποτελείται από ένα μεγάλο αριθμό

επιμέρους δυαδικών ψηφίων, καθένα από τα οποία μπορεί να ενεργοποιηθεί ή να απενεργοποιηθεί. Η έννοια ενός δακτυλικού αποτυπώματος καθορίζεται από τη συμπεριφορά αυτών των δυαδικών ψηφίων, με κάθε bit να συνεισφέρει μια μικρή ποσότητα στο συνολικό νόημα. Η Σημασιολογική Πτυσσόμενη Θεωρία (Semantic Folding Theory) είναι χτισμένη πάνω από την Ιεραρχική Θεωρητική Θεωρία Μνήμης. Η προσέγγιση της θεωρίας αυτής στην κατανόηση του τρόπου με τον οποίο λειτουργεί η επεξεργασία πληροφορίας, ενώ παραμένει στενά συσχετισμένη με τα βιολογικά δεδομένα, είναι κάπως διαφορετική από τα πιο επικρατέστερα έργα που έχουν μια κυρίως ανατομική ή κυρίως λειτουργική προσέγγιση αντιστοιχίας. Οι νευροεπιστήμονες που εργάζονται σε μικροανατομικά μοντέλα έχουν αναπτύξει εξελιγμένες τεχνικές για την παρακολούθηση της πραγματικής δομής 3D του νευρικού πλέγματος του φλοιού μέχρι το μικροσκοπικό επίπεδο των δενδριτών, των αξόνων και των συνάψεών τους. Αυτό επιτρέπει τη δημιουργία ενός πλήρους και ακριβούς χάρτη όλων των νευρώνων και των διασυνδέσεών τους στον εγκέφαλο. Με αυτό το διάγραμμα καλωδίωσης ελπίζουν να καταλάβουν τη λειτουργία των εγκεφάλων. Η έρευνα σε λειτουργική αντιστοιχία, από την άλλη πλευρά, έχει αναπτύξει πολύ προηγμένα μοντέλα απεικόνισης και υπολογιστικής μεθόδου για τον προσδιορισμό του τρόπου διασύνδεσης των διαφόρων επιθεμάτων του φλοιού ιστού για τη δημιουργία λειτουργικών οδών. Με την πλήρη απογραφή όλων των υφιστάμενων οδών και τις λειτουργικές περιγραφές τους, οι επιστήμονες ελπίζουν να αποκαλύψουν τη γενική αρχιτεκτονική πληροφόρησης του εγκεφάλου. Σε αντίθεση με αυτές τις πρωτογενείς προσεγγίσεις που βασίζονται σε δεδομένα, για την προαναφερθείσα θεωρία κάθε χαρακτηριστικό που προσδιορίζεται μπορεί να συνδυαστεί με στοιχεία από τη νευροανατομική, νευροφυσιολογική και συμπεριφορική έρευνα. Η Semantic Folding Theory προσπαθεί να ικανοποιήσει όλους τους περιορισμούς που καθορίζονται για την τρέχουσα μηχανική μάθηση ενώ παραμένει βιολογικώς εύλογη και εξηγεί όσο το δυνατόν περισσότερα χαρακτηριστικά της ανθρώπινης γλώσσας. Αυτό επιτυγχάνεται με την πρόταση μιας νέας προσέγγισης στο πρόβλημα της αναπαραστάσεως, δηλαδή την ικανότητα να αντιπροσωπεύει το νόημα με τρόπο που να υπολογίζεται από την υποδομή επεξεργασίας του Cortical.io. Η δυνατότητα επεξεργασίας των γλωσσικών πληροφοριών στο επίπεδο της έννοιας τους θα επιτρέψει την καλύτερη κατανόηση της φύσης της νοημοσύνης, ένα φαινόμενο στενά συνδεδεμένο με την ανθρώπινη γλώσσα.

ΚΕΦΑΛΑΙΟ 4

Προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων

Περιεχόμενα

4.1. Αρχικά Στάδια Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων

4.1.1. Στάδιο Μεταφοράς Γνώσεων συγκεκριμένου τομέα σε έναν άγνωστο τομέα, κατά μήκος του Κύκλου Ζωής των Δεδομένων

4.1.2. Στάδιο Συσσωμάτωσης Ετερογενών δεδομένων υγείας μέσω μίας Οντολογικής Κοινής Γλώσσας Υγείας

4.1.3. Στάδιο απόκτησης της Οντολογικής Αναπαράστασης των Δεδομένων της Υγείας μέσω των Τεχνικών Μεταμοντέλων

4.1.4. Στάδιο Οντολογικού Αντιστοιχιστή FHIR (FHIR Ontology Mapper - FOM): Συγκέντρωση Συντακτικών και Σημασιολογικών Ομοιοτήτων Οντολογιών προς την αντιστοιχισή τους σε HL7 FHIR

4.2. Προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων

4.2.1. Αλγόριθμος Οντολογιών και Συσχετίσεων

4.2.2. Γνωσιακή Βάση Δεδομένων

4.2.3. Αλγόριθμος Συσχέτισης Δομής

4.2.4. Τεχνική Μετάφρασης Δομής FHIR

4.3. Ολιστική Λύση Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων

Περίληψη

Στο **ΚΕΦΑΛΑΙΟ 4**, αναλύεται η παρούσα προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων, όπου αρχικά πραγματοποιείται αναφορά στις πιο κομβικές προσωπικές έρευνες στο πλαίσιο της Διδακτορικής Διατριβής, σε συνδυασμό με τα αποτελέσματά τους. Ακολούθως, περιγράφεται κάθε τεχνική της παρούσας προσέγγισης, αναφορικά με τη δημιουργία των οντολογιών, την αποθήκευσή τους, τη συντακτική και σημασιολογική αντιστοιχισή τους, καθώς και την τελική απεικόνισή τους. Το σύνολο αυτών των τεχνικών οδηγεί εν τέλει στη λειτουργικότητα της προσέγγισης, και την τελική αντιστοιχισή των δεδομένων υγείας στο πρότυπο υγείας HL7 FHIR.

4.1 Αρχικά Στάδια Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων

4.1.1 Στάδιο Μεταφοράς Γνώσεων συγκεκριμένου τομέα σε έναν άγνωστο τομέα, κατά μήκος του Κύκλου Ζωής των Δεδομένων

Η προτεινόμενη προσέγγιση επικεντρώθηκε στο πώς είναι δυνατόν να συγκεντρωθούν δεδομένα από μια συγκεκριμένη πηγή, και στο να κατανοηθεί πώς εκτελείται το τελευταίο στον κύκλο ζωής του. Κατόπιν, μελετήθηκε το πώς χρησιμοποιεί αυτό το σύνολο δεδομένων την εξαγόμενη γνώση σε συνδυασμό με ακατέργαστα δεδομένα που προέρχονται από διαφορετική πηγή, για διαφορετικό λόγο από αυτόν για τον οποίο τα ακατέργαστα δεδομένα δημιουργήθηκαν. Συνοπτικά, η παρούσα προσέγγιση αρχικά μελέτησε πώς εξελισσόμενοι από ένα δίκτυο διασυνδεδεμένων υπολογιστών και συσκευών δικτύωσης, όλο και περισσότερο προσεγγίζεται το διαδίκτυο ως Διαδίκτυο των Πάντων (Internet of Everything - IoE) [214], συμπεριλαμβάνοντας τις διασυνδέσεις συσκευών με διαφορετικά χαρακτηριστικά, προσφορές υπηρεσιών και ροές δεδομένων (π.χ. έξυπνες συσκευές, φορητές συσκευές, δίκτυα αισθητήρων, κάμερες, οχήματα κ.λπ.). Σαν αποτέλεσμα, το σημερινό τοπίο αντικατοπτρίζει ένα πλούσιο ψηφιακό περιβάλλον με προβλέψεις που αναφέρουν 21 δισεκατομμύρια συνδεδεμένες συσκευές έως το 2020 [215]. Στο πλαίσιο αυτό, το IoT εξελίσσεται σε μια πλατφόρμα πλατφόρμων, που αποτελείται από πολλαπλές συσκευές, υποδομές και υποστηρικτικά συστήματα συστημάτων.

Η απόκτηση πληροφοριών και επεξεργασίας αποτελεί ήδη μια πρόκληση που αντιπροσωπεύει ένα τεράστιο δυναμικό σε πολλούς τομείς όπως η μεταφορά, η ενέργεια, το περιβάλλον, η υγεία, κλπ. Υπάρχουν, για παράδειγμα, πολλές ιστορίες επιτυχίας για την επεξεργασία των δεδομένων που προέρχονται από το IoT. Αλλά τα δεδομένα μπορεί να είναι άχρηστα αν δεν μπορούν να «κατανοηθούν» και αν δεν μπορούν να αξιοποιηθούν σε διαφορετικούς τομείς εφαρμογών. Ο σημασιολογικός σχολιασμός συνδέεται συνήθως με τα δεδομένα, αλλά υπάρχουν πολλοί τρόποι επίτευξης της προαναφερθείσας εκμετάλλευσης και διαλειτουργικότητας σε όλους τους τομείς, γεγονός που αποτελεί βασική πρόκληση για το IoT και τις υψηλές απαιτήσεις διασύνδεσης ετερογενών αντικειμένων (συστημάτων, συστατικών κλπ.). Το αρχικό στάδιο της παρούσας προσέγγισης στόχευε στην αντιμετώπιση της σημασιολογικής πλευράς της διαλειτουργικότητας του διαδικτύου μέσω της μοντελοποίησης και της εκπροσώπησης της γνώσης.

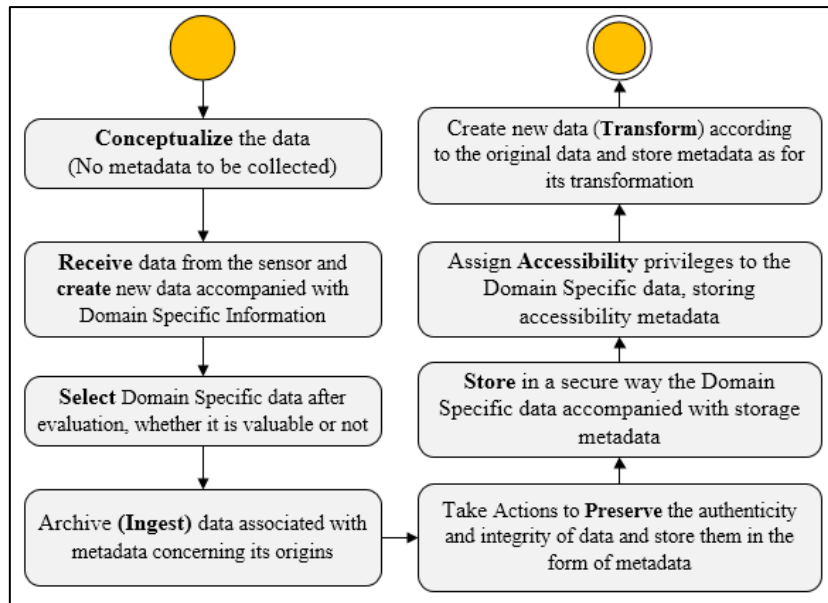
Το ΙοΤ έχει πολλούς ορισμούς [216][217] από διαφορετικούς οργανισμούς. Ωστόσο, ένα κοινό χαρακτηριστικό του ΙοΤ έχει να κάνει με τη διασυνδεσιμότητα των αντικειμένων. Σε κάθε περίπτωση, όλα πρέπει να συνδεθούν με το διαδίκτυο για να επικοινωνήσουν με άλλα αντικείμενα. Είναι μια αναδυόμενη περιοχή που απαιτεί όχι μόνο την ανάπτυξη της υποδομής αλλά και την ανάπτυξη νέων υπηρεσιών ικανών να υποστηρίξουν πολλαπλές, κλιμακούμενες και διαλειτουργικές εφαρμογές. Στο πλαίσιο του σχεδιασμού του ΙοΤ ως μέρος της μελλοντικής αρχιτεκτονικής διαδικτύου, οι ακαδημαϊκές κοινότητες και οι βιομηχανίες ΤΠΕ έχουν συνειδητοποιήσει ότι ένα κοινό πρόβλημα που πρέπει να αντιμετωπιστεί είναι η διαλειτουργικότητα των πληροφοριών και των υπηρεσιών.

Σήμερα, η πλειοψηφία των συσκευών έχει σχεδιαστεί για να ενσωματώνει έναν ή περισσότερους αισθητήρες. Αυτοί οι αισθητήρες έχουν τη δυνατότητα να χρησιμοποιούν τα δεδομένα ως είσοδο, προκειμένου να βελτιώσουν τη λήψη αποφάσεων αυτών των συσκευών ή να χρησιμοποιήσουν αυτά τα δεδομένα ως είσοδο σε διαφορετικό αισθητήρα του ίδιου τομέα για περαιτέρω χρήση. Ένα απλό παράδειγμα θα μπορούσε να είναι ο αισθητήρας του θερμομέτρου ενός αυτοκινήτου που λαμβάνει ως είσοδο την εξωτερική θερμοκρασία και επεξεργάζεται αυτές τις πληροφορίες με αποτέλεσμα την προειδοποίηση του οδηγού του αυτοκινήτου ότι υπάρχει πιθανότητα να συναντήσει παγωμένο δρόμο. Για το προαναφερθέν πρόβλημα έχουν συζητηθεί αρκετές προσεγγίσεις, αλλά πώς είναι δυνατόν να χρησιμοποιηθούν τα δεδομένα του αισθητήρα του θερμομέτρου του αυτοκινήτου ως είσοδος στον αισθητήρα ενός σταθμού πρόγνωσης καιρού για την ενημέρωση άλλων οδηγών για την παράκαμψη αυτού του συγκεκριμένου δρόμου λόγω της πιθανότητας πάγου;

Για το σκοπό αυτό, η αρχική προσέγγιση πρότεινε μια λύση με την αρχική ανασκόπηση των πρόσφατων τάσεων και προκλήσεων στη διαλειτουργικότητα, σχετικά με το είδος των σημασιολογικών τεχνολογιών, προσεγγίσεων και βέλτιστων πρακτικών που έχουν αναπτυχθεί και που μπορούν να υποστηρίξουν τη διαλειτουργικότητα των δεδομένων στο ΙοΕ. Ειδικότερα, προτάθηκε μια πιθανή προσέγγιση για την επίτευξη σημασιολογικής διαλειτουργικότητας μεταξύ διαφορετικών δεδομένων αισθητήρων (π.χ. μεταφορά, υγεία κλπ.) σε συνδυασμό με ένα συγκεκριμένο σενάριο όπου συζητήθηκε ο τρόπος με τον οποίο τα συγκεκριμένα δεδομένα μπορούν να μεταφραστούν σε αγνωστικό τομέα σε συνδυασμό με μη ετικετοποιημένα δεδομένα, έτσι ώστε αυτά να χρησιμοποιηθούν περαιτέρω, κερδίζοντας τη μέγιστη δυνατή χρήση και γνώση από αυτό.

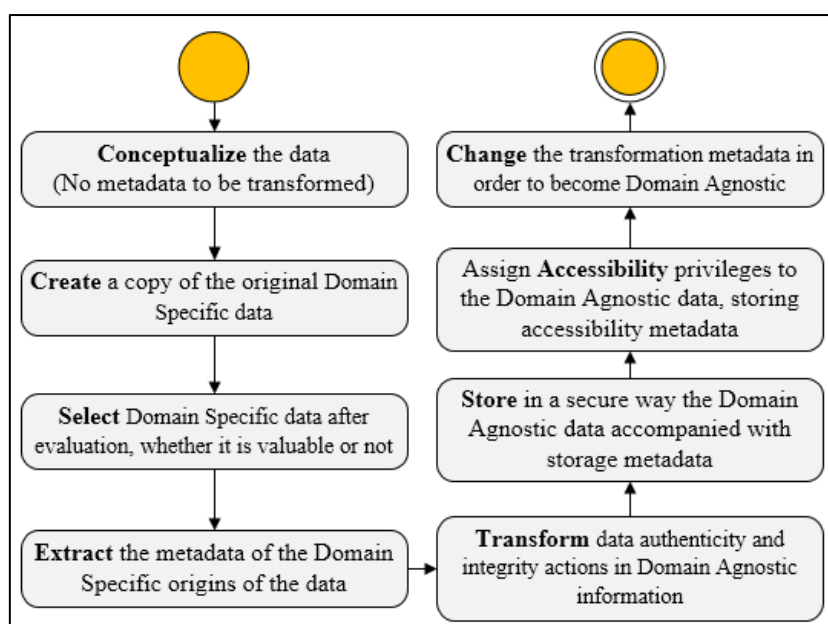
Αυτό σημαίνει ότι στην προτεινόμενη προσέγγιση θεωρούνταν διαθέσιμες μόνο οι πληροφορίες για το τι αντιπροσωπεύουν τα δεδομένα ενός συγκεκριμένου αισθητήρα, και ένας μηχανισμός που επέτρεπε να διαπιστωθεί πώς μπορούν να χρησιμοποιηθούν τα δεδομένα σε συνδυασμό με μη επισημασμένα δεδομένα από άγνωστη πηγή - με τον πιο διαλειτουργικό τρόπο - παρουσιάστηκε. Η προσέγγιση αντιμετώπιζε συγκεκριμένες προκλήσεις απαντώντας αρχικά στο ερώτημα εάν διαφορετικά είδη δεδομένων έχουν μια συγκεκριμένη συμπεριφορά σε μία ή περισσότερες φάσεις του κύκλου ζωής δεδομένων ή όχι. Επιπλέον, έδωσε μια απάντηση στο ερώτημα σχετικά με το πώς μπορούμε να αποκτήσουμε περισσότερες γνώσεις από άγνωστα δεδομένα εάν έχουμε ήδη πληροφορίες σχετικά με ένα συγκεκριμένο σύνολο δεδομένων, προσπαθώντας να αποδώσουμε τη μέγιστη δυνατή χρήση από αυτό.

Η προτεινόμενη λύση υλοποιήθηκε μέσω τριών (3) διαφορετικών βημάτων και περιγράφεται λεπτομερώς στην Εικόνα 4-1, Εικόνα 4-2, και Εικόνα 4-3. Στο 1^ο βήμα (Εικόνα 4-1), λαμβάνεται ένα συγκεκριμένο σύνολο δεδομένων από έναν γνωστό αισθητήρα ενός συγκεκριμένου τομέα, και αυτό που αντιπροσωπεύει έχει ήδη καθοριστεί, καθώς και η πηγή του. Για αυτό το σύνολο δεδομένων, αναλύουμε τον πλήρη κύκλο ζωής τους ως προς τον τρόπο με τον οποίο έχει διαμορφωθεί, δημιουργηθεί, επιλεγεί, διατηρηθεί, αποθηκευτεί, προσεγγιστεί και μετασχηματιστεί. Ειδικότερα, για κάθε φάση που παρέχει αξία για την επίτευξη σημασιολογικής διαλειτουργικότητας σε διάφορους αισθητήρες, αποθηκεύουμε τις συγκεκριμένες πληροφορίες των δεδομένων των τομέων με τη μορφή μεταδεδομένων, στα οποία θα αποκτήσουμε πρόσβαση κατά τη διάρκεια του 2ου βήματος της προτεινόμενης προσέγγισης. Εν ολίγοις, στο τέλος κάθε ξεχωριστής φάσης του κύκλου ζωής δεδομένων αυτού του συνόλου δεδομένων, αποθηκεύουμε όλες τις προαναφερθείσες πληροφορίες ως μεταδεδομένα, χρησιμοποιώντας την έννοια της γλώσσας συγκεκριμένου τομέα (Domain Specific Language - DSL) [218]. Ως αποτέλεσμα, επιτυγχάνεται η αποθήκευση των πληροφοριών που περιγράφουν την αξία, την προέλευση και τη φύση των δεδομένων του αισθητήρα στο συγκεκριμένο τομέα στον οποίο ανήκει, για τις διάφορες φάσεις του κύκλου ζωής δεδομένων.



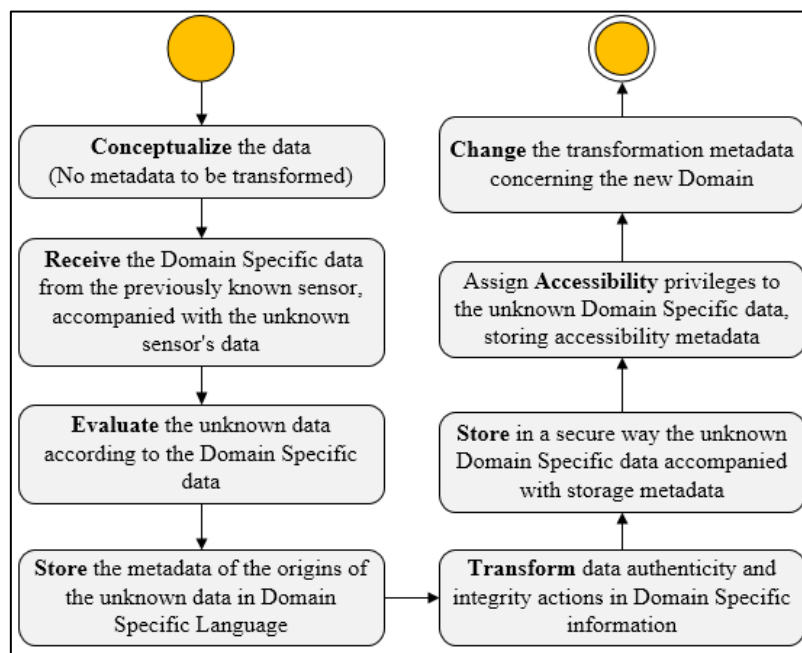
Εικόνα 4-1 - 1ο Βήμα μεταφοράς δεδομένων σε DSL

Στο 2ο βήμα (Εικόνα 4-2), με τις συγκεκριμένες πληροφορίες που έχουν αποκτηθεί από το 1ο βήμα, εξάγονται πληροφορίες από τα μεταδεδομένα που χαρακτηρίζουν αυτά τα δεδομένα ως συγκεκριμένου τομέα από κάθε φάση του μοντέλου του κύκλου ζωής δεδομένων. Έτσι, τα δεδομένα σχετίζονται με λιγότερο ακριβή και γενικευμένα μεταδεδομένα χρησιμοποιώντας την έννοια της γλώσσας αγνώστου τομέα (Domain Agnostic Language – DAL) [219]. Για το σκοπό αυτό, αναπτύσσεται μια πιο αφηρημένη άποψη σχετικά με τη φύση και την αξία αυτών των δεδομένων για κάθε φάση του κύκλου ζωής των δεδομένων, χωρίς πλέον να υπάρχει προηγούμενη γνώση του τομέα/προέλευσης αυτού του συνόλου δεδομένων.



Εικόνα 4-2 - 2ο Βήμα μεταφοράς δεδομένων σε DAL

Στο 3ο βήμα (Εικόνα 4-3), οι πληροφορίες που αναφέρθηκαν προηγουμένως περιγράφονται με την έννοια του DAL, και θα χρησιμοποιηθούν σε ένα τεχνητό νευρωνικό δίκτυο [220] παράλληλα με άλλα μη επισημασμένα δεδομένα που προέρχονται από άγνωστους αισθητήρες, προσπαθώντας να βρεθούν ομοιότητες, κοινά μέρη, συμπεριφορές, και πρότυπα, ώστε να γίνει κατανοητή η φύση του τελευταίου, μετατρέποντας τες, χρησιμοποιώντας την προσέγγιση DAL. Επιπλέον, κατά τη διάρκεια αυτού του βήματος, εστιάζουμε στην εξεύρεση συγκεκριμένων χαρακτηριστικών μεταξύ του συγκεκριμένου πεδίου και των άγνωστων δεδομένων (που περιγράφονται στο DAL), προκειμένου να ταιριάζει και να χρησιμοποιεί τα άγνωστα δεδομένα ως είσοδο για έναν διαφορετικό αισθητήρα. Αυτή η ιδέα θα χρησιμοποιηθεί ξεχωριστά για κάθε διαφορετική φάση του κύκλου ζωής των δεδομένων, με αποτέλεσμα πολύτιμες πληροφορίες να χρησιμοποιηθούν περαιτέρω για την κατανόηση της φύσης των πρώτων δεδομένων και να δοθούν ως είσοδος σε έναν αισθητήρα. Μετά την ολοκλήρωση αυτής της σύγκρισης, συγκεντρώνουμε πληροφορίες σχετικά με το άγνωστο σύνολο δεδομένων και η προσέγγιση του DSL θα χρησιμοποιηθεί και πάλι, με αποτέλεσμα την αναγνώριση αυτού του συνόλου δεδομένων.



Εικόνα 4-3 - 3ο Βήμα μεταφοράς και αναγνώρισης δεδομένων σε DSL

Ως αποτέλεσμα αυτής της προσέγγισης, τα δεδομένα αυτά θα χρησιμοποιηθούν με λειτουργικό τρόπο ως συνεισφορά σε μια διαφορετική περίπτωση εφαρμογής, επιτυγχάνοντας τη μέγιστη χρήση από παλαιότερα γνωστά δεδομένα.

4.1.2 Στάδιο Συσσωμάτωσης Ετερογενών δεδομένων υγείας μέσω μιας Οντολογικής Κοινής Γλώσσας Υγείας

Σε αυτό το στάδιο, στην παρούσα προσέγγιση προτάθηκε μία λύση για τη χρήση μιας Κοινής Γλώσσας Υγείας (Common Health Language - CHL), η οποία θα διευκόλυνε τη συγκέντρωση των κατανεμημένων ετερογενών δεδομένων που προέρχονταν από πολλαπλά σημεία για την καλύτερη ενημέρωση των αποφάσεων που σχετίζονται με την υγεία. Έτσι, οι στόχοι ήταν: (i) η υποστήριξη της ανταλλαγής δεδομένων για τη λήψη καλύτερων αποφάσεων και (ii) η εξαγωγή πολύτιμων πληροφοριών από ετερογενή δεδομένα που προέρχονται από διαφορετικά EHR. Η παρούσα προσέγγιση αρχικά μελέτησε πως λόγω της έκρηξης των διαθέσιμων υπηρεσιών ΤΠΕ, υπάρχουν αρκετοί αισθητήρες και εφαρμογές που υποστηρίζουν την εξατομικευμένη περίθαλψη, ενώ πολλές εφαρμογές [221] έχουν αναπτυχθεί για την ανίχνευση και την πρόβλεψη των ανωμαλιών της υγείας των ασθενών και τη διαχείριση της θεραπείας. Η Gartner [222] αναφέρει ότι η παγκόσμια αγορά της φορητής τεχνολογίας έχει φθάσει σε 275 εκατομμύρια συσκευές. Επί του παρόντος, όλες αυτές οι υπηρεσίες και τα δεδομένα είναι ετερογενή και λειτουργούν ανεξάρτητα, ενώ η αξία που προκύπτει από την εκμετάλλευσή τους είναι περιορισμένη. Έτσι, κάθε χρόνο υπάρχουν επιπλέον θάνατοι και τραυματισμοί εξαιτίας της κακής επικοινωνίας μεταξύ των επαγγελματιών του τομέα της υγείας [223]. Τα δεδομένα κατανέμονται σε διάφορα συστήματα, όπως για παράδειγμα δεδομένα σχετικά με τις συνθήκες υγείας που φυλάσσονται στο νοσοκομείο, δεδομένα για τις προηγούμενες νόσους, ή δεδομένα των ιατρικών αισθητήρων που αξιοποιούνται από συγκεκριμένα συστήματα. Λόγω αυτής της ανεπαρκούς ενσωμάτωσης της τεχνολογίας, καθώς και τη δημιουργία μεγάλων δεδομένων στο χώρο της υγείας, καθίσταται όλο και συχνότερη η αποτυχία σημαντικών γεγονότων, όπως οι έγκαιρες ενδείξεις ασθενειών. Σύμφωνα με την International Data Corporation (IDC) [224], τα προηγούμενα χρόνια έχει αναλυθεί λιγότερο από το 5% των πληροφοριών, επειδή ελάχιστα ήταν γνωστά για τα δεδομένα.

Επιπλέον, τα σημερινά αρχεία υγείας, καθώς και τα EHR απέχουν πολύ από το τι θεωρούν οι πολίτες ως πολύτιμα για την υγεία τους. Η συλλογή πληροφοριών, όπως το περιβάλλον ή ο τρόπος ζωής, καθώς και η σύνδεσή τους με άλλα δεδομένα θα ήταν επωφελής για τη μάθηση σχετικά με τα αποτελέσματα στρατηγικών πρόληψης και ασθενειών. Στο [225] τονίζεται ότι η ανταλλαγή πληροφοριών μεταβάλλει τη συνολική προσέγγιση των ασθενών προς την υγιή ζωή, ενώ το [226] δείχνει την αξία της ανταλλαγής πληροφοριών σχετικών με την υγεία. Η μελέτη στο [227] εκτιμά ότι θα μπορούσαν να επιτευχθούν ετήσιες εξοικονομήσεις περίπου 78 δισ. δολαρίων αν χρησιμοποιούνταν πρότυπα

ανταλλαγής δεδομένων σε ολόκληρο τον τομέα της υγειονομικής περίθαλψης, αποκαλύπτοντας ότι το σημαντικό πρόβλημα δεν είναι τα συστήματα αυτά καθαυτά αλλά η ανταλλαγή δεδομένων μεταξύ των συστημάτων.

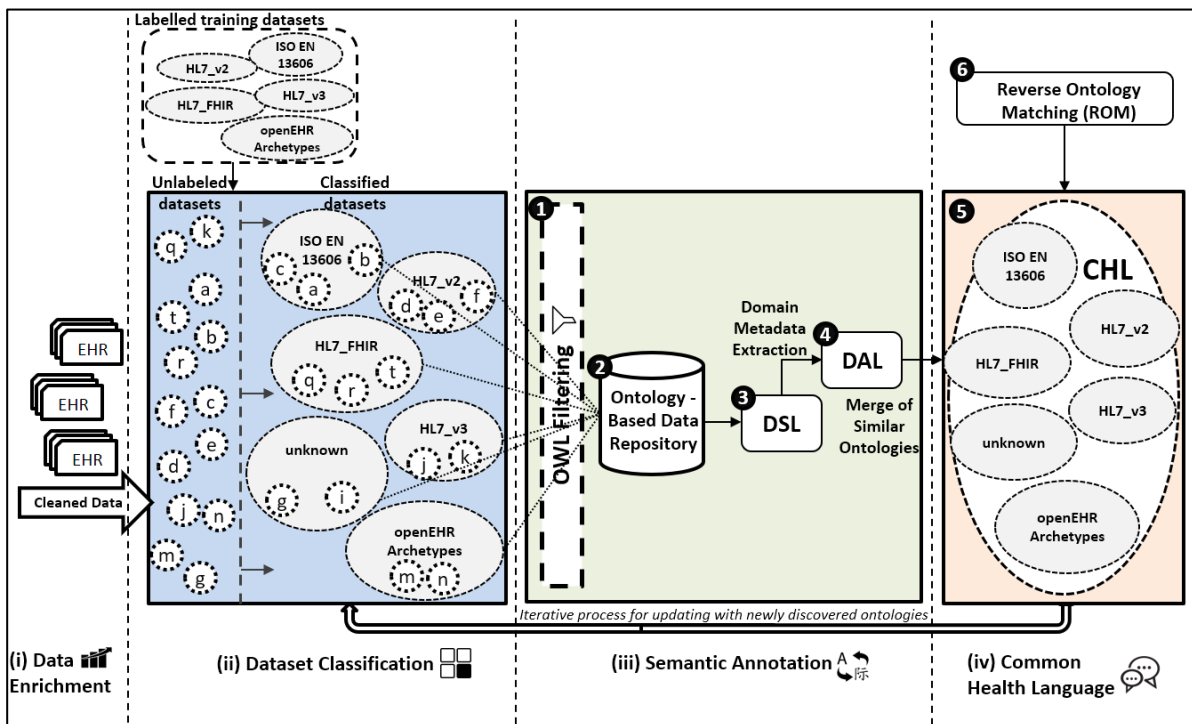
Καθίσταται σαφές ότι η βάση είναι τα ιατρικά αρχεία, τα οποία δεν πρέπει να απομονώνονται και να συνδέονται με τους πολίτες, αλλά να είναι «κοινά» με τα εξουσιοδοτημένα άτομα. Αυτός είναι ο λόγος για τον οποίο, το 2014, ο Νόμος περί Υγείας Τεχνολογίας Πληροφοριών για την Οικονομική και Κλινική Υγεία (Health Information Technology for Economic and Clinical Health Act - HITECH) [228] και ο Νόμος για την Προστασία και Προσιτή Φροντίδα των Ασθενών (Patient Protection and Affordable Care Act - PPACA) [229] έθεσαν την ανταλλαγή δεδομένων και πληροφοριών μεταξύ πολλών EHRs [230] μείζονος σημασίας. Επιπλέον, πρόσφατες συστάσεις της Ευρωπαϊκής Κοινότητας [231] ανέφεραν ότι η σημασιολογική διαλειτουργικότητα μεταξύ των πηγών δεδομένων για την υγεία είναι απαραίτητη για τη βελτίωση της ποιότητας και της ασφάλειας της περίθαλψης των ασθενών, της ιατρικής έρευνας και της διαχείρισης της υγειονομικής υπηρεσίας, ενώ παράλληλα προωθείται η βιωσιμότητα των συστημάτων πληροφοριών για την υγειονομική περίθαλψη [232].

Συνοπτικά, κάθε προτεινόμενη λύση στόχευε στην παροχή πρόσβασης στις κλινικές πληροφορίες των ασθενών για συγκεκριμένους κλινικούς οργανισμούς. Ωστόσο, ένα από τα σημαντικότερα μειονεκτήματα είναι ότι αυτές οι προσεγγίσεις δεν πρότειναν πλαίσια εύκολα εφαρμόσιμα σε διαφορετικά πρότυπα, ενώ έλειπε η έννοια ενός παγκόσμιου πλαισίου διαλειτουργικότητας. Συνεπώς, ήταν σαφές ότι υπήρχε μεγάλη ανάγκη για μια ολιστική προσέγγιση που θα είναι σε θέση να συγκεντρώνει πολλαπλές ομάδες ετερογενών δεδομένων EHR, να αποκτά γνώσεις και να προσφέρει μεγαλύτερη αξία στο οικοσύστημα υγείας των ασθενών.

Οι οντολογίες χρησιμοποιούνται στην υγειονομική περίθαλψη για την επίλυση όρων για ερωτήματα και άλλες γενικές στρατηγικές αναζήτησης, βελτιώνοντας έτσι τα αποτελέσματα αναζήτησης που παρουσιάζονται στους επαγγελματίες του τομέα υγείας. Επομένως, με την προσέγγισή μας, χρησιμοποιήθηκε ένας μηχανισμός φιλτραρίσματος για τον ορισμό των EHRs ως οντολογίες για να παράσχει ένα σημασιολογικό μοντέλο για να αντιπροσωπεύει κανόνες ορισμού πολλαπλών ιατρικών προτύπων. Αυτές οι οντολογίες κατασκευάστηκαν χρησιμοποιώντας το OWL. Επιπλέον, εφαρμόστηκαν ειδικές τεχνικές και σημασιολογικά μοντέλα για τη διαλειτουργική χρήση των δεδομένων, συνδυάζοντας τις τεχνικές διαλειτουργικότητας με μια οντολογική προσέγγιση.

Μεταξύ των καινοτομιών της προτεινόμενης προσέγγισης ήταν ένας μηχανισμός ταξινόμησης για την κατηγοριοποίηση πολλαπλών συνόλων δεδομένων ενός «αναγνωρισμένου» ιατρικού προτύπου σε κατηγορίες της ίδιας μορφής. Επιπλέον, υλοποιήθηκε ένας μηχανισμός εξαγωγής μεταδεδομένων για την εξαγωγή ειδικών πληροφοριών τομέα από ταξινομημένα σύνολα δεδομένων EHR, ώστε να τα μετασχηματίσει στο CHL, καθώς και ένας μηχανισμός αντιστοίχισης οντολογιών για την αντιστοίχιση και τη μετάφραση σε CHL των άγνωστων συνόλων δεδομένων.

Η προσέγγιση αυτή πρότείνει μια ολοκληρωμένη γενική σημασιολογική αρχιτεκτονική για να αντιμετωπίσει τα ετερογενή δεδομένα που προέρχονται από διαφορετικές πηγές, αποκτώντας την οντολογική αναπαράσταση των δεδομένων και μεταμορφώνοντάς τα με ολιστικό τρόπο (Εικόνα 4-4).



Εικόνα 4-4 - Αρχιτεκτονική προσέγγισης Κοινής Γλώσσας Υγείας

Όπως φαίνεται στην Εικόνα 4-4, η προτεινόμενη αρχιτεκτονική αποτελείται από τέσσερα κύρια στοιχεία: (i) το στοιχείο εμπλουτισμού δεδομένων (Data Enrichment) που χρησιμοποιείται για την εκτέλεση ενεργειών καθαρισμού δεδομένων, (ii) το στοιχείο ταξινόμησης (Dataset Classification), που χρησιμοποιείται για την ταξινόμηση των EHR σε ομάδες με τα πιο συνήθη ιατρικά πρότυπα, (iii) το στοιχείο σημασιολογικού σχολιασμού (Semantic Annotation) όπου τα σύνολα δεδομένων περιγράφονται σε μια γενικευμένη μορφή και (iv) το στοιχείο του CHL, που χρησιμοποιείται για τη μετάφραση των δεδομένων σε μία κοινή γλώσσα. Όσον αφορά το στοιχείο εμπλουτισμού

δεδομένων, εφαρμόζεται το OpenRefine [233], διασφαλίζοντας την αξιοπιστία των δεδομένων, αφαιρώντας σφάλματα, παρέχοντας ακρίβεια και διασφαλίζοντας πληρότητα. Στην περίπτωση μας, το OpenRefine λειτουργεί σε σειρές δεδομένων EHR (JSON, XML, κ.λπ.), όπου οι σειρές μπορούν να φιλτραριστούν για να εμφανίσουν τις όψεις που καθορίζουν τα κριτήρια φιλτραρίσματος, αφαιρώντας έτσι τιμές που ενδεχομένως περιέχουν σφάλματα. Στο τέλος, τα παλιά σύνολα δεδομένων αντικαθίστανται από τα νεοσυσταθέντα και καθαρισμένα σύνολα δεδομένων, τα οποία θα χρησιμοποιηθούν περαιτέρω στην ταξινόμηση των δεδομένων της αναπτυχθείσας προσέγγισης.

Λαμβάνοντας υπόψη ότι τα σύνολα δεδομένων EHR είναι διαφορετικής μορφής, λαμβάνει χώρα η ταξινόμηση των δεδομένων, όπου αυτά τα σύνολα δεδομένων εναρμονίζονται και κατηγοριοποιούνται σε ομάδες ίδιας μορφής (δηλ. ιατρικά πρότυπα). Έτσι, γίνεται ευκολότερο να αναπτυχθούν μηχανισμοί για τη μετατροπή αυτών των συνόλων δεδομένων σε μια ολιστική γλώσσα (δηλ. CHL). Επομένως, η ταξινόμηση του συνόλου δεδομένων πραγματοποιείται μέσω μηχανισμών διανυσματικής γραμμικής υποστήριξης πολλαπλών τάξεων (Support Vector Machines - SVM) [234], μαζί με το μοντέλο CBOW. Τέτοια μέθοδος εφαρμόστηκε, καθώς είναι μια γενική τεχνική που μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα προβλημάτων, παρέχοντας ακριβή αποτελέσματα. Πρέπει να σημειωθεί ότι για να επιτευχθεί μια τέτοια ταξινόμηση, ο αναπτυγμένος μηχανισμός, εκτός από τη συγκέντρωση των πολλαπλών μη ταξινομημένων συνόλων δεδομένων EHR διαφορετικών ιατρικών προτύπων, λαμβάνει επίσης ως είσοδο ένα σύνολο επισημασμένων συνόλων δεδομένων, όπου κάθε ένα περιγράφεται με τη μορφή διαφορετικού ιατρικού προτύπου (π.χ. HL7 v2, HL7 v3, κ.λπ.). Τα δεδομένα που δεν ανήκουν σε κανένα από τα γνωστά ιατρικά πρότυπα ταξινομούνται στην άγνωστη κατηγορία που θα εξηγηθεί στο τελικό βήμα του σημασιολογικού σχολιασμού.

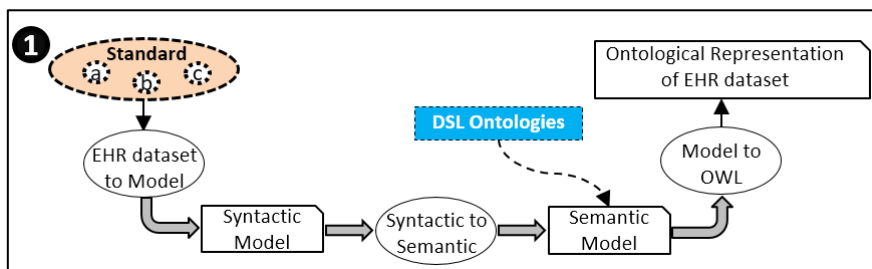
Το επόμενο βήμα είναι να εναρμονιστούν τα δεδομένα που συλλέγονται προκειμένου να μεταποιηθούν και να μετατραπούν σε ένα γενικό σχήμα, επιτρέποντας την ευκολότερη και πιο διαλειτουργική ανταλλαγή δεδομένων, κατά τη φάση σημασιολογικού σχολιασμού, που αποτελείται από τέσσερα (4) κύρια βήματα:

- **Βήμα 1:** Ο ανεπτυγμένος μηχανισμός παρέχει ένα επίσημο σημασιολογικό μοντέλο για την αντιπροσώπευση των ταξινομημένων συνόλων δεδομένων EHR των ιατρικών προτύπων, χρησιμοποιώντας το Protégé Framework [235] για τον καθορισμό οντολογιών OWL. Μέσω του μηχανισμού φιλτραρίσματος OWL (Εικόνα 4-5), όλα τα ταξινομημένα σύνολα

δεδομένων EHR εκφράζονται σε οντολογικούς κανόνες, συμπεριλαμβανομένων των λεπτομερών και ειδικών για τον τομέα ορισμών των κλινικών εννοιών. Ειδικότερα, η διαδικασία μετασχηματισμού αποτελείται από τα ακόλουθα βήματα:

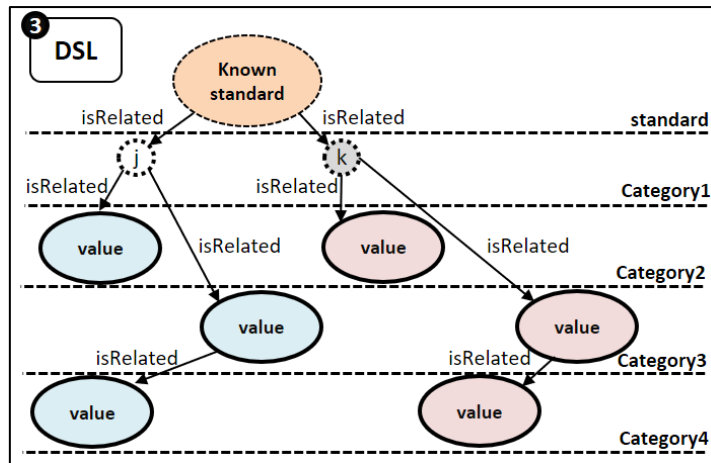
1. Κάθε σύνολο δεδομένων EHR κάθε διαφορετικού προτύπου εκφράζεται ως συντακτικό μοντέλο, μέσω ενός αναλυτή EHR (EHR dataset to Model).
2. Το συντακτικό μοντέλο μετασχηματίζεται σε ένα σημασιολογικό μοντέλο χρησιμοποιώντας ένα σύνολο κανόνων που έχουν προκαθοριστεί για την αντιστοίχιση μοντέλου από συντακτικό σε σημασιολογικό (Syntactic to Semantic).
3. Το σημασιολογικό μοντέλο μαζί με τις προκαθορισμένες οντολογίες συγκεκριμένου πεδίου (DSL ontologies), μετασχηματίζεται στην οντολογική αναπαράσταση του προτύπου του, μέσω ενός μετασχηματισμού από μοντέλο σε OWL οντολογίες (Model to OWL).

Αυτή η διαδικασία επαναλαμβάνεται για όλα τα δεδομένα των EHR των διαφόρων προτύπων, με αποτέλεσμα την οντολογική εκπροσώπηση κάθε EHR.



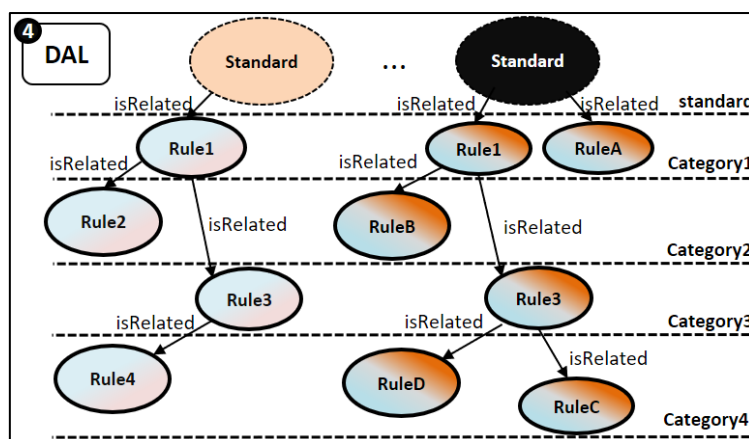
Εικόνα 4-5 - Μηχανισμός φιλτραρίσματος OWL οντολογιών

- **Βήμα 2:** Όλες οι κατασκευαζόμενες οντολογικές αναπαραστάσεις αποθηκεύονται στη συνέχεια σε ένα χώρο αποθήκευσης δεδομένων που βασίζεται στην οντολογία (GraphDB [236]), ακολουθώντας την προσέγγιση που προτείνεται στο [237] για την αποθήκευση ιατρικών δεδομένων, που περιέχουν πληροφορίες μεταδεδομένων και περιεχομένου για τα διάφορα EHR.
- **Βήμα 3:** Κάθε ταξινομημένο σύνολο δεδομένων της συνιστώσας ταξινόμησης μετατρέπεται σε οντολογίες OWL, χρησιμοποιώντας την έννοια του DSL, προσδιορίζοντας σχέσεις, τιμές και έννοιες, μεταξύ των πολλαπλών τιμών των EHR που ανήκουν στην ίδια ομάδα ιατρικών προτύπων (Εικόνα 4-6).



Εικόνα 4-6 - Εύρεση κοινών σχέσεων, τιμών και εννοιών

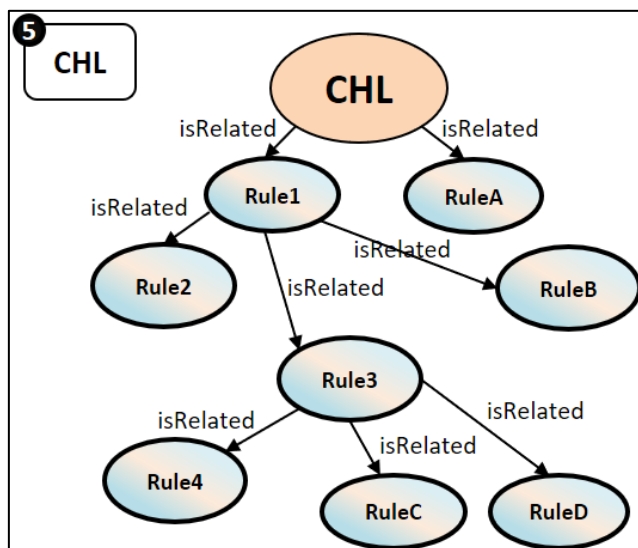
- Βήμα 4:** Το επόμενο βήμα είναι να ταξινομηθεί κάθε τιμή σε μια γενικευμένη κατηγορία, σύμφωνα με τη σημασιολογία της και τις σχέσεις μεταξύ των διαφορετικών αξιών, έτσι ώστε κάθε τιμή να ανήκει σε μια μοναδική κατηγορία. Στην περίπτωση αυτή, χρησιμοποιώντας την έννοια του DAL και την έρευνα που αποκτήθηκε από την προηγούμενη δουλειά μας [238], οι πληροφορίες μεταδεδομένων εξάγονται με αποτέλεσμα τον γενικό τρόπο αντιπροσώπευσης των δεδομένων EHR. Το τελευταίο βήμα είναι να δημιουργηθεί ένας κανόνας για κάθε διαφορετική τιμή που ανήκει στην ίδια κατηγορία, ώστε να προκύψει ένας συγκεκριμένος τρόπος με τον οποίο πρέπει να εκπροσωπείται κάθε διαφορετική τιμή (Εικόνα 4-7). Αυτή η διαδικασία επαναλαμβάνεται για όλα τα σύνολα δεδομένων των ταξινομημένων ιατρικών προτύπων, με αποτέλεσμα τους διαφορετικούς κανόνες αναπαράστασης αξίας της ίδιας κατηγορίας.



Εικόνα 4-7 - Δημιουργία κανόνων

Προκειμένου αυτή η γενική γνώση να μετατραπεί σε κοινή μορφή, εμφανίζεται το στοιχείο του CHL. Ο ανεπτυγμένος μηχανισμός εκτελεί τις τεχνικές αντιστοίχισης οντολογιών που παρέχονται στο [239] για κάθε ιατρικό πρότυπο

και κανόνα, προκειμένου να εντοπίσει ανόμοιες τιμές της ίδιας κατηγορίας. Οι οντολογίες της ίδιας κατηγορίας με τον ίδιο τύπο σχέσεων συγχωνεύονται με (α) την εξαγωγή του ίδιου τύπου κανόνων και (β) την ενημέρωση με κανόνες που υπάρχουν σε μια κατηγορία αλλά δεν υπάρχουν σε άλλη. Έτσι, ο ανεπτυγμένος μηχανισμός αναγνωρίζει πρότυπα κανόνων σχετικά με διαφορετικούς τύπους κανόνων και τα αποθηκεύει ως μεταδεδομένα. Η συλλεγμένη σημασιολογική γνώση αποθηκεύεται μέσω του GraphDB, ενώ οι κανόνες μετασχηματίζονται στην μορφή του CHL (Εικόνα 4-8).



Εικόνα 4-8 - Μετασχηματισμός κανόνων σε οντολογική μορφή

Όσον αφορά τα άγνωστα σύνολα δεδομένων EHR, έχει σχεδιαστεί μια προσέγγιση αντιστροφής οντολογικής αντιστοίχισης (Reverse Ontology Mapping - ROM) για τη μετάφρασή τους σε CHL. Το επόμενο βήμα είναι να συγκρίνουμε τα άγνωστα σύνολα δεδομένων με τα οντολογικά δεδομένα EHR, όπου με την εκτέλεση μίας ROM και μιας σύγκρισης με τα μεταδεδομένα των κανόνων που εκφράζονται ως οντολογίες, είναι δυνατόν να εκφραστεί η φύση των αγνώστων δεδομένων. Ο μηχανισμός εντοπίζει αρχικά ομοιότητες μεταξύ ενός αγνώστου στιγμιότυπου δεδομένων με έναν συγκεκριμένο κανόνα CHL και ακολουθώντας το ανάποδο οντολογικό μονοπάτι αυτού του κανόνα καταλήγει σε ένα γενικό κανόνα, κατηγοριοποιώντας το στην «οικογένεια» του γονικού κανόνα. Έτσι, το προτεινόμενο CHL μπορεί να ενσωματωθεί σε διαφορετικά συστήματα, έτσι ώστε τα ετερογενή σύνολα δεδομένων του EHR να μετασχηματιστούν σε κατανοητές και κοινά γνωστές γνώσεις που μπορούν να χρησιμοποιηθούν για την καλύτερη επικοινωνία μεταξύ των επαγγελματιών του τομέα της υγείας.

4.1.3 Στάδιο απόκτησης της Οντολογικής Αναπαράστασης των Δεδομένων της Υγείας μέσω των Τεχνικών Μεταμοντέλων

Σε αυτό το πλαίσιο, στην παρούσα προσέγγιση προτάθηκε μία λύση για τη μετατροπή των δεδομένων που συλλέγονται στον τομέα της υγειονομικής περίθαλψης σε ένα γενικό οντολογικό σχήμα, επιτρέποντας την ευκολότερη και πιο διαλειτουργική ανταλλαγή δεδομένων. Εν συντομία, ένα συντακτικό και σημασιολογικό μοντέλο υλοποιήθηκε για διαφορετικά σύνολα δεδομένων για τα ιατρικά πρότυπα, όπου αυτά τα σύνολα δεδομένων εκφράζονταν σε οντολογικούς κανόνες, αφού είχαν πρώτα μετασχηματιστεί σε συντακτικά μοντέλα. Η προτεινόμενη προσέγγιση διευκόλυνε τη συσσωμάτωση των ετερογενών δεδομένων που προέρχονταν από πολλαπλά σημεία ώστε να λαμβάνουν καλύτερα ενημερωμένες αποφάσεις σχετικά με την υγεία και να υποστηρίζουν την ανταλλαγή δεδομένων και την ολοκλήρωση για καλύτερη λήψη αποφάσεων. Συνοπτικά, η παρούσα προσέγγιση αρχικά μελέτησε πως στα τελευταία χρόνια, η έκρηξη των διαθέσιμων υπηρεσιών ΤΠΕ οδήγησε στη δημιουργία μιας σειράς αισθητήρων και εφαρμογών που υποστηρίζουν την εξατομικευμένη φροντίδα. Αυτήν τη στιγμή, η ζωή μας συνδέεται όλο και περισσότερο με τις συσκευές μας, με άλλους ανθρώπους και με διάφορα αντικείμενα (IoT). Οι έξυπνες μηχανές γίνονται πιο έξυπνες και μια νέα πραγματικότητα στον τομέα των τεχνολογιών πληροφορικής πρέπει να εξελιχθεί με τεχνολογικές αρχιτεκτονικές και πλατφόρμες για να υποστηρίξει την πρόοδο ενός ψηφιακά συνδεδεμένου κόσμου. Η Gartner [240] προβλέπει ότι 8.4 δισεκατομμύρια συνδεδεμένες συσκευές θα χρησιμοποιηθούν παγκοσμίως το 2017, αυξημένες κατά 31% από το 2016 και θα φθάσουν τα 20.4 δισεκατομμύρια μέχρι το 2020, ενώ οι συνολικές δαπάνες για τελικά σημεία και υπηρεσίες φτάνουν στα 2 τρισ. δολάρια το 2019. Οι συσκευές έχουν σχεδιαστεί για να παρακολουθούν τις καθημερινές δραστηριότητες των ασθενών, να ανιχνεύουν τα σχήματα ύπνου και τα συναισθήματα [241], και να μετρούνται τα ζωτικά σημεία [242]. Πολλές εφαρμογές και συστήματα [243] έχουν αναπτυχθεί βάσει δεδομένων που μπορούν να φορευθούν για την ανίχνευση και πρόβλεψη των ανωμαλιών της υγείας των ασθενών, τη μελέτη της ανθρώπινης συμπεριφοράς και τη διαχείριση της θεραπείας.

Επί του παρόντος, τα παρεχόμενα δεδομένα είναι ετερογενή και λειτουργούν ανεξάρτητα, ενώ η αξία που προκύπτει από την εκμετάλλευσή τους είναι περιορισμένη. Οι περισσότεροι από τους παραδοσιακούς αλγόριθμους εξόρυξης δεδομένων δεν είναι σε θέση να αντιμετωπίσουν τόσο την κλίμακα όσο και την ανομοιογένεια των δεδομένων που μπορούν να παρουσιαστούν στην υγειονομική περίθαλψη [244], καθιστώντας όλο και πιο κοινό το να

χάνονται προκαταρκτικές ενδείξεις ασθενειών. Σύμφωνα με το [245] τα ιατρικά λάθη όπως οι καθυστερήσεις στη διάγνωση, οι προληπτικές χειρουργικές επιπλοκές και οι υπερβολικές δόσεις φαρμάκων αποτελούν κύρια αιτία θανάτου και τραυματισμού στις Ηνωμένες Πολιτείες. Εκτιμάται ότι το 80% των σοβαρότερων ιατρικών σφαλμάτων μπορεί να συνδεθεί με την κακή επικοινωνία μεταξύ των κλινικών ιατρών.

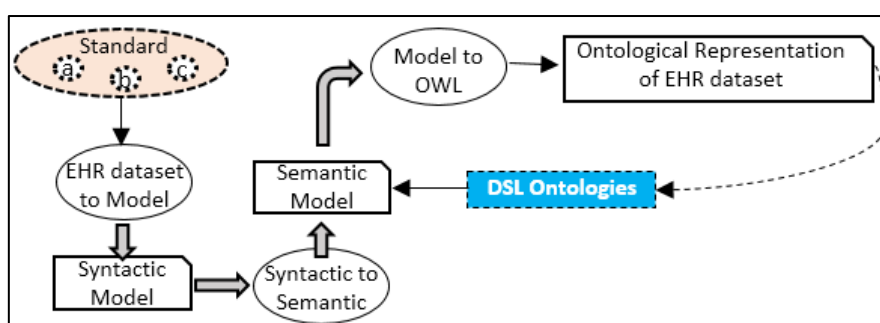
Επιπλέον, τα σημερινά EHR δεν προσφέρουν την επιθυμητή αξία για την υγεία των πολιτών [246], παρά το γεγονός ότι προσφέρουν τη δυνατότητα μετατροπής του συστήματος υγειονομικής περίθαλψης από μια βιομηχανία που βασίζεται κυρίως στον χαρτί, σε μια βιομηχανία που χρησιμοποιεί κλινικές και άλλες πληροφορίες για να βοηθήσουν τους παρόχους στην παροχή υψηλότερης ποιότητας περίθαλψης στους ασθενείς τους. Συνοπτικά, τα ηλεκτρονικά ιατρικά αρχεία φαίνεται να είναι η τρέχουσα τάση στην υγειονομική περίθαλψη, με ευρήματα πολλών ιατρών, συμμαχικών επαγγελματιών υγείας, φαρμακοποιών και νοσοκομείων που χρησιμοποιούν κάποια μορφή ηλεκτρονικής καταγραφής δεδομένων ασθενών. Η συλλογή πληροφοριών, όπως τα δημογραφικά στοιχεία των ασθενών, τα σημειώματα προόδου, τα προβλήματα, τα φάρμακα, τα ζωτικά σημάδια, το ιατρικό ιστορικό, οι εμβολιασμοί, τα εργαστηριακά δεδομένα και οι αναφορές ακτινολογίας, το να συνδέονται με άλλα δεδομένα στο EHR, θα ήταν επωφελές για την εκμάθηση των αποτελεσμάτων των στρατηγικών πρόληψης ασθενειών.

Ένα από τα μεγαλύτερα εμπόδια στην επικοινωνία συμβαίνει όταν υπάρχουν πολλοί τρόποι περιγραφής μιας ενιαίας έννοιας. Έτσι, είναι σαφές ότι το πρόβλημα είναι η δυσκολία ανταλλαγής δεδομένων μεταξύ συστημάτων και η ασυμβατότητα των δεδομένων. Η διαλειτουργικότητα είναι ο μόνος βιώσιμος τρόπος να βοηθηθούν οι εταιρείες να συνεργαστούν αρμονικά και να παράσχουν ποιοτική υγειονομική περίθαλψη [247]. Μια πρόσφατη μελέτη [248] υπολόγισε ότι θα μπορούσαν να επιτευχθούν ετήσιες εξοικονομήσεις περίπου 78 δισ. δολαρίων εάν χρησιμοποιούνταν πρότυπα ανταλλαγής δεδομένων σε ολόκληρο τον τομέα της υγειονομικής περίθαλψης. Αν και η διαλειτουργικότητα είναι τεχνολογικό ζήτημα, δεν πρόκειται μόνο για αυτό. Όπως συμβαίνει σε άλλες βιομηχανίες, η διαλειτουργικότητα απαιτεί από όλα τα μέρη να υιοθετήσουν ορισμένες αρχές διακυβέρνησης και εμπιστοσύνης και να δημιουργήσουν επιχειρηματικές συμφωνίες και λεπτομερείς οδηγούς για την εφαρμογή προτύπων. Τα μοναδικά ζητήματα εμπιστευτικότητας που αφορούν τα δεδομένα για την υγεία απαιτούν επίσης τη συμμετοχή νομοθετών και ρυθμιστικών αρχών. Η αντιμετώπιση αυτών των ζητημάτων απαιτεί συντονισμένη δράση από πολλούς μετόχους και αυτή η δράση θα συμβεί μόνο

αν προωθηθούν ισχυρά κίνητρα. Γίνεται σαφές ότι απαιτείται ένα τυποποιημένο λεξιλόγιο για την αναγνώριση της πολυπλοκότητας των διαδικασιών που περιβάλλουν την υγειονομική περίθαλψη όταν στοχεύουν στη διαλειτουργικότητα των πληροφοριών για την υγειονομική περίθαλψη.

Η διαλειτουργικότητα δεν έχει ακόμη καταστεί πραγματικότητα, παρά τη συντριπτική αυτή υποστήριξη για την ελεύθερη ροή δεδομένων ασθενών μεταξύ των φροντιστών. Πολλές προκλήσεις - τεχνικές, οικονομικές και διαδικαστικές - παραμένουν καθώς η υγειονομική περίθαλψη κινείται προς τη διαλειτουργικότητα. Ένα από τα μεγαλύτερα εμπόδια είναι να φτάσει η τεχνολογία στο σημείο όπου θα επιτρέψει στα διάφορα συστήματα να επικοινωνούν μεταξύ τους.

Συνοπτικά, κάθε συγκεκριμένη λύση βοήθησε στην παροχή πρόσβασης στις κλινικές πληροφορίες ασθενούς για συγκεκριμένους κλινικούς οργανισμούς. Ένα από τα μειονεκτήματα αυτών των προσεγγίσεων είναι ότι δεν πρότειναν πλαίσια και μεθόδους εύκολο να εφαρμοστούν σε διαφορετικά ιατρικά πρότυπα. Συνεπώς, σε αυτή την προσέγγιση εφαρμόστηκαν τεχνικές μετασχηματισμού για τη διαλειτουργική χρήση δεδομένων σε διαφορετικές υπηρεσίες, τοποθεσίες και περιβάλλοντα, σε συνδυασμό με μια οντολογική προσέγγιση που είχε ως στόχο την επίτευξη σημασιολογικής διαλειτουργικότητας. Εν ολίγοις, με την προσέγγιση αυτή, τα μεταμοντέλα και οι οντολογίες χρησιμοποιήθηκαν για να παράσχουν ένα επίσημο συντακτικό και σημασιολογικό μοντέλο, που αντιπροσώπευε σύνολα δεδομένων πολλαπλών ιατρικών προτύπων. Έτσι, κατασκευάστηκε ένας μηχανισμός μετάφρασης που μετασχημάτιζε ετερογενή ιατρικά δεδομένα διαφορετικής φύσης και μορφής σε μια κοινή οντολογική αναπαράσταση, χρησιμοποιώντας τη γλώσσα OWL. Η προσέγγισή προτείνει ένα περιεκτικό γενικό συντακτικό και σημασιολογικό μηχανισμό για να αντιμετωπίσει τα ετερογενή δεδομένα που προέρχονται από διαφορετικά EHR, αποκτώντας την οντολογική αναπαράσταση των δεδομένων του EHR.



Εικόνα 4-9 - Μεταφορά δεδομένων σε OWL οντολογίες

Όπως απεικονίζεται στην Εικόνα 4-9, η προτεινόμενη αρχιτεκτονική στοχεύει στην εναρμόνιση των δεδομένων ιατροφαρμακευτικής περίθαλψης πολλαπλών ιατρικών προτύπων προκειμένου να μεταποιηθεί και να μετατραπεί σε ένα γενικό οντολογικό σχήμα, επιτρέποντας την ευκολότερη και πιο διαλειτουργική ανταλλαγή δεδομένων. Με περισσότερες λεπτομέρειες, ο ανεπτυγμένος μηχανισμός παρέχει ένα επίσημο σημασιολογικό μοντέλο για την αντιπροσώπευση των συνόλων δεδομένων EHR των ιατρικών προτύπων, χρησιμοποιώντας το Protégé Framework για τον καθορισμό οντολογιών OWL. Με ένα τέτοιο μηχανισμό, τα διαφορετικά σύνολα δεδομένων του EHR εκφράζονται σε οντολογικούς κανόνες, συμπεριλαμβανομένων των λεπτομερών και ειδικών για τον τομέα ορισμών των κλινικών εννοιών.

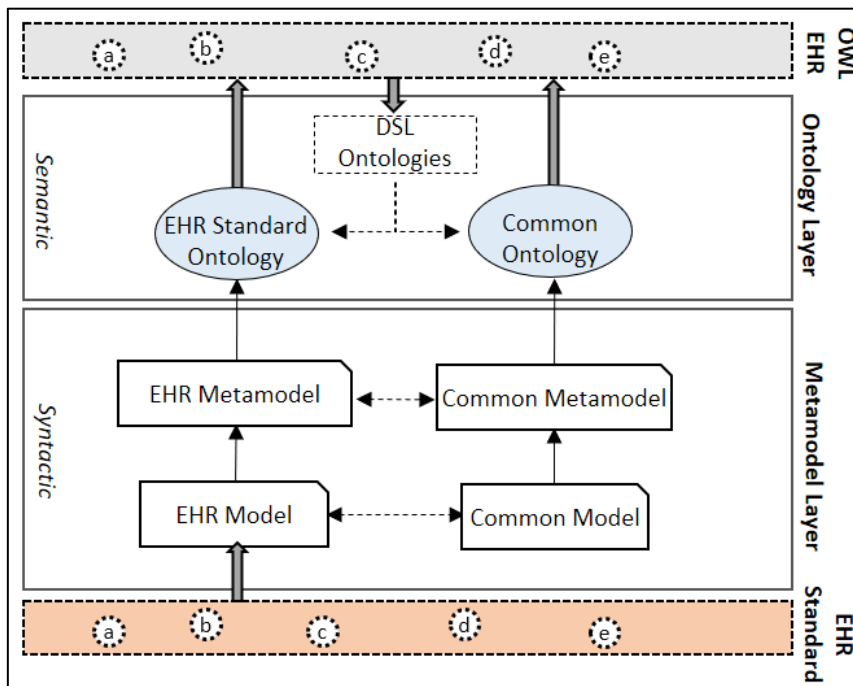
Ειδικότερα, η διαδικασία μετασχηματισμού αποτελείται από τα ακόλουθα βήματα:

1. Κάθε σύνολο δεδομένων EHR που ακολουθεί διαφορετικό πρότυπο εκφράζεται ως συντακτικό μοντέλο μέσω ενός αναλυτή EHR (EHR dataset to Model).
2. Το συντακτικό μοντέλο μετασχηματίζεται σε ένα σημασιολογικό μοντέλο χρησιμοποιώντας ένα σύνολο κανόνων που έχουν προκαθοριστεί για την αντιστοίχιση μοντέλου από συντακτικό σε σημασιολογικό (Syntactic to Semantic).
3. Το σημασιολογικό μοντέλο μαζί με τις προκαθορισμένες οντολογίες συγκεκριμένου τομέα (DSL Ontologies), μετασχηματίζεται στην οντολογική αναπαράσταση του προτύπου του, μέσω του μετασχηματισμού οντολογιών από μοντέλο σε OWL (Model to Owl).

Αυτή η διαδικασία επαναλαμβάνεται για όλα τα σύνολα δεδομένων των EHR των διαφόρων ιατρικών προτύπων, με αποτέλεσμα την οντολογική εκπροσώπηση κάθε EHR.

Η λύση αυτή συνδυάζει μια σειρά τεχνολογιών, δηλαδή οντολογίες και μεταμοντέλα. Η συνεισφορά στη διαδικασία μετασχηματισμού μας είναι ένα σύνολο δεδομένων EHR και το αποτέλεσμα είναι η οντολογική αναπαράσταση του συνόλου δεδομένων του EHR, αντίστοιχα. Η αρχιτεκτονική αυτής της λύσης απεικονίζεται στην Εικόνα 4-10. Γενικά, μπορούν να διακριθούν δύο διαφορετικά επίπεδα (layers), το επίπεδο μεταμοντέλου (Metamodel Layer) και το επίπεδο οντολογίας (Ontology Layer).

- Το Metamodel Layer περιέχει τα μοντέλα και τα μεταμοντέλα που αντιστοιχούν στα σύνολα δεδομένων του EHR. Σε αυτό το επίπεδο, πραγματοποιούνται οι αντιστοιχίσεις μετασχηματισμού.
- Το Ontology Layer περιλαμβάνει μια σειρά οντολογιών που διαμορφώνουν τη γνώση σχετικά με τα EHR για τα διαφορετικά ιατρικά πρότυπα. Σε αυτό το επίπεδο, γίνεται ο μετασχηματισμός των συνόλων δεδομένων EHR σε οντολογίες.



Εικόνα 4-10 - Αρχιτεκτονική επιπέδων μεταμοντέλου και οντολογιών

Με περισσότερες λεπτομέρειες, όπως απεικονίζεται στην Εικόνα 4-10, η αρχιτεκτονική της προτεινόμενης προσέγγισης αποτελείται από τα ακόλουθα:

Πρότυπο EHR

Το συστατικό πρότυπο EHR (EHR Standard) θεωρείται ότι περιέχει πολλαπλά σύνολα δεδομένων διαφορετικών ιατρικών προτύπων, τα οποία θα μετασχηματιστούν σε οντολογίες OWL μέσω του μετασχηματισμού τους από το Metamodel Layer και το Ontology Layer αντίστοιχα.

Metamodel Layer

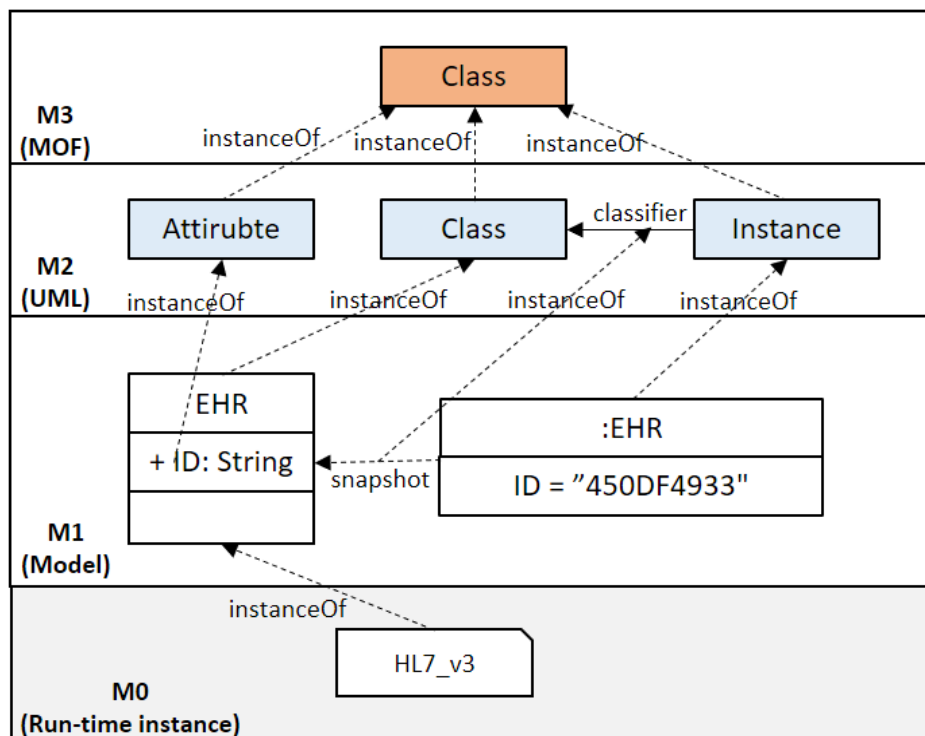
Το Metamodel Layer βασίζεται στην ιδέα της Μοντελοστραφούς Μηχανικής (MDE) για τη χρήση μοντέλων σε διαφορετικά επίπεδα αφαίρεσης για την ανάπτυξη συστημάτων. Γενικά, οι προσεγγίσεις αυτές διευκολύνουν την ανάπτυξη επίσημων και διατηρήσιμων λύσεων, αποτελώντας μια βέλτιστη τεχνολογική υποδομή για την επίτευξη του στόχου μας. Στο MDE, ο

μετασχηματισμός μοντέλου είναι η διαδικασία μετατροπής ενός μοντέλου M1 σύμφωνα με το μεταμοντέλο MM1, σε ένα μοντέλο M2 που συμμορφώνεται προς το μεταμοντέλο MM2. Τα τελευταία χρόνια, έχουν οριστεί αρκετές γλώσσες μετασχηματισμών μοντέλων, όπως το RubyTL [249], μια γλώσσα βασισμένη σε κανόνες υβριδικών μετασχηματισμών για τον καθορισμό των κανόνων μετασχηματισμού με πολλαπλούς τρόπους. Σε αυτή την προσέγγιση, μετασχηματισμοί μοντέλων που εκφράζονται σε RubyTL θα χρησιμοποιηθούν για τον μετασχηματισμό των μοντέλων, ενώ οι τελευταίοι θα εκπροσωπούνται σε οντολογίες OWL χρησιμοποιώντας το Protégé Framework, μαζί με το μοντέλο της γλώσσας MOFScript [250]. Πιο συγκεκριμένα, στην περίπτωση μας, υποθέτουμε ότι ένα μοντέλο είναι μια περίπτωση ενός συγκεκριμένου μεταμοντέλου, για αυτό και οι οντολογίες που αναπτύχθηκαν για τα διάφορα ιατρικά πρότυπα των EHR θα εκφραστούν από τα μεταμοντέλα. Για το σκοπό αυτό, θα χρησιμοποιηθεί το πρότυπο Ontology Definition Metamodel [251], το οποίο καθορίζει τη σημασιολογία του μετασχηματισμού μοντέλων σε οντολογίες OWL.

Μόλις επιτευχθούν συντακτικά τα μεταμοντέλα, ορίζονται οι αντιστοιχίες μεταξύ τους. Προκειμένου να μετασχηματιστούν τα δεδομένα EHR ενός συγκεκριμένου ιατρικού προτύπου στην οντολογική του αναπαράσταση (Ontology Layer), οι αντιστοιχίσεις μεταξύ του συγκεκριμένου προτύπου και του κοινού μεταμοντέλου ορίζονται τόσο στην έννοια όσο και στα επίπεδα ιδιοτήτων. Κάθε έννοια και ιδιότητα των μεταμοντέλων των ιατρικών προτύπων μπορεί να διαμορφωθεί σύμφωνα με την κοινή αναπαράσταση. Συνοπτικά, ο μετασχηματισμός ενός ιατρικού προτύπου EHR στην αναπαράσταση του μεταμοντέλου του αποτελείται από τις ακόλουθες φάσεις (Εικόνα 4-11):

1. Η είσοδος τυποποιημένου EHR (M0) μετατρέπεται σε αναπαράσταση μοντέλου (M1).
2. Η αναπαράσταση μοντέλου (M1) μετασχηματίζεται σε γενικευμένο τρόπο ακολουθώντας την αναπαράσταση μοντέλου UML (M2), όπου οι διαφορετικές περιπτώσεις της αναπαράστασης μοντέλου κατηγοριοποιούνται σε διαφορετικές κατηγορίες.
3. Όλες οι ταξινομημένες περιπτώσεις του μοντέλου UML κατηγοριοποιούνται σε μια γενική κλάση, παίρνοντας την τελική αναπαράσταση Metamodel χρησιμοποιώντας το πρότυπο Meta-Object Facility (MOF) [252] (M3), προκειμένου να μοντελοποιήσει την αφηρημένη σύνταξη του μοντέλου UML.

Όσον αφορά το κοινό μοντέλο, μετατρέπεται στο κοινό μεταμοντέλο ακολουθώντας τον ίδιο μηχανισμό. Και τα δύο μεταμοντέλα δίδονται στην συντακτική έκφρασή τους στο Ontology Layer, ώστε να μετασχηματιστούν στην οντολογική τους αναπαράσταση.



Εικόνα 4-11 - Μετασχηματισμός ενός ιατρικού προτύπου EHR στην αναπαράσταση του μεταμοντέλου

Ontology Layer

Το Ontology Layer παρέχει την επίσημη σημασιολογία του τομέα και αποτελείται από μια σειρά οντολογιών OWL που αναπτύχθηκαν για τα πολλαπλά ιατρικά πρότυπα του EHR. Με περισσότερες λεπτομέρειες, το Ontology Layer κατασκευάζεται με την αναγνώριση της κοινής και διακεκριμένης γνώσης που ορίζεται στις οντολογίες των πολλαπλών ιατρικών προτύπων του EHR, έτσι ώστε να μπορεί να θεωρηθεί ως παγκόσμια οντολογία EHR. Στο επίπεδο οντολογίας, η ανίχνευση των ισοδύναμων εννοιών και τύπων δεδομένων υποστηρίζεται από τη μεθοδολογία ενσωμάτωσης οντολογιών που αναπτύχθηκε στο [253]. Πιο συγκεκριμένα, μπορούν να παρατηρηθούν τρεις διαφορετικοί τύποι οντολογιών:

- Η **τυποποιημένη οντολογία EHR** αναφέρεται στις οντολογίες που κατασκευάζονται μέσω γνωστών και εύκολα αναγνωρίσιμων κατηγοριών των δεδομένων της υγειονομικής περίθαλψης που αντιπροσωπεύει κάθε ιατρικό πρότυπο. Για παράδειγμα, οι έννοιες όπως το ACTIVITY και το

ISM_TRANSITION ορίζονται μόνο στο OpenEHR. Έτσι, προστέθηκαν στην πρότυπη οντολογία της EHR ως ενιαία έννοια.

- Η **κοινή οντολογία** αναφέρεται στις οντολογίες που κατασκευάζονται μέσω των κοινών μερών και δομών των διαφόρων ιατρικών προτύπων, τα οποία εντοπίζονται κατά τη διάρκεια της διαδικασίας επεξεργασίας και της διαδικασίας μετασχηματισμού. Για παράδειγμα, οι έννοιες όπως FOLDER, COMPOSITION, SECTION ή ELEMENT είναι κοινές σε πολλά ιατρικά πρότυπα. Έτσι, προστέθηκαν στην κοινή οντολογία ως ενιαία έννοια.
- Η **οντολογία DSL** αναφέρεται στις οντολογίες που σχετίζονται με συγκεκριμένο τομέα υγειονομικής περίθαλψης και μπορεί να χρησιμοποιηθεί/επαναχρησιμοποιηθεί για τον εντοπισμό τυχόν δεδομένων της υγειονομικής περίθαλψης που δεν μπορούν να κατηγοριοποιηθούν ή προσδιοριστούν. Στην περίπτωση μας, τα μεταμοντέλα που συγκεντρώνονται από το Metamodel Layer μετασχηματίζονται στην οντολογική τους αναπαράσταση χρησιμοποιώντας το Protégé Framework, σε συνδυασμό με οντολογίες, πληροφορίες και μεταδεδομένα που ανήκουν στις οντολογίες DSL.

OWL EHR

Στο στοιχείο OWL EHR, τα σύνολα δεδομένων των διαφόρων ιατρικών προτύπων έχουν μετατραπεί στην οντολογική τους αναπαράσταση, ενώ παρέχουν πληροφορίες μεταδεδομένων για να διαμορφώσουν τις οντολογίες DSL.

Ως αποτέλεσμα της περιγραφόμενης διαδικασίας, διαμορφώθηκε η οντολογική αναπαράσταση διαφόρων ιατρικών προτύπων και μπορεί να ενσωματωθεί σε διαφορετικά συστήματα υγειονομικής περίθαλψης, έτσι ώστε τα ετερογενή σύνολα δεδομένων EHR να μετασχηματιστούν σε κατανοητές, διαλειτουργικές και ανταλλασσόμενες γνώσεις που μπορούν να χρησιμοποιηθούν για μια καλύτερη επικοινωνία μεταξύ των επαγγελματιών του τομέα της υγείας. Πρέπει να σημειωθεί ότι οι κατασκευασμένες οντολογικές αναπαραστάσεις αποθηκεύονται στη συνέχεια σε ένα χώρο αποθήκευσης δεδομένων που βασίζεται στην οντολογία (GraphDB), που μπορεί εύκολα να ανακτηθεί και να χρησιμοποιηθεί.

4.1.4 Στάδιο Οντολογικού Αντιστοιχιστή FHIR (FHIR Ontology Mapper - FOM): Συγκέντρωση Συντακτικών και Σηματολογικών Ομοιοτήτων Οντολογιών προς την αντιστοίχισή τους σε HL7 FHIR

Σε αυτό το πλαίσιο, στην παρούσα προσέγγιση προτάθηκε ένας μηχανισμός για την επίτευξη διαλειτουργικότητας μέσω της μετάφρασης των οντολογιών που σχετίζονται με την υγειονομική περίθαλψη στην αντιστοιχη δομή τους HL7 FHIR. Λεπτομερέστερα, παρουσιάστηκε ο μηχανισμός FOM, ο οποίος στοχεύει καταρχάς στην κατασκευή οντολογιών για την υγειονομική περίθαλψη, προσδιορίζοντας τις διαφορετικές έννοιες και σχέσεις μεταξύ δομημένων δεδομένων που σχετίζονται με την υγειονομική περίθαλψη. Στη συνέχεια, ο μηχανισμός αυτός προσδιορίζει τη συντακτική και τη σηματολογική ομοιότητα κάθε διαφορετικής οντολογίας με τις οντολογίες των πόρων HL7 FHIR. Κατόπιν, σύμφωνα με τον μέσο όρο τόσο της συντακτικής όσο και της σηματολογικής ανάλυσης, πραγματοποιείται η τελική αντιστοίχιση με τους πόρους HL7 FHIR, μετασχηματίζοντας τα στοιχεία της υγειονομικής περίθαλψης σε ένα ευρέως υιοθετημένο ιατρικό πρότυπο. Συνοπτικά, η παρούσα προσέγγιση αρχικά μελέτησε πώς μεταβαίνοντας στο 2019, τα νοσοκομεία και τα συστήματα υγειονομικής περίθαλψης εξακολουθούν να αντιμετωπίζουν πολλές προκλήσεις όσον αφορά την εφαρμογή, τη διατήρηση και την αναβάθμιση των συστημάτων υγειονομικής περίθαλψης. Τέτοιες προκλήσεις κυμαίνονται από την τεχνική μέχρι την ασφάλεια, τη στρατηγική και την ανθρώπινη αλληλεπίδραση. Ωστόσο, μεταξύ των μεγάλων προκλήσεων που αντιμετωπίζει σήμερα ο τομέας της υγειονομικής περίθαλψης είναι η συνεχής εστίαση στη διαλειτουργικότητα. Τα συστήματα πρέπει να είναι σε θέση να μιλούν μεταξύ τους για να αποκτήσουν με επιτυχία την πλήρη εικόνα ενός ασθενούς. Έχοντας κατά νου ότι η αγορά φορητών ιατρικών συσκευών αναμένεται να φθάσει τα 14.41 δισ. δολάρια μέχρι το 2022 από 6.22 δισ. δολάρια το 2017, είναι σαφές ότι μια αποτελεσματική λύση στην ασυμβατότητα των ανταλλαγών δεδομένων είναι εξαιρετικής σημασίας [254]. Σύμφωνα με το [255], το 2016, τα νοσοκομεία και τα ιατρεία σε εθνικό επίπεδο θα μπορούσαν να αποφύγουν σχεδόν 2000 θανάτους ασθενών, αν το ιατρικό προσωπικό και οι ασθενείς είχαν επικοινωνήσει καλύτερα. Αυτό έχει ως αποτέλεσμα να καταστεί ακόμη πιο δύσκολο το έργο διαλειτουργικότητας στον τομέα της υγείας, παρά το γεγονός ότι πολλές τεχνικές τυποποίησης επενδύονται ετησίως στη διαλειτουργικότητα των δεδομένων για την υγεία. Μια μελέτη [256] υπολόγισε ότι θα μπορούσαν να επιτευχθούν ετήσιες μειώσεις κόστους εάν χρησιμοποιούνταν πρότυπα ανταλλαγής δεδομένων στον τομέα της υγείας.

Είναι αναμφισβήτητο ότι η παροχή και η έρευνα για την υγειονομική περίθαλψη πολλαπλών εγκαταστάσεων απαιτεί την αναδιάρθρωση των δεδομένων που σχετίζονται με την υγειονομική περίθαλψη σε κοινή μορφή και τυποποιημένες ορολογίες που συνδέονται με άλλες πηγές δεδομένων. Ο τελευταίος διατίθεται σήμερα μέσω του προτύπου HL7 FHIR, το οποίο υιοθετείται ευρέως μεταξύ των διαφόρων φορέων παροχής υγειονομικής περίθαλψης για την επίτευξη διαλειτουργικότητας, καθώς είναι απαραίτητο να βελτιωθεί η ποιότητα και η ασφάλεια της περίθαλψης, της έρευνας και της υγείας διαχείριση υπηρεσιών [257]. Ωστόσο, η υπόσχεση ενός παγκόσμιου ιατρικού προτύπου είναι ακόμη χρόνια μακριά, όπου παρά το γεγονός ότι το HL7 FHIR χρησιμοποιείται ως επί το πλείστον, χρειάζεται ακόμα πολύς χρόνος για να υιοθετηθεί, καθώς υπάρχουν πολλά συστήματα, συσκευές και πλατφόρμες που παράγουν δεδομένα που δεν σχετίζονται με το HL7 FHIR Πρότυπο. Στην περίπτωση αυτή, η διαλειτουργικότητα μπορεί να επιτευχθεί με την κατασκευή οντολογιών ιατρικού τομέα για την απεικόνιση συστημάτων ιατρικής ορολογίας [258]. Για το λόγο αυτό, οι οντολογίες έχουν γίνει ένα δημοφιλές ερευνητικό θέμα στο πλαίσιο της υγειονομικής περίθαλψης [259]. Συνεπώς, καθίσταται σαφές ότι το κλειδί για την επίτευξη και τη βελτίωση της διαλειτουργικότητας της υγειονομικής περίθαλψης μπορεί να βρεθεί μεταξύ οντολογιών, με τη μορφή μόνιμου αντικειμένου των προδιαγραφών κάθε ιατρικής έννοιας. Ωστόσο, παρά τη δημιουργία οντολογιών που σχετίζονται με την υγειονομική περίθαλψη, πρέπει επίσης να πραγματοποιηθεί αντιστοίχιση οντολογιών προκειμένου να μειωθεί το σημασιολογικό χάσμα μεταξύ των διαφόρων επικαλυπτόμενων αναπαραστάσεων του ίδιου τομέα. Οι οντολογίες και οι οντολογικές τεχνικές αντιστοίχισης είναι μια αυξανόμενη τάση, καθώς οι οντολογίες παρέχουν ίσως την πιο ενδιαφέρουσα ευκαιρία να κωδικοποιήσουν το νόημα των πληροφοριών που σχετίζονται με την υγειονομική περίθαλψη.

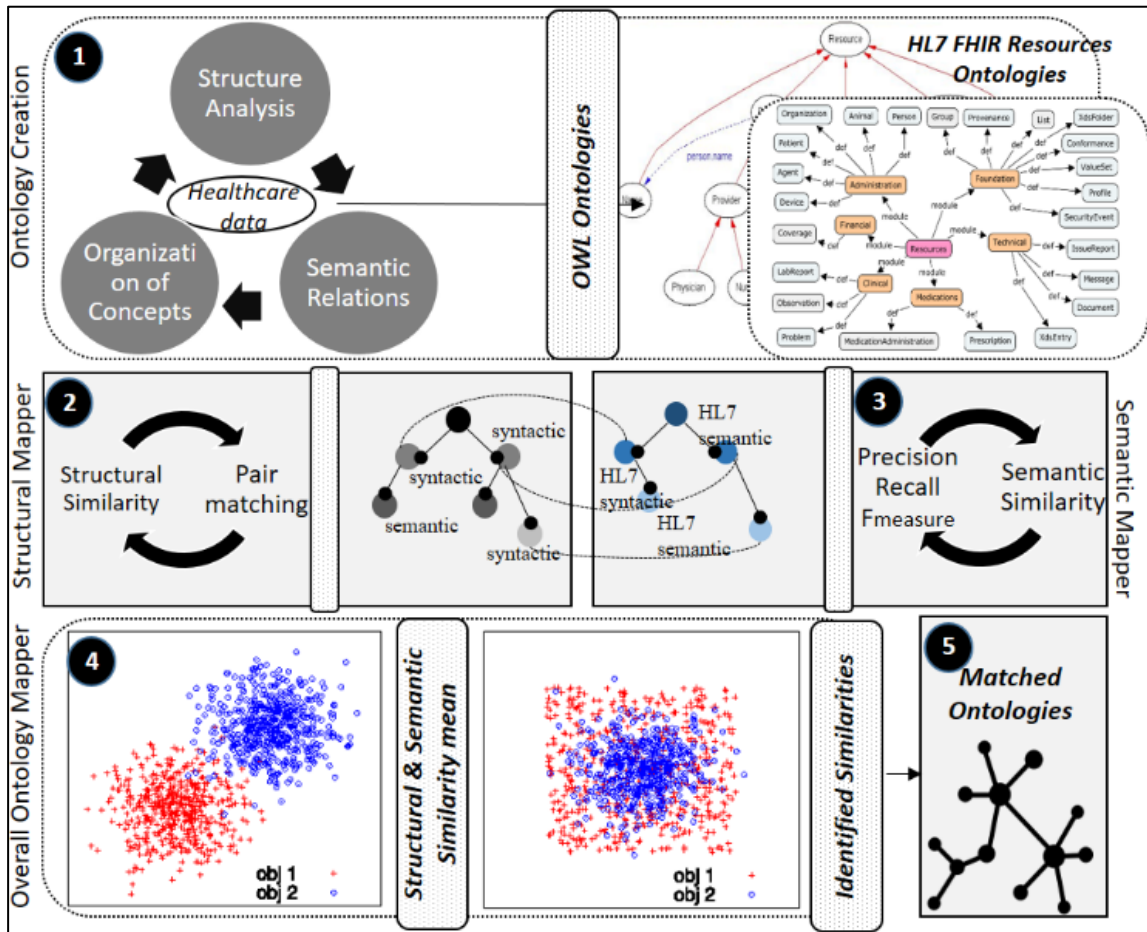
Είναι αναμφισβήτητο ότι έχουν προταθεί διάφορες λύσεις που επιτρέπουν την ευθυγράμμιση και αντιστοίχιση των οντολογιών που σχετίζονται με την υγειονομική περίθαλψη που εμπεριέχουν διαφορετικές κλάσεις, ιδιότητες και στιγμιότυπα, προς τη διαλειτουργικότητα για την επίτευξη ταχύτερης και αποτελεσματικότερης λήψης αποφάσεων. Ωστόσο, ένα από τα βασικά μειονεκτήματα αυτών των προσεγγίσεων είναι ότι δεν προτείνουν λύσεις που μπορούν να εφαρμοστούν σε διαφορετικά ιατρικά πρότυπα και εισερχόμενα δεδομένα. Αυτό που λείπει αναφέρεται σε ένα γενικευμένο μηχανισμό αντιστοίχισης οντολογιών, ως γενική προσέγγιση που μπορεί να αντιμετωπίσει διαφορετικούς τύπους οντολογιών σχετιζόμενων με την υγειονομική περίθαλψη, που προέρχονται τόσο από γνωστές όσο και από άγνωστες

συσκευές που μπορούν να συνδυαστούν άμεσα με ένα παγκοσμίως χρησιμοποιούμενο ιατρικό πρότυπο (HL7 FHIR) . Έτσι, θα μπορούσαν να συγκεντρωθούν πολλαπλές ομάδες ετερογενών δεδομένων που σχετίζονται με την υγειονομική περίθαλψη, να αποκτήσουν γνώσεις και να προσδώσουν μεγαλύτερη αξία στο οικοσύστημα της υγείας. Για να αντιμετωπιστεί αυτό το κενό, προτάθηκε το FOM, μια γενικευμένη προσέγγιση που είναι ικανή να εντοπίσει και να κατασκευάσει οντολογίες που σχετίζονται με την υγειονομική περίθαλψη από διαφορετικά σύνολα δεδομένων και στη συνέχεια να βρει ομοιότητες και κοινές συνδέσεις μεταξύ αυτών των οντολογιών και οντολογιών που αντιπροσωπεύουν τους πόρους HL7 FHIR . Σε αυτό το πλαίσιο, όχι μόνο η αντιστοίχιση θα πραγματοποιηθεί με βάση τις σημασιολογικές σημασίες των οντολογιών, αλλά η συντακτική αναπαράσταση των οντολογιών θα είναι επίσης ένα μέρος αυτής της αντιστοίχισης, όπου διαφορετικές πιθανότητες ομοιότητας θα αποδοθούν τόσο στο πλαίσιο των συντακτικών όσο και των σημασιολογικών αντιστοιχίσεων. Στο τέλος, η προτεινόμενη προσέγγιση συγκεντρώνει τα αποτελέσματα των δύο μετρήσεων ομοιότητας και αντιστοιχεί μια οντολογία με έναν συγκεκριμένο πόρο HL7 FHIR, σύμφωνα με το μέσο όρο της συντακτικής και σημασιολογικής τους ομοιότητας.

Με περισσότερες λεπτομέρειες, τα προκαταρκτικά βήματα της αναπτυγμένης προσέγγισης (Ontology Creation) ασχολούνταν με την αυτόματη μετάφραση του συνόλου δεδομένων που σχετίζεται με την υγειονομική περίθαλψη και των πόρων HL7 FHIR στην οντολογική δομή τους. Θα πρέπει να αναφερθεί ότι η οντολογική δομή περιέχει τόσο τη δομή όσο και τη σημασιολογία κάθε διαφορετικής οντολογίας που θα χρησιμοποιηθεί ως είσοδος στα εναπομείναντα στάδια της προσέγγισης FOM. Μόλις κατασκευαστούν οι οντολογίες, εκτελούνται οι δύο ξεχωριστοί υπο-μηχανισμοί, που παρέχουν (i) τη συντακτική ομοιότητα (Structural Mapper) και (ii) τη σημασιολογική ομοιότητα (Semantic Mapper) μεταξύ δύο διαφορετικών οντολογιών, τις αναπαραστάσεις και τις σημασιολογικές γνώσεις τους ανάλογα. Το τελικό στάδιο της προσέγγισης FOM υλοποιεί ένα επιπρόσθετο υπο-μηχανισμό (Overall Ontology Mapper) που συγκεντρώνει και συγχωνεύει τα αποτελέσματα των προαναφερθέντων υπο-μηχανισμών, παρέχοντας τα συνολικά αποτελέσματα αντιστοίχισης. Η Εικόνα 4-12 απεικονίζει τη συνολική αρχιτεκτονική του FOM, απεικονίζοντας τους τέσσερις προαναφερθέντες υπο-μηχανισμούς.

Δημιουργία Οντολογιών

Η μέθοδος Δημιουργίας Οντολογιών παρουσιάζει έναν αυτόματο τρόπο απόκτησης μιας αρχικής οργάνωσης των εννοιών από τη συλλογή οποιωνδήποτε εγγράφων που μπορούν να σχηματιστούν μόνο με κείμενο ή δομή και κείμενο. Αυτή η μέθοδος βελτιώνει ορισμένες λύσεις που χρησιμοποιούν τη συσσώματωση και τις στατιστικές μεθόδους για να αποκτήσουν οντολογίες που είναι πιο σημαντικές, βελτιώνοντας την ανάπτυξη των εννοιών και των σημασιολογικών σχέσεων.



Εικόνα 4-12 - Αρχιτεκτονική Οντολογικού Αντιστοιχιστή FHIR

Στην περίπτωση μας, για να δημιουργηθούν οι οντολογίες, ακολουθούνται τα παρακάτω βήματα:

- Στο πρώτο βήμα, προετοιμάζονται σύνολα δεδομένων για την απόκτηση των εννοιών. Σε αυτήν την περίπτωση, είναι απαραίτητο να αναλυθεί η συλλογή των συνόλων δεδομένων προκειμένου να προσδιοριστούν τα σύνολα δεδομένων που έχουν δομή (π.χ. έγγραφα XML) ή όχι. Για σύνολα δεδομένων που έχουν δομή, η δομή αναλύεται για να επαληθεύσει εάν τα στοιχεία έχουν ορισμούς που μπορούν να θεωρηθούν ως έννοιες στην

οντολογία. Τα στοιχεία αγνοούνται εάν η δομή τους δεν είναι σχετική, διαφορετικά αποθηκεύονται. Για τα δομημένα σύνολα δεδομένων, ο ορισμός μπορεί να περιέχει στοιχεία δομής που σχετίζονται με τη διαμόρφωση των εννοιών.

- Το επόμενο βήμα είναι να αποκτηθούν οι σημασιολογικές σχέσεις για κάθε μια από τις οντολογίες που θα δημιουργηθούν, αναφερόμενες στις ιδιότητες, τις έννοιες και τα αξιώματα. Μπορούν να οριστούν δύο τύποι ιδιοτήτων: (α) οι ιδιότητες αντικειμένων που σχετίζονται με άλλες περιπτώσεις που καθορίζουν περιορισμούς και συμπεριφορές και (β) οι ιδιότητες τύπων δεδομένων που αναφέρονται σε ιδιότητες που εκφράζουν μόνο τιμές, π.χ., συμβολοσειρές ή αριθμούς. Οι κλάσεις μπορούν να έχουν υπερ και υποκλάσεις, παρέχοντας έναν μηχανισμό εξορθολογισμού και κληρονομιάς της ιδιότητας. Για το σκοπό αυτό, τα στιγμιότυπα χρησιμοποιούνται για την παροχή πληροφοριών σχετικά με τις κλάσεις και τις ιδιότητες, όπως για παράδειγμα, για να προσδιοριστεί η ισοδυναμία δύο κλάσεων ή το εύρος μιας ιδιότητας. Υπάρχουν πολλές σημασιολογικές σχέσεις που μπορούν να ληφθούν υπόψη, συμπεριλαμβανομένων των ιδιοτήτων όπως είναι η υποκλάση μεταξύ των κλάσεων των διαφορετικών επιπέδων που αποτελούν την οντολογία. Επιπλέον, μπορούν να οριστούν σχέσεις όπως, *equivalent_to* (μεταξύ συνωνύμων ή παρόμοιων κλάσεων), *disjoin_of* (μεταξύ αντωνυμικών κλάσεων), *part_of* (μεταξύ όρων που συμπληρώνουν άλλους όρους) και *inverse_of* (μεταξύ αντωνύμων και συνωνύμων). Για να καθορίσουμε αυτές τις σχέσεις, οι έννοιες αναλύονται χρησιμοποιώντας το Wordnet, ελέγχοντας πιθανές συσχετίσεις μεταξύ των θεωρημένων εννοιών. Για αυτές τις συσχετίσεις, αναλύονται εκείνες που είναι κατάλληλες για χρήση στον ορισμό της οντολογίας όπου στην περίπτωση που οι έννοιες είναι συνώνυμες, τους δίνεται ένα αξίωμα που καθορίζεται με ισοδυναμία.
- Το επόμενο βήμα έχει να κάνει με την οργάνωση των κλάσεων και των σημασιολογικών σχέσεων που οργανώνονται σε οντολογίες που αποθηκεύονται σε αρχεία που κωδικοποιούνται σε γλώσσα OWL. Αυτή η γλώσσα χρησιμοποιείται για τον ορισμό οντολογιών και παρέχει μηχανισμούς για τη δημιουργία συστατικών: κλάσεις, στιγμιότυπα, ιδιότητες και αξιώματα. Συνεπώς, χρησιμοποιώντας όλες τις κλάσεις και τις ιδιότητες που αποκτήθηκαν, μπορεί να δημιουργηθεί μια ενιαία οντολογία που περιγράφει ολόκληρη τη συλλογή. Πρέπει να αναφερθεί ότι η ίδια διαδικασία επαναλαμβάνεται για όλα τα διαφορετικά σύνολα δεδομένων που σχετίζονται με την υγειονομική περίθαλψη, προκειμένου να αποκτηθεί η οντολογική τους δομή.

Όσον αφορά τους πόρους του HL7 FHIR, έχουν ήδη δομηθεί στην οντολογική τους μορφή μέσω της FHIR Linked Data Module [260], η οποία αποτελεί μια τεχνική που περιγράφει τα RDF για τους πόρους FHIR (FHIR / RDF), συμπεριλαμβανομένης μιας οντολογίας OWL για το FHIR / RDF και μιας γραμματικής ShEx [261] για την επικύρωση του FHIR / RDF.

Συντακτική Αντιστοίχιση

Η μέθοδος συντακτικής αντιστοίχισης παρέχει έναν τρόπο για την αντιστοίχιση και την αναγνώριση της συντακτικής ομοιότητας μεταξύ δύο διαφορετικών οντολογιών, με βάση τη συντακτική τους μορφή. Ο στόχος αυτής της μεθόδου είναι να προσδιορίσει το μέτρο ομοιότητας μεταξύ των διαφορετικών οντολογιών και να παράσχει την πιθανότητα μια συγκεκριμένη οντολογία να είναι η ίδια - από την άποψη της συντακτικής της ερμηνείας - με μια άλλη οντολογία. Στην περίπτωση μας, προκειμένου να μετρηθεί η συντακτική ομοιότητα μεταξύ δύο οντολογιών, ακολουθούνται τα εξής βήματα:

- Αρχικά, γίνεται ο προσδιορισμός των δύο οντολογιών, των οποίων η συντακτική ομοιότητα θα μετρηθεί. Με περισσότερες λεπτομέρειες, ο Structural Mapper αναγνωρίζει αυτόματα και επαναλαμβάνεται πάνω από κάθε διαφορετική οντολογία που έχει δημιουργηθεί από το προηγούμενο βήμα και παρέχει κάθε οντολογία ως είσοδο στον αναπτυγμένο μηχανισμό.
- Το επόμενο βήμα ασχολείται με τη μετατροπή τόσο της δομικής αναπαράστασης των οντολογιών στους κεφαλαίους χαρακτήρες τους, οι οποίοι στη συνέχεια χωρίζονται σε διαφορετικά ζεύγη χαρακτήρων (π.χ. : PATIENT διαιρείται σε {PA, TI, ENT}). Ένα μέρος του μηχανισμού επαναλαμβάνεται πάνω από τη συντακτική μορφή κάθε οντολογίας και προσδιορίζει τα πρότυπα λέξεων - από λέξεις που συχνά συναντώνται και επαναλαμβάνονται. Αφού εντοπιστούν αυτά τα πρότυπα, τότε χωρίζονται και αποθηκεύονται σε διαφορετικούς πίνακες για κάθε οντολογία, με αποτέλεσμα την αναγνώριση των διαφορετικών ζευγών χαρακτήρων που πληρούνται.
- Το επόμενο βήμα αφορά τον έλεγχο των ζευγών πολλαπλών χαρακτήρων, προκειμένου να προσδιοριστούν οι χαρακτήρες που μπορούν να βρεθούν και στις δύο διαχωρισμένες σειρές. Με περισσότερες λεπτομέρειες, αυτό το συγκεκριμένο τμήμα του Structural Mapper επαναλαμβάνεται πάνω από κάθε διαφορετικό προσδιορισμένο ζεύγος και βρίσκει για κάθε οντολογία όλα τα ζεύγη που έχουν ακριβή αντιστοιχία με ένα διαφορετικό ζεύγος. Το αποτέλεσμα αυτού του βήματος είναι τα συνολικά ζεύγη διαφορετικών

χαρακτήρων που είναι παρόμοια μεταξύ των οντολογιών που συγκρίνονται μεταξύ τους.

- Το τελευταίο βήμα αφορά την αναγνώριση της πιθανότητας ομοιότητας σύμφωνα με τον τύπο που παρουσιάζεται στο (8). Συνοπτικά, η συντακτική ομοιότητα (St_Simil) μεταξύ δύο δεδομένων οντολογιών $S1$ και $S2$ είναι διπλάσια από τον αριθμό των ζευγών χαρακτήρων που είναι κοινά και στις δύο οντολογίες, διαιρούμενο με το άθροισμα του αριθμού των ζευγών χαρακτήρων που αναγνωρίζονται και στις δύο οντολογίες.

$$St_Simil = \frac{2 * char_pairs(S1) \cap char_pairs(S2)}{char(S1) + char(S2)} \quad (8)$$

Σημασιολογική Αντιστοιχισή

Η μέθοδος της σημασιολογικής αντιστοιχισής παρέχει τα μέσα ευθυγράμμισης και αντιστοιχισής των διαφορετικών οντολογιών, σύμφωνα με τη σημασιολογική σημασία τους. Η διαδικασία περιλαμβάνει τη διεξαγωγή αρκετών λειτουργιών αντιστοιχισής, ανάλογα με την ομοιότητα μεταξύ των αποθηκευμένων οντολογιών, και στη συνέχεια το φιλτράρισμα των αποτελεσμάτων τους για να βρεθεί μια συνολική ευθυγράμμιση. Μόλις ευθυγραμμιστεί, γίνεται άμεση σύνδεση και αντιστοιχισή μεταξύ των δημιουργημένων οντολογιών. Στην περίπτωση της σημασιολογικής αντιστοιχισής, οι σημασιολογικές αντιστοιχίες μεταξύ οντολογιών βασίζονται στις σχέσεις και τις εξαρτήσεις μεταξύ της δομής των κλάσεων και της τοποθέτησης των στιγμιότυπων. Στην περίπτωση μας, για να μετρήσουμε τη σημασιολογική ομοιότητα μεταξύ δύο οντολογιών, ακολουθούνται τα εξής βήματα:

- Το πρώτο βήμα ασχολείται με την αναγνώριση της ομοιότητας ονόματος που υπάρχει ανάμεσα στις οντολογίες, χρησιμοποιώντας σημασιολογικές προσεγγίσεις που στοχεύουν στον εντοπισμό ομοιότητας μεταξύ των ονομάτων των διαφόρων στοιχείων.
- Το επόμενο βήμα αναφέρεται στην αναγνώριση της δομικής ομοιότητας μεταξύ των οντολογιών και υπολογίζεται στην περίπτωση που και οι δύο οντολογίες έχουν κοινά μέλη ή σχέσεις.
- Το ακόλουθο βήμα έχει να κάνει με τον προσδιορισμό της ομοιότητας των παρουσιών που γίνεται με την εξέταση των περιπτώσεων δεδομένων από τις δύο οντολογίες. Μέχρι τη στιγμή που τα βήματα αυτά εκτελούνται για κάθε διαφορετική οντολογία, παρέχονται διαφορετικές πιθανότητες της

αντιστοιχίας μεταξύ των πολλαπλών οντολογιών. Μόλις ολοκληρωθεί η αντιστοίχιση μεταξύ των σημασιολογικών σημασιών των διαφορετικών οντολογιών, προσδιορίζονται οι μετρήσεις της ακρίβειας και της ανάκλησης των τελικών αποτελεσμάτων, προκειμένου να υπολογιστεί τελικά ο αρμονικός μέσος όρος (μέτρο F-measure) αυτών των δύο μέτρων, με σεβασμό των τεχνικών UMLS. Σε περισσότερες λεπτομέρειες, υπολογίζονται οι ακόλουθες μετρήσεις:

- Η ακρίβεια των αποτελεσμάτων (precision) που αναφέρεται στο κλάσμα των ανακτηθέντων δεδομένων που σχετίζονται με το ερώτημα, εμφανίζοντας τον αριθμό των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό όλων των αποτελεσμάτων που επιστράφηκαν (9).

$$precision = \frac{(relevant_data \cap retrieved_data)}{retrieved_data} \quad (9)$$

- Τα αποτελέσματα της ανάκλησης (recall) αναφέρονται στο κλάσμα των σχετικών δεδομένων που ανακτώνται, εμφανίζοντας τον αριθμό των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό των αποτελεσμάτων που θα έπρεπε να έχουν επιστραφεί (10).

$$recall = \frac{(relevant_data \cap retrieved_data)}{relevant_data} \quad (10)$$

- Το μέτρο σημασιολογικής ομοιότητας Fmeasure (Sem_Simil) που αναφέρεται στον αρμονικό μέσο (δηλαδή συνδυασμό) της ακρίβειας και της ανάκλησης (11).

$$Sem_Simil = \frac{2 * (precision * recall)}{precision + recall} \quad (11)$$

Αυτό το τμήμα του μηχανισμού παρέχει τελικά τον αριθμό που υποδεικνύει πόσο μια συγκεκριμένη οντολογία είναι σημασιολογικά η ίδια με μια άλλη οντολογία. Μέχρι τη στιγμή που ο Semantic Mapper ολοκληρώνει τη δουλειά του, αποθηκεύονται τα συνολικά αποτελέσματα αντιστοίχισης και παρέχονται μαζί με τα αποτελέσματα του Structural Mapper στον Ολικό Αντιστοιχιστή Οντολογιών (Overall Ontology Mapper).

Ολική Οντολογική Αντιστοίχιση

Η μέθοδος ολικής οντολογικής αντιστοίχισης παρέχει έναν τρόπο που είναι σε θέση να συγκεντρώσει και να συγχωνεύσει τα αποτελέσματα που προέκυψαν από τους προηγούμενους μηχανισμούς (Structural Mapper και Semantic Mapper), προκειμένου να προσδιοριστούν τελικά οι οντολογίες που έχουν αντιστοιχηθεί σωστά. Η συνολική προσέγγιση ασχολείται με την αντιστοίχιση και τον εντοπισμό των οντολογιών δεδομένων σχετικά με την υγειονομική περίθαλψη και των οντολογιών των πόρων HL7 FHIR, προκειμένου να εντοπιστεί η ομοιότητα που υπάρχει μεταξύ οντολογιών διαφορετικών ιατρικών προτύπων. Ο στόχος της ολικής οντολογικής αντιστοίχισης είναι να βρει την προαναφερθείσα ομοιότητα και τελικά να μετασχηματίσει τις οντολογίες σε μορφή HL7 FHIR, η οποία υιοθετείται ευρέως μεταξύ διαφορετικών οργανισμών υγειονομικής περίθαλψης για την επίτευξη διαλειτουργικότητας. Στην περίπτωση μας, για να υπολογίσουμε το τελικό αποτέλεσμα των προαναφερθέντων μηχανισμών, ακολουθούνται τα ακόλουθα βήματα:

- Αρχικά, η μέθοδος ολικής οντολογικής αντιστοίχισης διερευνά τις μετρήσεις και τις τιμές που έχουν υπολογιστεί για κάθε διαφορετική οντολογία και παρέχει ένα μέσο μεταξύ των συντακτικών και σημασιολογικών ομοιοτήτων. Ο μέσος όρος αυτός υπολογίζεται ως το σύνολο της συντακτικής και της σημασιολογικής ομοιότητας, διαιρούμενος με δύο (12):

$$Overall_Mean = \frac{St_Simil + Sem_Simil}{2} \quad (12)$$

- Το επόμενο βήμα έχει να κάνει με τον τελικό εντοπισμό των οντολογιών των δεδομένων που σχετίζονται με την υγειονομική περίθαλψη και των πόρων του HL7 FHIR. Με περισσότερες λεπτομέρειες, σύμφωνα με τον υπολογισμό της συντακτικής και σημασιολογικής ομοιότητας, μια οντολογία χαρακτηρίζεται ότι αντιστοιχεί σε έναν συγκεκριμένο πόρο HL7 FHIR στην περίπτωση που υπερβαίνει το κατώτατο όριο των 0,9 (δηλ. 90% συντακτική και σημασιολογική ομοιότητα). Στην περίπτωση που ο μέσος όρος είναι μικρότερος από αυτό το όριο, τότε ο πόρος HL7 FHIR με υψηλότερη πιθανότητα ομοιότητας, θεωρείται αυτόματα ότι αντιπροσωπεύει την συγκεκριμένη οντολογία. Ο ίδιος μηχανισμός επαναλαμβάνεται για όλες τις αποθηκευμένες οντολογίες, προκειμένου να προσδιοριστούν τελικά οι οντολογίες των δεδομένων που σχετίζονται με την υγειονομική περίθαλψη και να μετασχηματιστούν σε HL7 FHIR.

4.2 Προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων

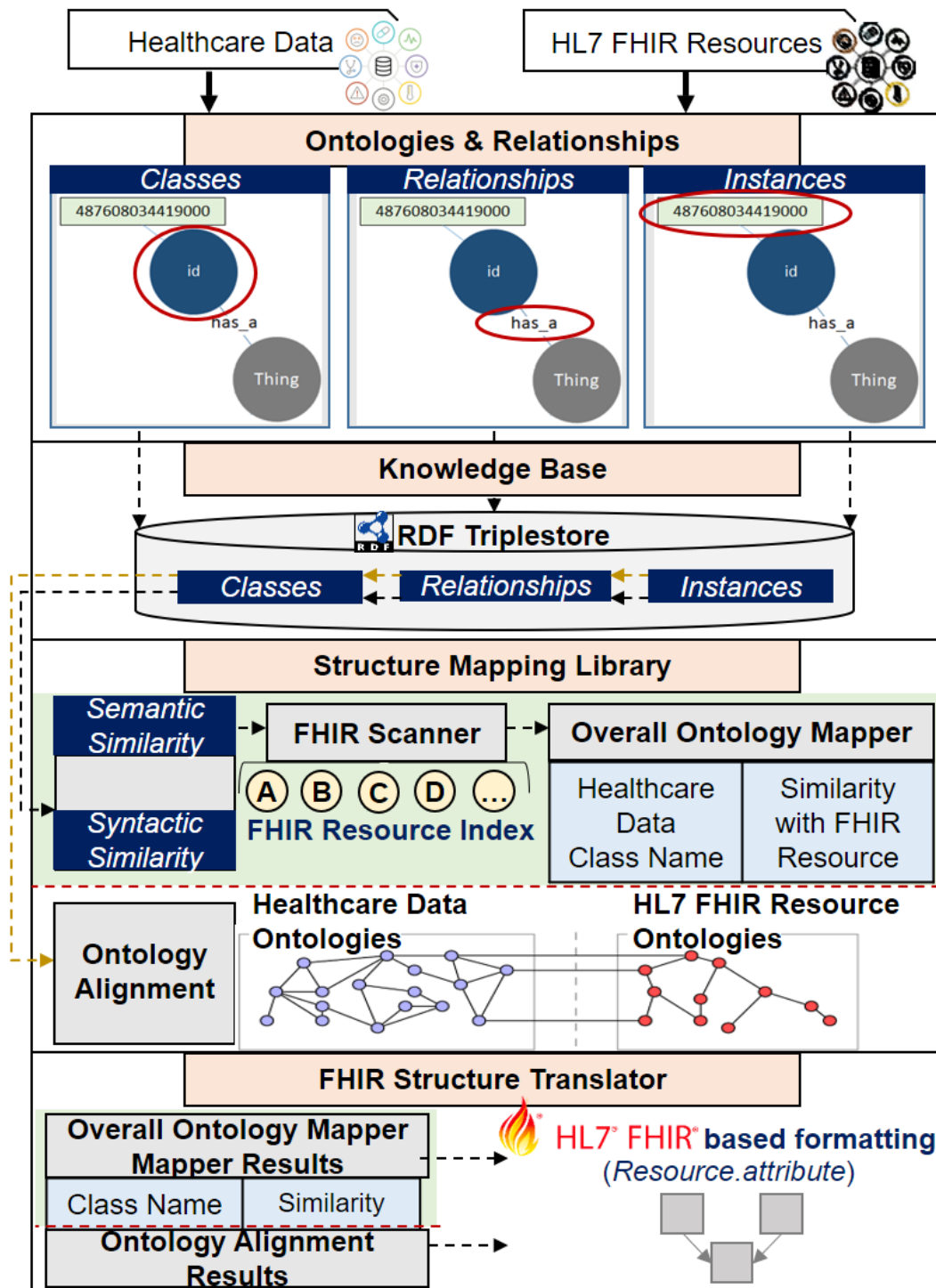
Σε σχέση με τις προσεγγίσεις που αναλύθηκαν στο υπο-Κεφάλαιο 4.1, η τελική προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων, αποτελεί έναν συνδυασμό των προηγούμενων προσεγγίσεων, καθώς και μια βελτιστοποιημένη εξέλιξη τους. Ειδικότερα, κατόπιν σχετικών πειραματικών δοκιμών και ερευνών, κάθε μία από τις προσεγγίσεις του υπο-Κεφαλαίου 4.1 κατέληξαν σε ερευνητικά αδιέξοδα, τα οποία απαιτούσαν έναν επανασχεδιασμό της κάθε φορά προτεινόμενης προσέγγισης. Ειδικότερα, καταλήξαμε στα ακόλουθα ερευνητικά αδιέξοδα:

- Για την προσέγγιση του υπο-Κεφαλαίου 4.1.1, εμβαθύνοντας και μελετώντας την αξία και σημασία των προτύπων και των οργανισμών προτυποποίησης, θεωρήθηκε σχεδόν αδύνατο να μπορούν να συμπεριληφθούν τα πρότυπα στην προσέγγιση του μηχανισμού, μιας και είχε σχεδιαστεί ώστε να ακολουθεί μια σειρά προκαθορισμένων βημάτων, χωρίς την απαραίτητη ελαστικότητα και ευελιξία.
- Για την προσέγγιση του υπο-Κεφαλαίου 4.1.2, τα περισσότερα από τα πλέον χρησιμοποιούμενα πρότυπα υγείας λήφθηκαν υπόψιν. Ωστόσο, αναφορικά με τα δεδομένα τα οποία δε μπορούσαν να αντιστοιχηθούν σε κάποιο από τα γνωστά πρότυπα υγείας είτε λόγω σφάλματος είτε λόγω της δομής τους, ο μηχανισμός αντίστροφης αντιστοίχισης ROM, δε μπορούσε να εξάγει σωστά αποτελέσματα στην περίπτωση που τα δεδομένα ξεπερνούσαν ένα συγκεκριμένο όγκο πληροφοριών. Παράλληλα, ήδη η HL7 FHIR είχε υιοθετηθεί σε πολλά συστήματα, με αποτέλεσμα η προτεινόμενη γλώσσα CHL να μην είναι σε θέση να την αντικαταστήσει, λόγω αδυναμιών σε περιπτώσεις μεγάλου όγκου δεδομένων.
- Για την προσέγγιση του υπο-Κεφαλαίου 4.1.3, η χρήση οντολογιών αποτέλεσε μία πολλά υποσχόμενη προσέγγιση προς μία ενιαία αναπαράσταση και κατανόηση των δεδομένων υγειονομικής περιθαλψής. Ωστόσο, το Metamodel Layer δεν ήταν τόσο αποδοτικό στις περιπτώσεις που η σύνταξη των δεδομένων EHR δεν ακολουθούσε κάποιο από τα γνωστά πρότυπα αναπαράστασης δεδομένων. Αυτό είχε σαν αποτέλεσμα κάποιες πληροφορίες να χάνονται, και να μην υπολογίζονται στον τελικό σχηματισμό των οντολογιών.
- Για την προσέγγιση του υπο-Κεφαλαίου 4.1.4, η τελική προσέγγιση στηρίχτηκε στα τρέχοντα βήματα της λύσης, παραμετροποιώντας τους

αλγορίθμους της προσέγγισης. Η προσέγγιση ήταν αποδοτική σε πολλές περιπτώσεις, καλύπτοντας το κενό άλλων ερευνών μιας και μπορούσε να μετασχηματίσει σχεδόν όλα τα ήδη των δεδομένων σε οντολογίες και να τα αντιστοιχίσει σε HL7 FHIR μορφή. Ωστόσο, δημιουργήθηκαν κάποια κενά αναφορικά με τον τρόπο δημιουργίας των οντολογιών, αλλά και των αλγορίθμων που χρησιμοποιήθηκαν για την εύρεση τόσο της συντακτικής όσο και της σημασιολογικής ομοιότητας, αφού τα αποτελέσματα της προσέγγισης δε συμφωνούσαν πάντα με τα αναμενόμενα αποτελέσματα.

Βασικός στόχος της προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων, είναι να μετατρέπει ετερογενή ιατρικά δεδομένα σε HL7 FHIR μορφή, εντοπίζοντας κοινά στοιχεία και ομοιότητες μεταξύ διαφορετικών δεδομένων υγείας και πόρων HL7 FHIR. Συνοπτικά, η προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων εμπεριέχει αρχικά την εκτέλεση λειτουργιών οντολογικών μετασχηματισμών μέσω του επιπέδου Οντολογιών και Συσχετίσεων (Ontologies & Relationships) στα σύνολα δεδομένων υγείας που πρόκειται να μετασχηματιστούν, καθώς και στους αντίστοιχους πόρους HL7 FHIR. Έπειτα, οι παραγμένες οντολογίες καταχωρούνται σε μία Γνωσιακή Βάση Δεδομένων (Knowledge Base) σε τριαδική μορφή πληροφοριών, όπου επερωτήματα εφαρμόζονται για την πρόσβασή τους, καθώς και για την ανάκτηση των σωστών πληροφοριών συνταιριάσματος και μετατροπής. Κατόπιν, μέσω του Αλγορίθμου Συσχέτισης Δομής (Structure Mapping Library) σε αρχικό στάδιο, η σημασιολογία των δεδομένων υγείας εντοπίζεται μέσω του υπολογισμού της ομοιότητας των οντολογιών του συνόλου των δεδομένων υγείας σε σχέση με τους πόρους HL7 FHIR, σύμφωνα με τη συντακτική αλλά και τη σημασιολογική τους ομοιότητα. Σε δεύτερο στάδιο, διαδικασίες οντολογικής ευθυγράμμισης διεξάγονται για τα δεδομένα που δεν έχουν ταυτοποιηθεί σωστά μέσω της συντακτικής ή της σημασιολογικής τους ομοιότητας, ώστε να γνωστοποιηθεί ο συσχετισμός μεταξύ των δεδομένων που έχουν εισαχθεί, και των πόρων HL7 FHIR. Σε τελικό στάδιο, η Τεχνική Μετάφρασης Δομής FHIR (FHIR Structure Translator) λαμβάνει ως όρισμα για κάθε οντολογική ονομασία, τον πόρο HL7 FHIR με το οποίο έχουν εξαχθεί τα μέγιστα αποτελέσματα συντακτικής και σημασιολογικής ομοιότητας ή οντολογικής ευθυγράμμισης, μετασχηματίζοντάς το στη δομή του HL7 FHIR (δηλαδή Resource.attribute). Αξίζει να σημειωθεί πως τα εν λόγω στάδια εκτελούνται είτε σειριακά είτε παράλληλα, προκειμένου να ανακαλύψουν τη σημασιολογία και/ή τη φύση των στοιχείων του συνόλου των δεδομένων, ώστε να μεταφερθούν τα δεδομένα αυτά σε ένα μοναδικό και διεθνώς αναγνωρισμένο πρότυπο, για τη λήψη των βέλτιστων ιατρικών αποφάσεων, καθώς και τη δημιουργία αξιόλογων πληροφοριών. Προκειμένου να επιτευχθεί

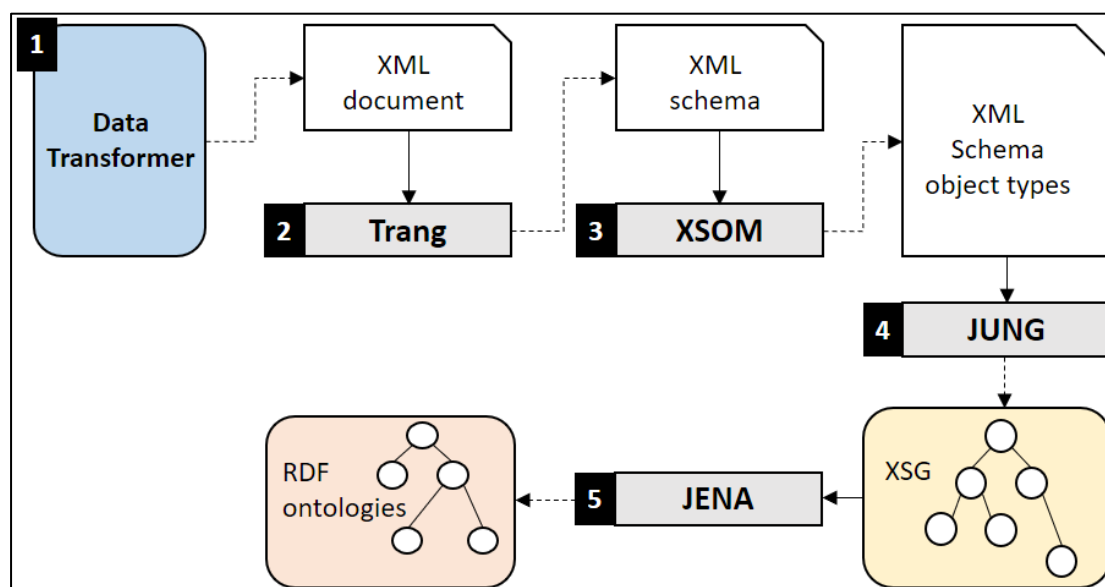
αυτό, η προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων αποτελείται από τα τέσσερα προαναφερθέντα διαφορετικά επίπεδα (Εικόνα 4-13), όπου το κάθε ένα είναι υπεύθυνο να παρέχει διαφορετικού τύπου υπηρεσίες. Τα επίπεδα αυτά είναι το επίπεδο των Οντολογιών και Συσχετίσεων (Ontologies & Relationships), η Γνωσιακή Βάση Δεδομένων (Knowledge Base), ο Αλγόριθμος Συσχέτισης Δομής (Structure Mapping Library), και η Τεχνική Μετάφρασης Δομής FHIR (FHIR Structure Translator).



Εικόνα 4-13 - Προσέγγιση υπηρεσιών διαλειτουργικότητας δεδομένων

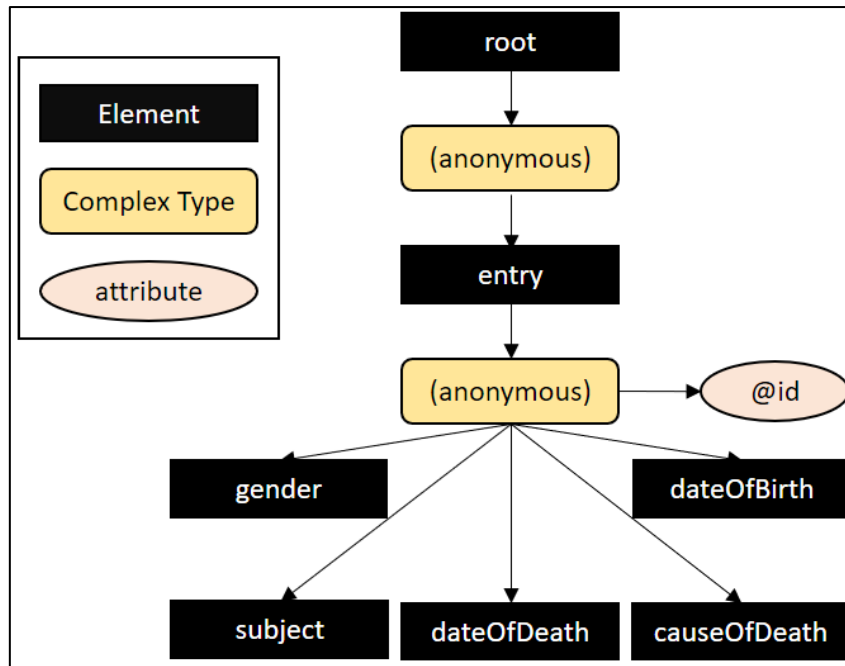
4.2.1 Επίπεδο Οντολογιών και Συσχετίσεων

Ο στόχος του επιπέδου Οντολογιών και Συσχετίσεων είναι να συγκεντρώσει το σύνολο των δεδομένων που πρέπει να μετατραπούν σε ένα XML Schema και στη συνέχεια να υλοποιηθεί η αντίστοιχη οντολογία. Στην περίπτωση που το XML Schema δεν υπάρχει ήδη, μπορεί να παραχθεί αυτομάτως από το πηγαίο XML αρχείο, αντιμετωπίζοντας όλες τις πολύπλοκες περιπτώσεις που μπορεί να υπάρχουν λόγω των διαφορετικών σχεδιαστικών προτύπων των XML Schemas. Η δημιουργία των οντολογιών από XML δεδομένα περιγράφεται στα ακόλουθα πέντε βήματα (Εικόνα 4-14):



Εικόνα 4-14 - Δημιουργία οντολογιών από XML δεδομένα

- Το αρχείο με τα δεδομένα υγείας διαβάζεται μέσω ενός Μετασχηματιστή Δεδομένων (Data Transformer) προκειμένου να αναγνωριστεί η μορφή με την οποία το αρχείο είναι αποθηκευμένο (π.χ. JSON, CSV, κτλ.). Σε κάθε περίπτωση, ο Data Transformer μετατρέπει τον τύπο του εν λόγω αρχείου σε XML, αναγνωρίζοντας τα διαφορετικά στοιχεία και κόμβους του αρχείου.
- Το XML αρχείο που δημιουργείται μετατρέπεται σε XML schema χρησιμοποιώντας το Trang API [262].
- Το XML Schema αναλύεται μέσω του XML Schema Object Model (XSOM) [263], το οποίο παρέχει ένα σύνολο κλάσεων που επιτρέπει την ανάγνωση ενός XML Schema και την εξαγωγή πληροφοριών, απαραίτητων για τη δημιουργία των οντολογιών (Εικόνα 4-15).

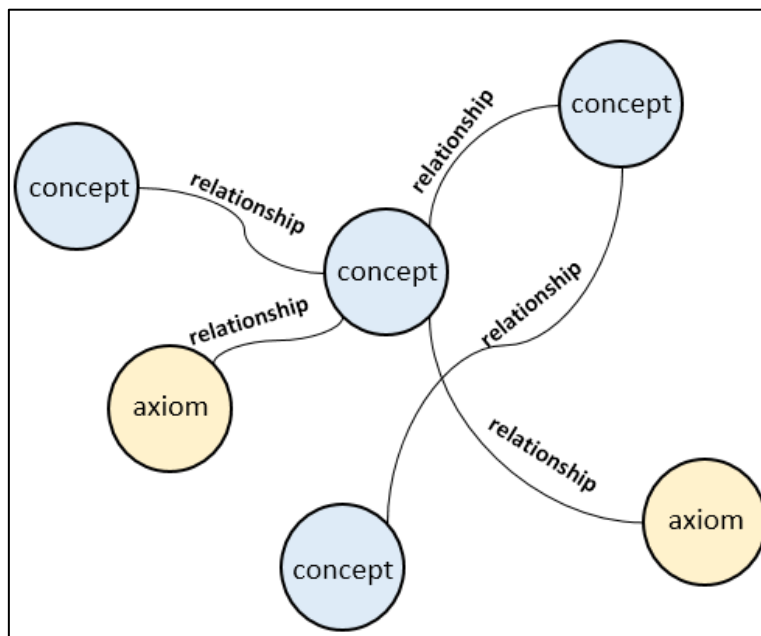


Εικόνα 4-15 - Παράδειγμα ανάλυσης XML Schema μέσω του XSOM

- Το εξαγόμενο αποτέλεσμα του XSOM χρησιμοποιείται σαν εισόδος στο Java Universal Network/Graph framework (JUNG) [264] το οποίο παράγει ένα XML Schema Graph (XSG) [265] που περιγράφει το Schema με τη χρήση γράφων.
- Το XSG χρησιμοποιείται σαν εισόδος στο Apache Jena, προκειμένου να παράγει τις RDF οντότητες και κλάσεις, μέσω των οποίων θα υλοποιηθούν οι απαιτούμενες οντολογίες των εισαχθέντων αρχείων.

4.2.2 Γνωσιακή Βάση Δεδομένων

Ο στόχος της Γνωσιακής Βάσης Δεδομένων (Knowledge Base - Ontology Triplestore) είναι να παρέχει μία βάση δεδομένων που βασίζεται στις συσχετίσεις, προκειμένου να αποθηκεύει τις οντολογίες και τις ιδιότητες που εντοπίστηκαν από τις υλοποιημένες οντολογίες (Εικόνα 4-16).



Εικόνα 4-16 - Δομή γνωσιακής βάσης δεδομένων

Πιο συγκεκριμένα, η RDF triplestore [151] χρησιμοποιείται, η οποία αποτελεί μία βάση δεδομένων για την καταχώρηση των αναγνωρισμένων (i) κλάσεων των οντολογιών (classes/concepts), (ii) συσχετίσεων και ιδιοτήτων (relationships), και (iii) στιγμιότυπων (instances/axioms), διευκολύνοντας τη δημιουργία ερωτημάτων μέσω των συλλεγμένων δεδομένων που μπορεί να εμπεριέχουν γνώση αναφορικά με μία ή παραπάνω καταχωρημένες πληροφορίες.

Γενικότερα, το RDF triplestore είναι ένας τύπος βάσης δεδομένων γραφημάτων που αποθηκεύει σημασιολογικά γεγονότα. Το RDF, είναι ένα μοντέλο για την έκδοση και ανταλλαγή δεδομένων στον παγκόσμιο ιστό τοποποιημένο από το W3C. Δεδομένου ότι είναι μια βάση δεδομένων με γραφήματα, τα triplestores αποθηκεύουν δεδομένα ως ένα δίκτυο αντικειμένων με υλοποιημένους δεσμούς μεταξύ τους. Αυτό καθιστά τα RDF triplestores την προτιμώμενη επιλογή για διαχείριση εξαιρετικά διασυνδεδεμένων δεδομένων. Τα triplestores είναι πιο ευέλικτα και λιγότερο δαπανηρά από μια σχεσιακή βάση δεδομένων. Η βάση δεδομένων RDF triplestores, συχνά αποκαλούμενη βάση δεδομένων σημασιολογικών γραφημάτων, είναι επίσης ικανή να χειρίζεται

ισχυρά σημασιολογικά ερωτήματα και να χρησιμοποιεί συμπεράσματα για την αποκάλυψη νέων πληροφοριών από τις υπάρχουσες σχέσεις. Σε αντίθεση με άλλους τύπους βάσεων δεδομένων γραφημάτων, οι μηχανισμοί RDF triplestores υποστηρίζουν προαιρετικά μοντέλα σχήματος, που ονομάζονται οντολογίες. Οι οντολογίες επιτρέπουν μια τυπική περιγραφή των δεδομένων και καθορίζουν και τις δύο κλάσεις αντικειμένων και ιδιότητες σχέσεων και την ιεραρχική σειρά τους.

Τα δεδομένα σε ένα RDF triplestore αποθηκεύονται σε τρία συνδεδεμένα κομμάτια δεδομένων, όπως παρακάτω:

RDF triple (υποκείμενο (*subject*) → ρήμα (*predicate*) → αντικείμενο (*object*))

Το αποτέλεσμα αυτό ονομάζεται τριάδα (*triple*), εξ' ου και το triplestore όνομα. Τα triples αναφέρονται επίσης ως «δηλώσεις» ή «δηλώσεις RDF».

Η μορφή **subject → predicate → object** μπορεί να λάβει οποιοδήποτε αντικείμενο και να συνδεθεί με οποιοδήποτε άλλο αντικείμενο χρησιμοποιώντας το predicate για να δείξει τον τύπο σχέσης που υπάρχει ανάμεσα στο subject και το object. Να σημειωθεί πως η βασική ιδέα της μορφής RDF triplestore είναι το αναγνωριστικό URI. Το URI είναι ένα ενιαίο παγκόσμιο σύστημα αναγνώρισης που χρησιμοποιείται στον Ιστό - ένα είδος μοναδικού αναγνωριστικού.

Οι βάσεις δεδομένων RDF triplestores χρησιμοποιούνται με επιτυχία για τη διαχείριση των Ανοικτών Δεδομένων (Open Data) [266], όπως τα DBPedia [267] και GeoNames [268], τα οποία δημοσιεύονται ως RDF και είναι διασυνδεδεμένα. Τα δυασυνδεδεμένα ανοικτά δεδομένα επιτρέπουν την ταχύτερη αναζήτηση και απάντηση σε ενοποιημένα ερωτήματα και την απόκτηση άκρως συναφών αποτελεσμάτων αναζήτησης. Τα RDF triplestores διερευνούν ποικίλα και εξελισσόμενα δεδομένα από διαφορετικές πηγές, τα οποία είναι πιο οικονομικά, αποδοτικά, και λιγότερο χρονοβόρα. Τα RDF triplestores διευκολύνουν επίσης πολλές τεχνικές ανάλυσης κειμένου, όπως η εξαγωγή πληροφοριών από μη δομημένα δεδομένα και τον εμπλουτισμό περιεχομένου. Με τον όρο «μάθηση» τόσο το νόημα όσο και το πλαίσιο μέσα στο οποίο χρησιμοποιούνται οι οντότητες, οι αλγόριθμοι μηχανικής μάθησης μπορούν να ταξινομήσουν οντότητες και να αποσαφηνιστούν μεταξύ τους. Εκτός από τον ορισμό των σχέσεων, τα RDF triplestores επιτρέπουν επίσης συνδέσεις μεταξύ βάσεων δεδομένων με δομημένα δεδομένα και έγγραφα που περιέχουν αδόμητο, ελεύθερο κείμενο. Με αυτό τον τρόπο, τα RDF triplestores

συνδέουν οντότητες από βάσεις δεδομένων με έγγραφα που αναφέρουν αυτές τις οντότητες.

Σε αυτό το επίπεδο της προσέγγισης, οι οντολογίες αποθηκεύονται σαν δικτυακά αντικείμενα με συσχετίσεις μεταξύ τους, ώστε να είναι πιο ευέλικτα και λιγότερο κοστοβόρα σε σχέση με μία σχεσιακή βάση δεδομένων. Επιπλέον, με το επίπεδο αυτό, καθίσταται ευκολότερο να εντοπιστούν μέσω πιθανοτήτων οι συσχετίσεις που λείπουν ή είναι εσφαλμένες, μέσω μίας Τεχνικής Εντοπισμού Συσχετίσεων (Relationship Matching). Συνοπτικά, η τεχνική αυτή είναι εκπαιδευμένη με τη χρήση Γραμμικής Παλινδρόμησης (Linear Regression) [269] σε πραγματικό χρόνο με τα δεδομένα και μεταδεδομένα που έχουν ήδη αποθηκευτεί στη Γνωσιακή Βάση Δεδομένων, με αποτέλεσμα να έχει αναπτύξει τη δυνατότητα να εντοπίζει τις προβληματικές συσχετίσεις. Έτσι, διαφορετικοί βαθμοί συσχέτισης παρέχονται μεταξύ των διαφορετικών συσχετίσεων και κλάσεων/στιγμιότυπων. Ανάλογα με τους βαθμούς αυτούς, η Τεχνική Εντοπισμού Συσχετίσεων ανιχνεύει και εντοπίζει εσφαλμένες αντιστοιχίες, ελέγχοντας επαναλαμβανόμενα όλη τη Γνωσιακή Βάση Δεδομένων, με αποτέλεσμα να αναθέτει εκ νέου τις συσχετίσεις που έχουν μεγαλύτερο βαθμό συσχέτισης σε άλλη κλάση/στιγμιότυπο, σε σχέση με αυτή που ήταν σε προηγούμενο στάδιο.

4.2.3 Αλγόριθμος Συσχέτισης Δομής

Ο στόχος του Αλγορίθμου Συσχέτισης Δομής είναι να παρέχει έναν τρόπο αντιστοίχισης και αναγνώρισης της Συντακτικής και Σημασιολογικής Ομοιότητας μεταξύ δύο διαφορετικών τύπων δεδομένων (attributes), βάσει της συντακτικής τους αλλά και της σημασιολογικής τους μορφής. Παράλληλα, στην περίπτωση που τα εξαγόμενα αποτελέσματα δεν είναι ικανά ώστε να γίνει σωστή αντιστοίχιση μεταξύ των δεδομένων υγείας και των πόρων HL7 FHIR, μία επιπλέον Τεχνική Οντολογικής Ευθυγράμμισης (Ontology Alignment) έχει υλοποιηθεί, με σκοπό την κάλυψη των περιπτώσεων που η Τεχνική Συντακτικής και Σημασιολογικής Ομοιότητας δεν αποδώσουν ικανά αποτελέσματα. Για το λόγο αυτό, ο Αλγόριθμος Συσχέτισης Δομής διαιρείται σε δύο υπο-Τεχνικές αναφορικά με τη (α) Συντακτική και τη Σημασιολογική Ομοιότητα, καθώς και την (β) Οντολογική Ευθυγράμμιση.

4.2.3.1 Τεχνικές Συντακτικής και Σημασιολογικής Ομοιότητας

4.2.3.1.1 Τεχνική Συντακτικής Ομοιότητας

Πρωτού αναφερθούμε στην τεχνική Συντακτικής Ομοιότητας, αξίζει να αναφερθεί η συγκριτική έρευνα που πραγματοποιήθηκε, ώστε να καταλήξουμε στο βέλτιστο αλγόριθμο εύρεσης συντακτικής ομοιότητας. Για το λόγο αυτό, πραγματοποιήθηκε η ακόλουθη σύγκριση μεταξύ της απόστασης Levenshtein, της ομοιότητας Cosine, της ομοιότητας Jaccard και της απόστασης Jaro-Winkler.

Στην περίπτωση αυτή, για την αξιολόγηση των τεσσάρων διαφορετικών μετρήσεων ομοιότητας συμβολοσειρών, θα τις εφαρμόσουμε στο ίδιο σενάριο χρήσης όπου θα υπολογίσουμε την ακρίβεια, την ανάκληση, το Fmeasure και τα συνολικά σφάλματα των αποτελεσμάτων τους, λαμβάνοντας υπόψη τους κανόνες ευθυγράμμισης που βασιζονται στο UMLS. Εν συντομία, αυτά τα κριτήρια σύγκρισης ορίζονται ως εξής:

(i) Η ακρίβεια των αποτελεσμάτων (σε εκατοστιαία ποσοστά): Το κλάσμα των δεδομένων που ανακτάται και σχετίζεται με το ερώτημα, απεικονίζοντας τον αριθμό των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό όλων των αποτελεσμάτων που επιστράφηκαν, όπως αναφέρθηκε προηγουμένως στο (9).

$$precision = \frac{|{\text{relevant data}} \cap {\text{retrieved data}}|}{|{\text{retrieved data}}|} \quad (9)$$

(ii) Ανάκληση αποτελεσμάτων (σε εκατοστιαία βάση): Το κλάσμα των σχετικών δεδομένων που ανακτάται επιτυχώς, απεικονίζοντας τον αριθμό των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό των αποτελεσμάτων που θα έπρεπε να έχουν παρασχεθεί, όπως αναφέρθηκε προηγουμένως στο (10).

$$recall = \frac{|\{relevant\ data\} \cap \{retrieved\ data\}|}{|\{relevant\ data\}|} \quad (10)$$

(iii) Το μέτρο F: Ο αρμονικός μέσος όρος (δηλαδή ο συνδυασμός) ακρίβειας και ανάκλησης, όπως στο (15).

$$F = \frac{2 * precision * recall}{precision + recall} \quad (13)$$

(iv) Σφάλματα (σε ποσοστό): Το ποσοστό που εμφανίζει το συνολικό ποσό των ελαττωματικών δεδομένων σε σύγκριση με το συνολικό ποσό των δεδομένων, όπως στο (16).

$$errors = \frac{|faulty\ data|}{|\{total\ data\}|} \quad (14)$$

Μετά από αυτόν τον υπολογισμό, θα γίνει μια σύγκριση μεταξύ των αποτελεσμάτων προκειμένου να συμπεράνει κανείς τον αλγόριθμο συντακτικής ομοιότητας που είναι πιο κατάλληλος για τον υπολογισμό της ομοιότητας των συμβολοσειρών στον τομέα της υγειονομικής περίθαλψης. Για το λόγο αυτό, πραγματοποιήθηκε η παρακάτω σύγκριση λαμβάνοντας υπόψιν την ακόλουθη περίπτωση χρήσης.

Η περίπτωση χρήσης εκμεταλλεύεται ένα σύνολο δεδομένων που έχει διαμορφωθεί με μορφή Comma Separated Values (CSV) (Εικόνα 4-17). Το χρησιμοποιηθέν σύνολο δεδομένων για την αξιολόγηση της αποτελεσματικότητας της προτεινόμενης προσέγγισης είναι ένα υποσύνολο προσωπικών πληροφοριών ανώνυμων πολιτών, που προέρχονται από το Ινστιτούτο Karolinska [270]. Αναλυτικότερα, αποτελείται από 5000 περιπτώσεις προσωπικών πληροφοριών, ορισμένων πολιτών σχετικά με: (i) το προσωπικό αναγνωριστικό (subject), (ii) το φύλο (gender), (iii) την ημερομηνία γέννησης (dateOfBirth), (iv) την ημερομηνία του θανάτου (dateOfDeath), και (v) την αιτία θανάτου (causeOfDeath).

1	subject;gender;dateOfBirth;dateOfDeath;causeOfDeath
2	ID0054619;1;1/6/1954;NA;NA
3	ID1012175;0;1/6/1968;NA;NA
4	ID0332172;0;1/11/1961;NA;NA
5	ID0186244;0;1/7/1963;NA;NA
6	ID0782623;1;1/5/1957;NA;NA
7	ID0944842;1;1/5/1961;NA;NA
8	ID0131218;1;1/7/1976;NA;NA
9	ID0131218;1;1/12/1951;NA;NA
10	ID0911095;1;1/11/1945;NA;NA
11	ID0951273;0;1/5/1928;20060212;C2299
12	ID0695153;1;1/4/1980;NA;NA
13	ID0695153;0;1/4/1977;NA;NA
14	ID0695153;1;1/4/1959;NA;NA
15	ID0695153;1;1/6/1951;NA;NA
16	ID1164801;0;1/1/1978;NA;NA
17	ID1409311;0;1/9/1960;NA;NA
18	ID0097871;0;1/8/1950;NA;NA
19	ID1095873;1;1/6/1924;NA;NA
20	ID0250381;0;1/10/1954;NA;NA

Εικόνα 4-17 – Δεδομένα περίπτωσης χρήσης για εύρεση βέλτιστου αλγορίθμου συντακτικής ομοιότητας

Μετά την εκτέλεση των βημάτων που αναλύονται στην Τεχνική Συντακτικής Ομοιότητας (της παρούσας προσέγγισης), προκύπτουν οι ακόλουθοι πίνακες (Πίνακας 4-1, Πίνακας 4-2,

Πίνακας 4-3, και Πίνακας 4-4). Σε αυτούς τους πίνακες, απεικονίζονται οι πόροι HL7 FHIR με υψηλότερους βαθμούς συντακτικής ομοιότητας, σε σύγκριση με τα ονόματα του συνόλων δεδομένων της περίπτωσης χρήσης, μέσω των τεσσάρων διαφορετικών αλγορίθμων μέτρησης συντακτικής ομοιότητας.

Πίνακας 4-1 – Βέλτιστα αποτελέσματα συντακτικής αντιστοίχισης για την απόσταση Levenshtein

Use Case Dataset Attribute	HL7 FHIR Resources	Levenshtein Distance
subject	<i>Patient.contact</i>	0.237 (~24%)
gender	<i>Patient.gender</i>	0.743 (~74%)
dateOfBirth	<i>Patient.birthDate</i>	0.881 (~88%)
dateOfDeath	<i>Patient.contact.relationship</i>	0.250 (25%)
causeOfDeath	<i>Patient.contact.gender</i>	0.278 (~28%)

Πίνακας 4-2 - Βέλτιστα αποτελέσματα συντακτικής αντιστοίχισης για την ομοιότητα Cosine

Use Case Dataset Attribute	HL7 FHIR Resources	Cosine Similarity
subject	<i>Patient.contact</i>	0.267 (~27%)
gender	<i>Patient.gender</i>	0.743 (~75%)
dateOfBirth	<i>Patient.multipleBirth</i>	0.591 (~59%)
dateOfDeath	<i>Patient.animal.breed</i>	0.250 (25%)
causeOfDeath	<i>Patient.contact.address</i>	0.128 (~13%)

Πίνακας 4-3 - Βέλτιστα αποτελέσματα συντακτικής αντιστοίχισης για την ομοιότητα Jaccard

Use Case Dataset Attribute	HL7 FHIR Resources	Jaccard Similarity
subject	<i>Patient.animal.species</i>	0.293 (~29%)
gender	<i>Patient.gender</i>	0.620 (~62%)
dateOfBirth	<i>Patient.multipleBirth</i>	0.241 (~24%)
dateOfDeath	<i>Patient.contact.address</i>	0.221 (~22%)
causeOfDeath	<i>Patient.multipleBirth</i>	0.218 (~22%)

Πίνακας 4-4 - Βέλτιστα αποτελέσματα συντακτικής αντιστοίχισης για την απόσταση Jaro-Winkler

Use Case Dataset Attribute	HL7 FHIR Resources	Jaro-Winkler Distance
subject	<i>Patient.photo</i>	0.462 (~46%)
gender	<i>Patient.gender</i>	0.543 (~54%)
dateOfBirth	<i>Patient.birthDate</i>	0.781 (~78%)
dateOfDeath	<i>Patient.gender</i>	0.186 (~19%)
causeOfDeath	<i>Patient.contact.gender</i>	0.198 (~20%)

Για την αξιολόγηση αυτών των αποτελεσμάτων σε σχέση με τις μετρήσεις της ακρίβειας, της ανάκλησης, της μέτρησης F και των σφαλμάτων, το συγκεκριμένο σύνολο δεδομένων μετασχηματίστηκε χειροκίνητα στο HL7 FHIR, έτσι ώστε να συγκριθούν τα αποτελέσματα της συντακτικής ομοιότητας με τα πραγματικά αποτελέσματα. Αυτός ήταν ο βασικός λόγος για τον οποίο επιλέχθηκε ένα μικρό δείγμα δεδομένων για την αξιολόγηση της προσέγγισης, προκειμένου να μετασχηματιστεί χειροκίνητα πιο εύκολα το προαναφερθέν αποτέλεσμα. Για τα αποτελέσματα που προέκυψαν χειροκίνητα, θεωρήθηκε ότι ήταν υψηλής ακρίβειας, τηρώντας τις κατευθυντήριες γραμμές της μορφής HL7 FHIR, που θεωρήθηκε ως σημείο αναφοράς υψηλής ποιότητας και ακρίβειας. Ο Πίνακας 4-5 απεικονίζει τα αποτελέσματα του χειροκίνητου μετασχηματισμού. Θα πρέπει να σημειωθεί ότι στην περίπτωση του χαρακτηριστικού *causeOfDeath*, τα αποτελέσματα του χειροκίνητου μετασχηματισμού δεν παρέχουν κανένα αντίστοιχο πόρο HL7 FHIR, καθώς στην τρέχουσα έκδοση του HL7 FHIR (v3.0.1) δεν υπάρχει κανένας πόρος που να περιγράφει αυτό το ειδικό χαρακτηριστικό ακόμη.

Πίνακας 4-5 – Αποτελέσματα χειροκίνητης συντακτικής αντιστοίχισης

Use Case Dataset Attribute	Manual Results	Similarity
subject	<i>Patient.identifier</i>	100%
gender	<i>Patient.gender</i>	100%
dateOfBirth	<i>Patient.birthDate</i>	100%
dateOfDeath	<i>Patient.deceased</i>	100%
causeOfDeath	NO MATCHING	NO MATCHING

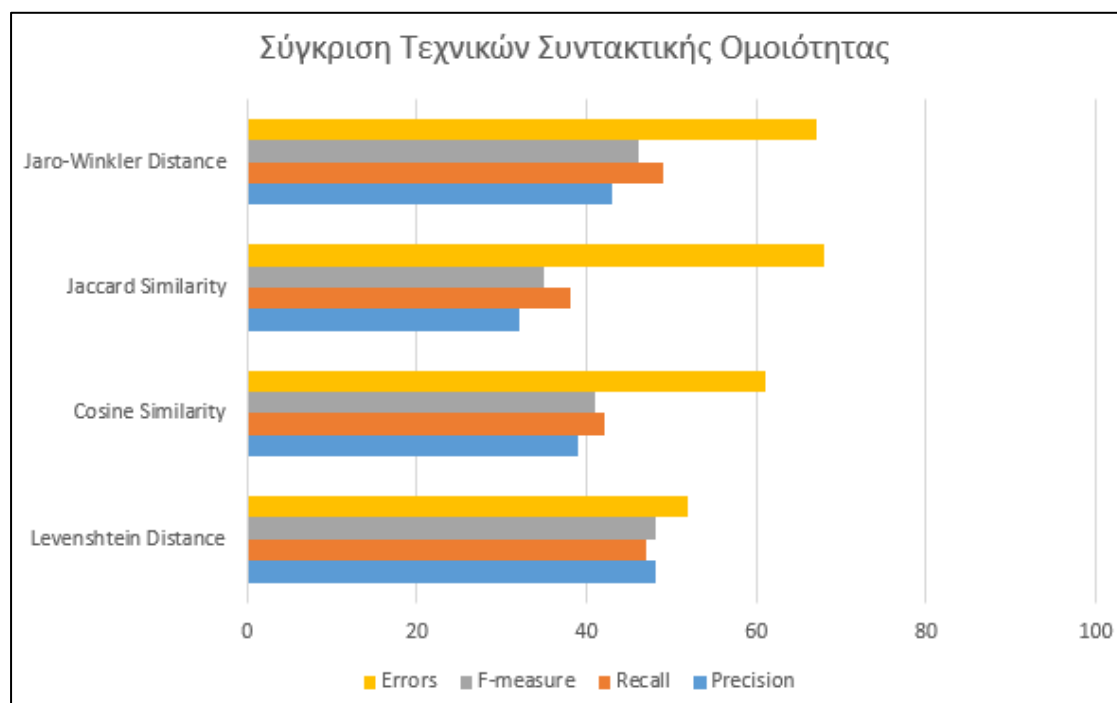
Βάσει των αποτελεσμάτων του Πίνακα 4-5, μπορούμε εύκολα να υπολογίσουμε τα διαφορετικά κριτήρια σύγκρισης για κάθε αλγόριθμο συντακτικής ομοιότητας.

Ο Πίνακας 4-6 μας παρέχει αυτά τα αποτελέσματα.

Πίνακας 4-6 - Κριτήρια σύγκρισης αλγορίθμων συντακτικής ομοιότητας

	Precision	Recall	F-measure	Errors
Manual Results	100%	100%	100%	0%
Levenshtein Distance	48%	47%	48%	52%
Cosine Similarity	39%	42%	41%	61%
Jaccard Similarity	32%	38%	35%	68%
Jaro-Winkler Distance	43%	49%	46%	67%

Στο Εικόνα 4-18 απεικονίζεται ένα ραβδόγραμμα των προαναφερθέντων αποτελεσμάτων, που δείχνει πώς συμπεριφέρονται οι τέσσερις διαφορετικοί αλγόριθμοι συντακτικής ομοιότητας συμβολοσειρών στην ολική εύρεση της συντακτικής ομοιότητας, αναφορικά με τα μη αυτόματα παρεχόμενα αποτελέσματα που δεν περιείχαν σφάλματα (0% σφάλματα), και η ακρίβειά τους, η ανάκληση και το μέτρο F είχαν τιμή ίση με το 100%.

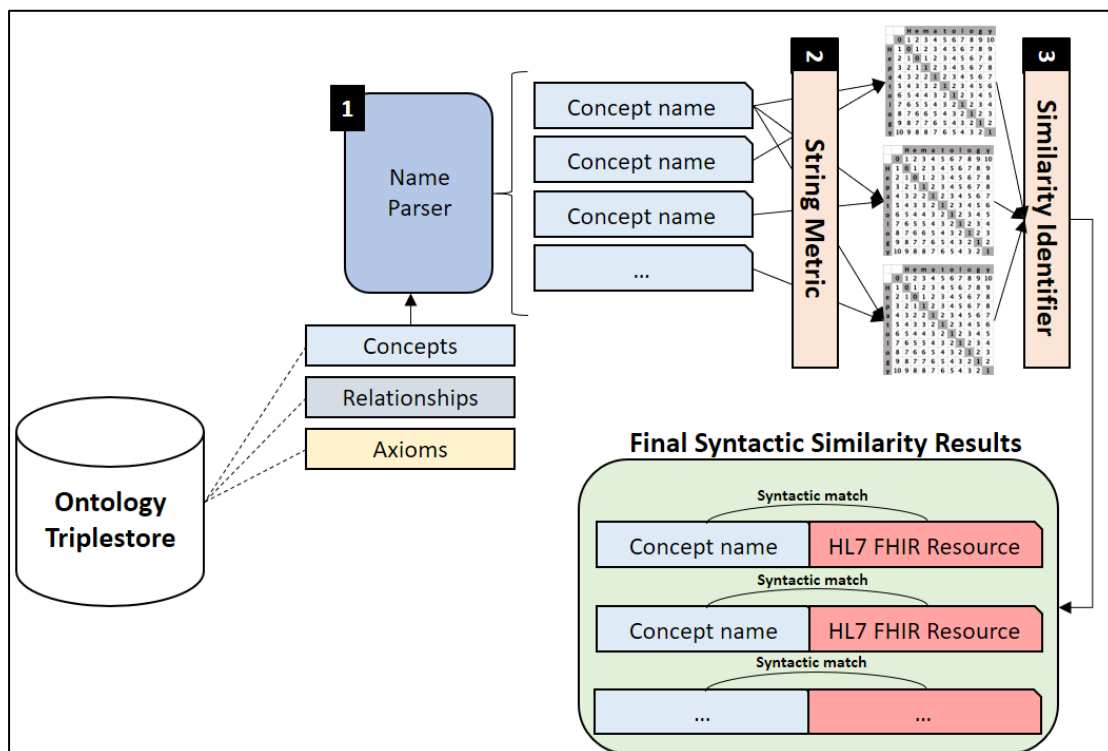


Εικόνα 4-18 - Συμπεριφορά αλγορίθμων συντακτικής ομοιότητας συμβολοσειρών

Είναι σαφές ότι δεν υπήρχε κανένας αλγόριθμος συντακτικής ομοιότητας συμβολοσειρών που να παρέχει 100% ακριβή αποτελέσματα. Είναι ξεκάθαρο

ότι με τον Πίνακα 4-6 και την Εικόνα 4-18, μεταξύ των διαφόρων μετρήσεων, η απόσταση Levenshtein συμπεριφέρεται καλύτερα, δεδομένου ότι παρέχει λιγότερα σφάλματα (52%) σε σχέση με τους άλλους αλγορίθμους συντακτικής ομοιότητας (61%, 68% και 67% αντίστοιχα). Επιπλέον, το μέτρο F (αποτέλεσμα τόσο της ακρίβειας όσο και της ανάκλησης) της απόστασης Levenshtein είχε τη μεγαλύτερη αξία (48%) σε σύγκριση με τις άλλες μετρήσεις ομοιότητας σειράς (41%, 35% και 46% αντίστοιχα). Συνεπώς, αυτό οδηγεί στο συμπέρασμα ότι για την Τεχνική Συντακτικής Ομοιότητας, ο πιο κατάλληλος αλγόριθμος συντακτικής ομοιότητας συμβολοσειρών που χρησιμοποιείται είναι η απόσταση Levenshtein, δεδομένου ότι παρέχει πιο αποτελεσματικά και αξιόπιστα αποτελέσματα. Για το λόγο αυτό, η απόσταση Levenshtein θα χρησιμοποιηθεί ολοκληρωτικά στην Τεχνική Συντακτικής Ομοιότητας.

Στόχος της Τεχνικής Συντακτικής Ομοιότητας (Syntactic Similarity Identifier) είναι να αναγνωρίσει το μέτρο της συντακτικής ομοιότητας μεταξύ των διαφορετικών ονομασιών των οντολογιών, και να παρέχει την πιθανότητα ώστε μία συγκεκριμένη οντολογία να είναι η ίδια -συντακτικά - με μία άλλη οντολογία. Στην περίπτωση μας, προκειμένου να υπολογίσουμε τη συντακτική ομοιότητα μεταξύ δύο διαφορετικών ονομασιών των οντολογιών, τα ακόλουθα τρία βήματα ακολουθούνται (Εικόνα 4-19):



Εικόνα 4-19 – Βήματα τεχνικής συντακτικής ομοιότητας

- Οι διαφορετικές οντολογίες που αποθηκεύονται στη Γνωσιακή Βάση Δεδομένων διαβάζονται μέσω του Σαρωτή Ονομάτων (Name Parser) ο οποίος είναι υπεύθυνος να λαμβάνει τις ονομασίες των διαφορετικών οντολογιών που εμπεριέχονται στις οντολογίες.
- Οι ονομασίες των διαφορετικών οντολογιών συγκρίνονται μία-προς-μία, προκειμένου να υπολογιστεί η συντακτική τους ομοιότητα. Για το λόγο αυτό, η απόσταση Levenshtein (Levenshtein Distance) μετριέται, αναφερόμενη στο πλήθος των διαγραφών, εισαγωγών ή αντικαταστάσεων που απαιτούνται για τη μετατροπή του ονόματος μιας οντολογίας, ώστε να ταυτιστεί με το όνομα της οντολογίας με την οποία συγκρίνεται.
- Από τη στιγμή που όλοι οι διαφορετικοί συνδυασμοί έχουν υπολογιστεί, δημιουργούνται πίνακες συντακτικής ομοιότητας μεταξύ των διαφορετικών οντολογιών. Στόχος του συγκεκριμένου βήματος της προσέγγισης είναι να εντοπίσει τις ονομασίες των οντολογιών των δεδομένων υγείας, οι οποίες έχουν μεγαλύτερο βαθμό ομοιότητας με τους πόρους HL7 FHIR. Συνεπώς, μεταξύ των διαφορετικών βαθμών ομοιότητας, στο συγκεκριμένο βήμα τελικά αποθηκεύονται και αντιστοιχίζονται συντακτικά οι ονομασίες των οντολογιών και των πόρων HL7 FHIR, οι οποίοι ταυτίζονται περισσότερο συντακτικά.

4.2.3.1.2 Τεχνική Σημασιολογικής Ομοιότητας

Πρωτού αναφερθούμε στην τεχνική Σημασιολογικής Ομοιότητας, αξίζει να αναφερθεί η συγκριτική έρευνα που πραγματοποιήθηκε, ώστε να καταλήξουμε στη βέλτιστη τεχνική εύρεσης σημασιολογικής ομοιότητας. Για το λόγο αυτό, πραγματοποιήθηκε η ακόλουθη σύγκριση μεταξύ του Word2Vec, του Wordnet, και του Cortical.io. Στην περίπτωση αυτή, για την αξιολόγηση των τριών διαφορετικών τεχνικών μέτρησης σημασιολογικής ομοιότητας συμβολοσειρών, θα τις εφαρμόσουμε στο ίδιο σενάριο χρήσης όπου θα υπολογίσουμε την ακρίβεια, την ανάκληση, το μέτρο F και τα συνολικά σφάλματα των αποτελεσμάτων τους. Εν συντομία, αυτά τα κριτήρια σύγκρισης ορίζονται ως εξής:

(i) Η ακρίβεια των αποτελεσμάτων (σε εκατοστιαία ποσοστά): Το κλάσμα των δεδομένων που ανακτάται και σχετίζεται με το ερώτημα, απεικονίζοντας τον αριθμό των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό όλων των αποτελεσμάτων που επιστράφηκαν, όπως αναφέρθηκε προηγουμένως στο (9).

$$precision = \frac{|{\textit{relevant data}} \cap {\textit{retrieved data}}|}{|{\textit{retrieved data}}|} \quad (9)$$

(ii) Ανάκληση αποτελεσμάτων (σε εκατοστιαία βάση): Το κλάσμα των σχετικών δεδομένων που ανακτάται επιτυχώς, απεικονίζοντας τον αριθμό των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό των αποτελεσμάτων που θα έπρεπε να έχουν παρασχεθεί, όπως αναφέρθηκε προηγουμένως στο (10).

$$recall = \frac{|{\textit{relevant data}} \cap {\textit{retrieved data}}|}{|{\textit{relevant data}}|} \quad (10)$$

(iii) Το μέτρο F: Ο αρμονικός μέσος όρος (δηλαδή ο συνδυασμός) ακρίβειας και ανάκλησης, όπως αναφέρθηκε προηγουμένως στο (13).

$$F = \frac{2 * precision * recall}{precision + recall} \quad (13)$$

(iv) Σφάλματα (σε ποσοστό): Το ποσοστό που εμφανίζει το συνολικό ποσό των ελαττωματικών δεδομένων σε σύγκριση με το συνολικό ποσό των δεδομένων, όπως αναφέρθηκε προηγουμένως στο (14)

$$errors = \frac{|{\textit{faulty data}}|}{|{\textit{total data}}|} \quad (14)$$

Μετά από αυτόν τον υπολογισμό, θα πραγματοποιηθεί μια σύγκριση μεταξύ των αποτελεσμάτων προκειμένου να συμπεράνει κανείς την τεχνική σημασιολογικής ομοιότητας που είναι πιο κατάλληλη για τον υπολογισμό της σημασιολογικής ομοιότητας των συμβολοσειρών στον τομέα της υγειονομικής περίθαλψης. Για το λόγο αυτό, πραγματοποιήθηκε η παρακάτω σύγκριση λαμβάνοντας υπόψιν την ακόλουθη περίπτωση χρήσης.

Η περίπτωση χρήσης εκμεταλλεύεται το ίδιο σύνολο δεδομένων που χρησιμοποιήθηκε για την Τεχνική Συντακτικής Ομοιότητας, σε μορφή CSV (Εικόνα 4-20). Το χρησιμοποιηθέν σύνολο δεδομένων για την αξιολόγηση της αποτελεσματικότητας της προτεινόμενης προσέγγισης είναι ένα υποσύνολο προσωπικών πληροφοριών ανώνυμων πολιτών, που προέρχονται από το Ινστιτούτο Karolinska. Αναλυτικότερα, αποτελείται από 5000 περιπτώσεις προσωπικών πληροφοριών, ορισμένων πολιτών σχετικά με: (i) το προσωπικό αναγνωριστικό (subject), (ii) το φύλο (gender), (iii) την ημερομηνία γέννησης (dateOfBirth), (iv) την ημερομηνία του θανάτου (dateOfDeath), και (v) την αιτία θανάτου (causeOfDeath).

	subject;gender;dateOfBirth;dateOfDeath;causeOfDeath
1	ID0054619;1;1/6/1954;NA;NA
2	ID1012175;0;1/6/1968;NA;NA
3	ID0332172;0;1/11/1961;NA;NA
4	ID0186244;0;1/7/1963;NA;NA
5	ID0782623;1;1/5/1957;NA;NA
6	ID0944842;1;1/5/1961;NA;NA
7	ID0131218;1;1/7/1976;NA;NA
8	ID0131218;1;1/12/1951;NA;NA
9	ID0911095;1;1/11/1945;NA;NA
10	ID0951273;0;1/5/1928;20060212;C2299
11	ID0695153;1;1/4/1980;NA;NA
12	ID0695153;0;1/4/1977;NA;NA
13	ID0695153;1;1/4/1959;NA;NA
14	ID0695153;1;1/6/1951;NA;NA
15	ID1164801;0;1/1/1978;NA;NA
16	ID1409311;0;1/9/1960;NA;NA
17	ID0097871;0;1/8/1950;NA;NA
18	ID1095873;1;1/6/1924;NA;NA
19	ID0250381;0;1/10/1954;NA;NA
20	

Εικόνα 4-20 - Δεδομένα περίπτωσης χρήσης για εύρεση βέλτιστου αλγορίθμου σημασιολογικής ομοιότητας

Μετά την εκτέλεση των βημάτων που εισάγονται στην Τεχνική Σημασιολογικής Ομοιότητας (της παρούσας προσέγγισης), δημιουργούνται οι ακόλουθοι πίνακες (Πίνακας 4-7, Πίνακας 4-8, και Πίνακας 4-9). Σε αυτούς τους πίνακες, απεικονίζονται οι πόροι HL7 FHIR με υψηλότερους βαθμούς σημασιολογικής ομοιότητας, σε σύγκριση με τα σύνολα δεδομένων της περίπτωσης χρήσης, μέσω των τριών διαφορετικών τεχνικών μέτρησης σημασιολογικής ομοιότητας.

Πίνακας 4-7 - Βέλτιστα αποτελέσματα σημασιολογικής αντιστοίχισης για το Cortical.io

Use Case Dataset Attribute	HL7 FHIR Resources	Cortical.io
subject	<i>Patient.identifier</i>	0.947 (~95%)
gender	<i>Patient.gender</i>	0.863 (~86%)
dateOfBirth	<i>Patient.birthDate</i>	0.971 (~97%)
dateOfDeath	<i>Patient.deceased</i>	0.860 (86%)
causeOfDeath	<i>Patient.deceased</i>	0.458 (~46%)

Πίνακας 4-8 - Βέλτιστα αποτελέσματα σημασιολογικής αντιστοίχισης για το Word2Vec

Use Case Dataset Attribute	HL7 FHIR Resources	Word2Vec
subject	<i>Patient.contact.address</i>	0.357 (~36%)
gender	<i>Patient.gender</i>	0.663 (~66%)
dateOfBirth	<i>Patient.birthDate</i>	0.891 (~89%)
dateOfDeath	<i>Patient.deceased</i>	0.452 (~45%)
causeOfDeath	-	0 (0%)

Πίνακας 4-9 - Βέλτιστα αποτελέσματα σημασιολογικής αντιστοίχισης για το Wordnet

Use Case Dataset Attribute	HL7 FHIR Resources	Wordnet
subject	<i>Patient.gender</i>	0.772 (~77%)
gender	<i>Patient.gender</i>	0.861 (~86%)
dateOfBirth	<i>Patient.multipleBirth</i>	0.561 (~56%)
dateOfDeath	<i>Patient.deceased</i>	0.721 (~72%)
causeOfDeath	<i>Patient.deceased</i>	0.619 (~62%)

Για την αξιολόγηση αυτών των αποτελεσμάτων σε σχέση με τις μετρήσεις της ακρίβειας, της ανάκλησης, της μέτρησης F και των σφαλμάτων, το συγκεκριμένο σύνολο δεδομένων μετασχηματίστηκε χειροκίνητα στο HL7 FHIR, έτσι ώστε να συγκριθούν τα αποτελέσματα της αναπτυγμένης προσέγγισης με τα πραγματικά αποτελέσματα. Αυτός ήταν ο βασικός λόγος για τον οποίο επιλέχθηκε ένα μικρό δείγμα δεδομένων για την αξιολόγηση της προσέγγισης, προκειμένου να δημιουργηθεί χειροκίνητα πιο εύκολα το προαναφερθέν αποτέλεσμα. Για τα αποτελέσματα που προέκυψαν χειροκίνητα, θεωρήθηκε ότι ήταν υψηλής ακρίβειας, τηρώντας τις κατευθυντήριες γραμμές της μορφής HL7 FHIR, που θεωρήθηκε ως σημείο αναφοράς υψηλής ποιότητας και ακρίβειας. Ο Πίνακας 4-10 απεικονίζει τα αποτελέσματα του χειροκίνητου μετασχηματισμού. Θα πρέπει να σημειωθεί ότι στην περίπτωση του χαρακτηριστικού `causeOfDeath`, τα αποτελέσματα του χειροκίνητου μετασχηματισμού δεν παρέχουν κανένα αντίστοιχο πόρο HL7 FHIR, καθώς στην τρέχουσα έκδοση του HL7 FHIR (v3.0.1) δεν υπάρχει κανένας πόρος που να περιγράφει αυτό το ειδικό χαρακτηριστικό ακόμη.

Πίνακας 4-10 - Αποτελέσματα χειροκίνητης σημασιολογικής αντιστοίχισης

Use Case Dataset Attribute	Manual Results	Similarity
subject	<i>Patient.identifier</i>	100%
gender	<i>Patient.gender</i>	100%
dateOfBirth	<i>Patient.birthDate</i>	100%
dateOfDeath	<i>Patient.deceased</i>	100%
causeOfDeath	NO MATCHING	NO MATCHING

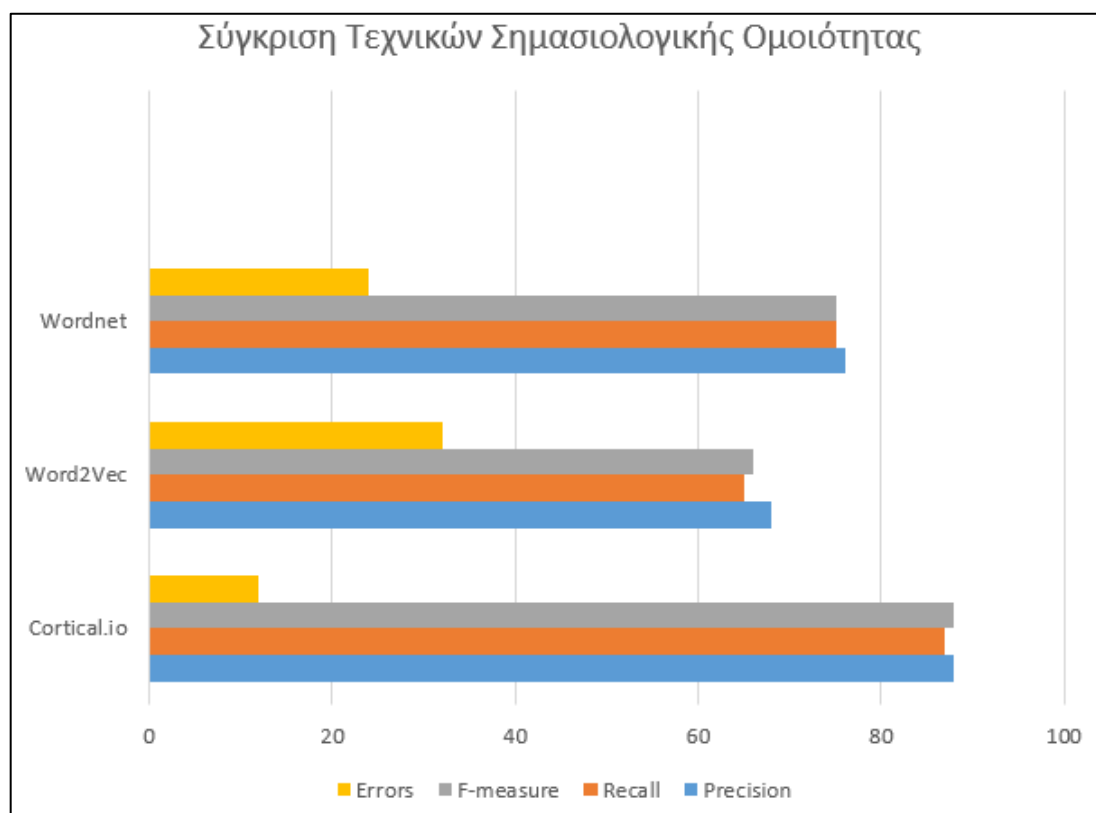
Βάσει των αποτελεσμάτων του Πίνακας 4-10, μπορούμε εύκολα να υπολογίσουμε τα διαφορετικά κριτήρια σύγκρισης για κάθε τεχνική σημασιολογικής ομοιότητας.

Ο Πίνακας 4-11 μας παρέχει αυτά τα αποτελέσματα.

Πίνακας 4-11 - Κριτήρια σύγκρισης αλγορίθμων σημασιολογικής ομοιότητας

	Precision	Recall	F-measure	Errors
Manual Results	100%	100%	100%	0%
Cortical.io	88%	87%	88%	12%
Word2Vec	68%	65%	66%	32%
Wordnet	76%	75%	75%	24%

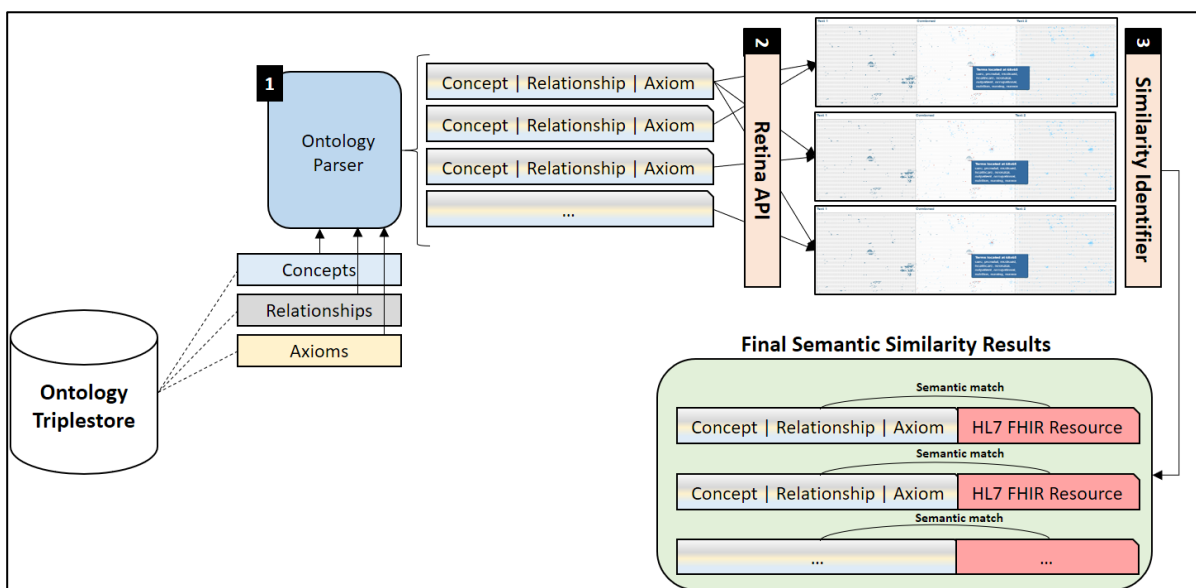
Στην Εικόνα 4-21 απεικονίζεται ένα διάγραμμα ράβδων των προαναφερθέντων αποτελεσμάτων, που δείχνει πώς συμπεριφέρονται οι τρεις διαφορετικές τεχνικές σημασιολογικής ομοιότητας συμβολοσειρών στην ολική εύρεση της σημασιολογικής ομοιότητας, αναφορικά με τα μη αυτόματα παρεχόμενα αποτελέσματα που δεν περιείχαν σφάλματα (0% σφάλματα), και η ακρίβειά τους, η ανάκληση και το μέτρο F είχαν τιμή ίση με το 100%.



Εικόνα 4-21 - Συμπεριφορά αλγορίθμων συντακτικής ομοιότητας συμβολοσειρών

Είναι σαφές ότι δεν υπήρχε καμία τεχνική σημασιολογικής ομοιότητας συμβολοσειρών που να παρέχει 100% ακριβή αποτελέσματα. Είναι ξεκάθαρο ότι με τον Πίνακα 4-11 και την Εικόνα 4-21, μεταξύ των διαφόρων μετρήσεων, το Cortical.io συμπεριφέρεται καλύτερα, δεδομένου ότι παρέχει λιγότερα σφάλματα (12%) σε σχέση με τις άλλες τεχνικές σημασιολογικής ομοιότητας (32%, και 24% αντίστοιχα). Επιπλέον, το μέτρο F (αποτέλεσμα τόσο της ακρίβειας όσο και της ανάκλησης) του Cortical.io είχε τη μεγαλύτερη αξία (88%) σε σύγκριση με τις άλλες τεχνικές (66%, και 75% αντίστοιχα). Συνεπώς, αυτό οδηγεί στο συμπέρασμα ότι για την Τεχνική Σημασιολογικής Ομοιότητας, η πιο κατάλληλη τεχνική σημασιολογικής ομοιότητας συμβολοσειρών που χρησιμοποιείται είναι το Cortical.io, δεδομένου ότι παρέχει πιο αποτελεσματικά και αξιόπιστα αποτελέσματα. Για τον λόγο αυτό η τεχνική Cortical.io θα χρησιμοποιηθεί ολοκληρωτικά στην Τεχνική Σημασιολογικής Ομοιότητας.

Στόχος της Τεχνικής Σημασιολογικής Ομοιότητας (Semantic Similarity Identifier) είναι να παρέχει τον τρόπο αντιστοίχισης των διαφορετικών οντολογιών, βάσει της σημασιολογικής τους ερμηνείας. Η συγκεκριμένη διαδικασία εμπεριέχει ενέργειες αντιστοίχισης, βάσει της ομοιότητας μεταξύ των αποθηκευμένων οντολογιών, καθώς και ενέργειες ανάλυσης των αποτελεσμάτων με στόχο την ανεύρεση μιας ολικής σημασιολογικής ομοιότητας. Στην περίπτωση αυτή, οι σημασιολογικές αντιστοιχίσεις βασίζονται στις συσχετίσεις και τις εξαρτίσεις μεταξύ των διαφορετικών οντολογιών. Στην περίπτωσή μας, προκειμένου να υπολογίσουμε τη σημασιολογική ομοιότητα μεταξύ δύο οντολογιών, τα ακόλουθα τρία βήματα ακολουθούνται (Εικόνα 4-22):

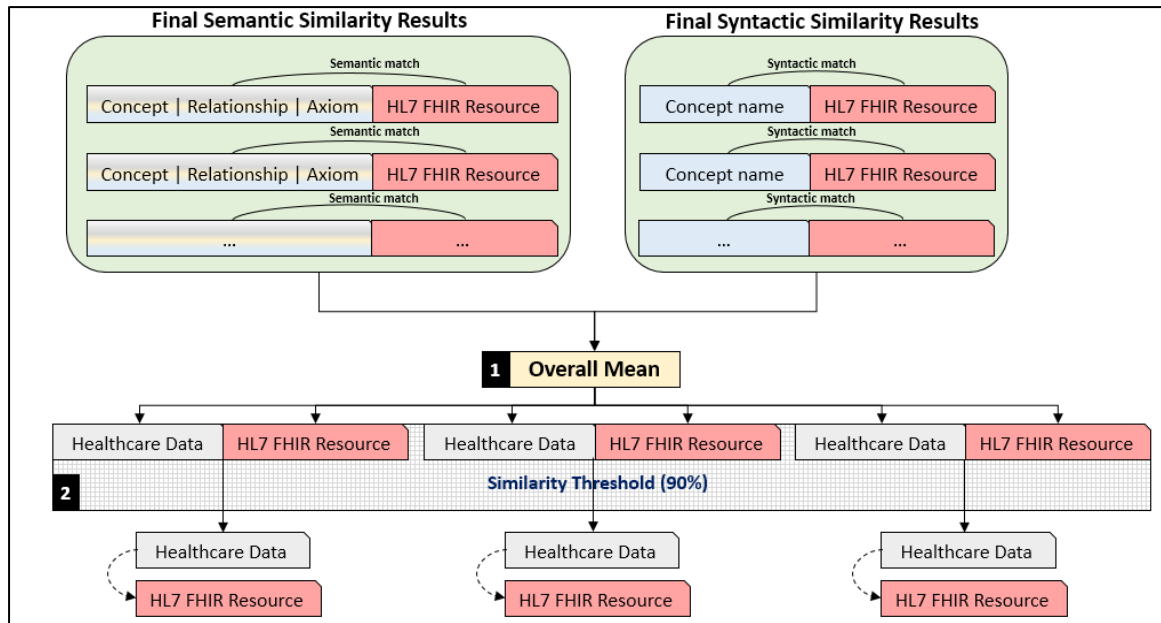


Εικόνα 4-22 - Βήματα τεχνικής σημασιολογικής ομοιότητας

- Οι διαφορετικές οντολογίες που αποθηκεύονται στη Γνωσιακή Βάση Δεδομένων διαβάζονται μέσω του Σαρωτή Οντολογιών (Ontology Parser), όπου είναι υπεύθυνος να λαμβάνει τις διαφορετικές ονομασίες των οντολογιών (κλάσεις), μαζί με τις συσχετίσεις (ιδιότητες) και τα στιγμιότυπα.
- Αυτή η τριάδα πληροφοριών των διαφορετικών οντολογιών συγκρίνονται μία-προς-μία, προκειμένου να υπολογιστεί η σημασιολογική τους ομοιότητα. Για το λόγο αυτό, το Cortical.io χρησιμοποιείται, το οποίο βασίζεται στη χρήση κατανεμημένης μνήμης. Ειδικότερα, μέσω του εν λόγω API οι λέξεις δεν αποθηκεύονται σαν μία απλή μονάδα, αλλά κάθε λέξη αποτελείται από συγκεκριμένα μοτίβα νευρώνων, με αποτέλεσμα να αναπαρίστανται στο χώρο, σαν σημασιολογικά δακτυλικά αποτυπώματα. Κατόπιν, το γενικό Αγγλικό Λεξικό Retina (English Retina Database) χρησιμοποιείται, και μετατρέπει τις τριάδες πληροφοριών σε σημασιολογικά δακτυλικά αποτυπώματα, προκειμένου να συγκριθεί η σημασία τους.
- Από τη στιγμή που όλοι οι διαφορετικοί συνδυασμοί έχουν υπολογιστεί, δημιουργούνται πίνακες σημασιολογικής ομοιότητας μεταξύ των διαφορετικών οντολογιών. Στόχος του συγκεκριμένου βήματος της Τεχνικής είναι να εντοπίσει τις τριάδες πληροφοριών των δεδομένων υγείας, οι οποίες έχουν μεγαλύτερο βαθμό ομοιότητας με τους πόρους HL7 FHIR. Συνεπώς, μεταξύ των διαφορετικών βαθμών ομοιότητας, στο συγκεκριμένο βήμα τελικά αποθηκεύονται και αντιστοιχίζονται σημασιολογικά οι τριάδες των πληροφοριών και των πόρων HL7 FHIR, οι οποίοι ταυτίζονται περισσότερο σημασιολογικά.

4.2.3.1.3 Τεχνική Ολικής Οντολογικής Αντιστοίχισης

Στόχος της Τεχνικής Ολικής Οντολογικής Αντιστοίχισης (Overall Ontology Mapper) είναι να παρέχει έναν τρόπο ώστε να προσθέτει και να συγχωνεύει τα αποτελέσματα που έχουν εξαχθεί από τις προηγούμενες Τεχνικές Συντακτικής και Σημασιολογικής Ομοιότητας, ώστε να καταλήξει στην τελική αντιστοίχιση των δεδομένων υγείας με τους αντίστοιχους πόρους HL7 FHIR. Στην περίπτωση μας, προκειμένου να υπολογίσουμε το τελικό αποτέλεσμα αντιστοίχισης, τα ακόλουθα δύο βήματα ακολουθούνται (Εικόνα 4-23):

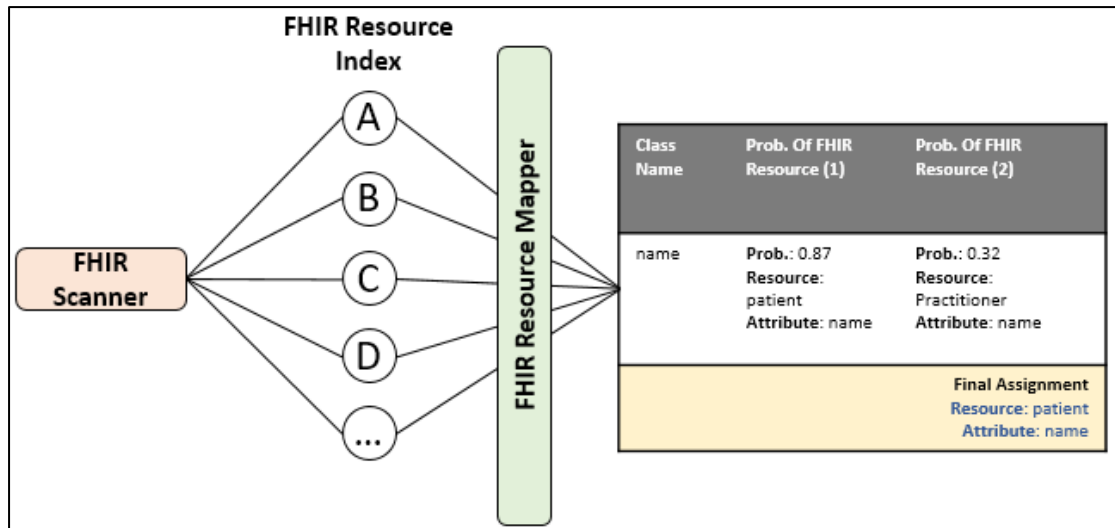


Εικόνα 4-23 - Βήματα τεχνικής ολικής οντολογικής αντιστοίχισης

- Η Τεχνική Ολικής Οντολογικής Αντιστοίχισης πραγματοποιεί επερωτήματα μέσω των τιμών ομοιότητας που υπολογίστηκαν για κάθε διαφορετική οντολογία, και παρέχει το μέσο όρο μεταξύ της συντακτικής και της σημασιολογικής ομοιότητας, κατανέμοντας την αντίστοιχη βαρύτητα του 50% στα αποτελέσματα της συντακτικής ομοιότητας, και του 50% στα αποτελέσματα της σημασιολογικής ομοιότητας. Αυτός ο μέσος όρος (Overall Mean) υπολογίζεται με τον ακόλουθο τύπο (15):

$$\text{Overall Mean} = \frac{\text{Syntactic Simil} + \text{Semantic Simil}}{2} \quad (15)$$

- Εφόσον το αποτέλεσμα της ομοιότητας υπολογιστεί μεγαλύτερο του 90%, το συγκεκριμένο επίπεδο εισάγει μία επιπλέον Τεχνική, τον Σαρωτή FHIR (FHIR Scanner) ο οποίος σαρώνει μεταξύ των διαφορετικών πόρων HL7 FHIR (Εικόνα 4-24). Η τεχνική αυτή συσχετίζει πιθανολογικά τη σημασιολογική ερμηνεία των αποθηκευμένων κλάσεων με ένα συγκεκριμένο πόρο HL7 FHIR, σε συνδυασμό με την τρέχουσα τιμή του. Στην περίπτωση αυτή, διαφορετικές πιθανότητες ομοιότητας με τους διαφορετικούς πόρους HL7 FHIR συνδέονται με κάθε διαφορετική κλάση, και ακολουθώντας τα στάδια που περιγράφονται από την Τεχνική Μετάφρασης Δομής FHIR (Κεφάλαιο 4.2.4), ο πόρος HL7 FHIR με την υψηλότερη πιθανότητα ομοιότητας, ανατίθεται στη συγκεκριμένη κλάση.



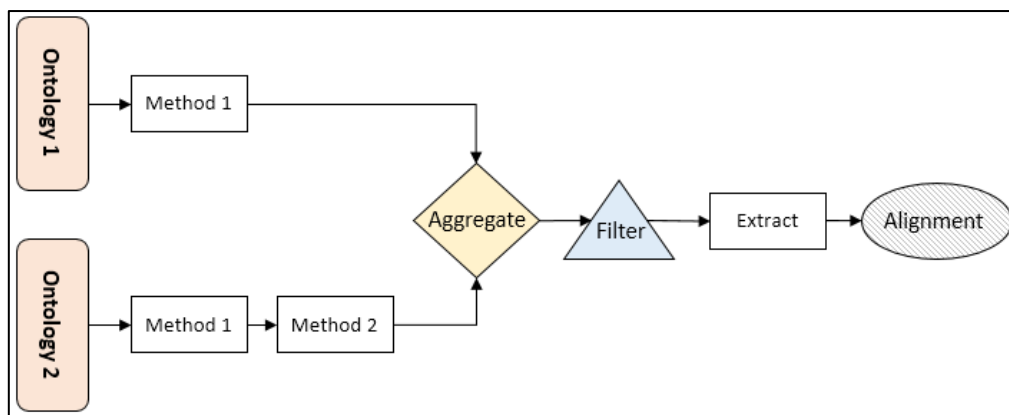
Εικόνα 4-24 – Τεχνική σαρωτή FHIR

Στην περίπτωση που ο μέσος όρος (*Overall Mean*) δε ξεπεράσει το 90%, η τεχνική έχει σαν αποτέλεσμα να θεωρεί αναξιόπιστα τα τελικά αποτελέσματα – κατόπιν σχετικού επερωτήματος προς τον χρήστη, και έτσι να τα αγνοεί. Συνεπώς, στην περίπτωση αυτή εκτελούνται τα βήματα που περιγράφονται στο πλαίσιο της Οντολογικής Ευθυγράμμισης, προκειμένου να πραγματοποιηθεί η αντιστοιχη μετάφραση των δεδομένων υγείας σε HL7 FHIR.

4.2.3.2 Τεχνική Οντολογικής Ευθυγράμμισης

Τα βήματα της Οντολογικής Ευθυγράμμισης (*Ontology Alignment*) έχουν στόχο να βοηθήσουν στην κατανόηση και διερμηνεία της σημασιολογίας των διαφορετικών οντολογικών ονομασιών που υπάρχουν στη Γνωσιακή Βάση Δεδομένων, για τα οποία η Τεχνική Συντακτικής & Σημασιολογικής Ομοιότητας δεν έδωσε έμπιστα και αξιόπιστα αποτελέσματα. Συνοπτικά, η Οντολογική Ευθυγράμμιση αποτελεί τη διαδικασία εύρεσης κοινών συσχετίσεων μεταξύ δύο διαφορετικών οντολογιών. Η εν λόγω διαδικασία περιλαμβάνει την εκτέλεση αλγορίθμων συνταιριάσματος παράλληλα με το *Alignment API* [271], σύμφωνα με την ομοιότητα ήδη καταχωρημένων οντολογιών, και στη συνέχεια την επεξεργασία των αποτελεσμάτων προκειμένου να βρεθεί η πλήρης αντιστοίχιση. Από τη στιγμή που δύο οντολογίες ευθυγραμμιστούν, μία απ' ευθείας σύνδεση μπορεί να δημιουργηθεί μεταξύ των δεδομένων που βρίσκονται στο σύνολο των δεδομένων υγείας, και των αντιστοιχών πόρων HL7 FHIR. Στην περίπτωση αυτή, οι συσχετίσεις μεταξύ των οντολογιών μπορούν να βασιστούν στις συσχετίσεις και τις εξαρτήσεις μεταξύ της δομής των ονομασιών των οντολογιών, καθώς και των αντιστοιχών στιγμιότυπων. Στην περίπτωση λοιπόν που πρέπει να πραγματοποιηθεί Οντολογική Ευθυγράμμιση,

διαφορετικές πιθανότητες ομοιότητας μεταξύ των διαφορετικών στοιχείων ανατίθενται, και ανάλογα με τα αποτελέσματα αυτά, τα δεδομένα υγείας που ανήκουν στις οντολογίες που έχουν ευθυγραμμιστεί καλύτερα με τις οντολογίες των πόρων HL7 FHIR, μετασχηματίζονται σε HL7 FHIR μορφή, ακολουθώντας τα στάδια που περιγράφονται από την Τεχνική Μετάφρασης Δομής FHIR (Κεφάλαιο 4.2.4). Ένα παράδειγμα της διαδικασίας που ακολουθείται για την οντολογική ευθυγράμμιση διαφαίνεται στην Εικόνα 4-25. Το πρώτο στάδιο σχετίζεται με την εφαρμογή των προαναφερθέντων βημάτων, ενώ ακολούθως τα εν λόγω αποτελέσματα συγχωνεύονται, επεξεργάζονται και εξάγονται αντιστοίχως, προκειμένου να αποφευχθούν εσφαλμένα συμπεράσματα.



Εικόνα 4-25 - Παράδειγμα οντολογικής ευθυγράμμισης

4.2.4 Τεχνική Μετάφρασης Δομής FHIR

Βασικός στόχος της Τεχνικής Μετάφρασης Δομής FHIR είναι να παρέχει μία τεχνική που προσφέρει τη δυνατότητα να παρέχει την τελική μορφή των οντολογιών σε HL7 FHIR μορφή. Επομένως, σύμφωνα με τα προαναφερθέντα βήματα, στην περίπτωση που η Τεχνική Συντακτικής & Σημασιολογικής Ομοιότητας εξάγει έμπιστα αποτελέσματα, τότε μία οντολογία χαρακτηρίζεται πως αντιστοιχεί σε έναν συγκεκριμένο πόρο HL7 FHIR στην περίπτωση που η ολική ομοιότητά τους είναι μεγαλύτερη του 90%. Η συγκεκριμένη τεχνική επαναλαμβάνεται για όλες τις αποθηκευμένες οντολογίες, προκειμένου να αναγνωρίσει τις οντολογίες των δεδομένων υγείας και να τα μετατρέψει στη HL7 FHIR μορφή τους. Αντίστοιχα, στην περίπτωση που η Τεχνική Συντακτικής & Σημασιολογικής Ομοιότητας δεν εξάγει έμπιστα αποτελέσματα, τότε ακολουθούνται τα βήματα της Οντολογικής Ευθυγράμμισης.

Σε κάθε περίπτωση, ανάλογα με τα προαναφερθέντα αποτελέσματα, η Τεχνική Μετάφρασης Δομής FHIR λαμβάνει ως όρισμα για κάθε οντολογική ονομασία, τον πόρο HL7 FHIR με τον οποίο έχουν εξαχθεί τα μέγιστα αποτελέσματα ομοιότητας (είτε μέσω της Τεχνικής Συντακτικής & Σημασιολογικής Ομοιότητας, είτε μέσω της Τεχνικής Οντολογικής Ευθυγράμμισης), μετασχηματίζοντάς το στη δομή του HL7 FHIR (δηλαδή Resource.attribute). Η ίδια διαδικασία επαναλαμβάνεται για όλες τις κλάσεις, παρέχοντας την τελική μορφή της προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων. Αξίζει να αναφερθεί πως προκειμένου να ανατεθούν εκ νέου τα στιγμιότυπα που είναι καταχωρημένα στη Γνωσιακή Βάση Δεδομένων, πραγματοποιούνται επερωτήματα μεταξύ των διαφορετικών στιγμιότυπων κάθε ξεχωριστής κλάσης ώστε να επανα-ταυτοποιηθούν και να επανα-τοποθετηθούν τα στιγμιότυπα αυτά, στους συγκεκριμένους πόρους HL7 FHIR.

ΚΕΦΑΛΑΙΟ 5

Αξιολόγηση

Περιεχόμενα

5.1. Περιβάλλον Δοκιμών

5.2. Αξιολόγηση Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων

5.2.1. 1^ο Πείραμα: Εφαρμογή Τεχνικών Συντακτικής και Σημασιολογικής Ομοιότητας

5.2.2. 2^ο Πείραμα: Συγκριτική Αξιολόγηση Απόδοσης Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων

Περίληψη

Στο **ΚΕΦΑΛΑΙΟ 5**, λαμβάνει μέρος η αξιολόγηση της προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων, μέσω δύο διαφορετικών περιπτώσεων χρήσης. Η πρώτη περίπτωση χρήσης σχετίζεται με ένα σύνολο δεδομένων υγείας από ανώνυμους ασθενείς ενός Ινστιτούτου Υγείας, στο οποίο περιγράφονται όλα τα εξαγόμενα αποτελέσματα της προσέγγισης, ακολουθούμενα από σχετική συζήτηση και ανάλυσή τους. Η δεύτερη περίπτωση χρήση σχετίζεται πάλι με ένα σύνολο δεδομένων από έναν ανώνυμο ασθενή ο οποίος προσφέρει τα δεδομένα του, όπως προήλθαν από μια φορητή συσκευή. Τα βήματα της προσέγγισης επίσης αναλύονται, καταλήγοντας σε παρόμοια συμπεράσματα, ενώ παράλληλα πραγματοποιείται και αναλύεται μία συγκριτική αξιολόγηση της παρούσας προσέγγισης με μία άλλη προσέγγιση μετασχηματισμού και αντιστοίχισης δεδομένων.

5.1 Περιβάλλον Δοκιμών

Η προτεινόμενη προσέγγιση αναπτύχθηκε σε Java SE χρησιμοποιώντας το εργαλείο NetBeans IDE v8.0.2 [272]. Η παρούσα μελέτη χρησιμοποίησε ένα περιβάλλον επεξεργασίας μνήμης RAM 16GB, επεξεργαστή Intel i7-4790 @ 3.60 GHz x 8 CPU Cores, 2TB χώρο αποθήκευσης και Windows 10 λειτουργικό σύστημα.

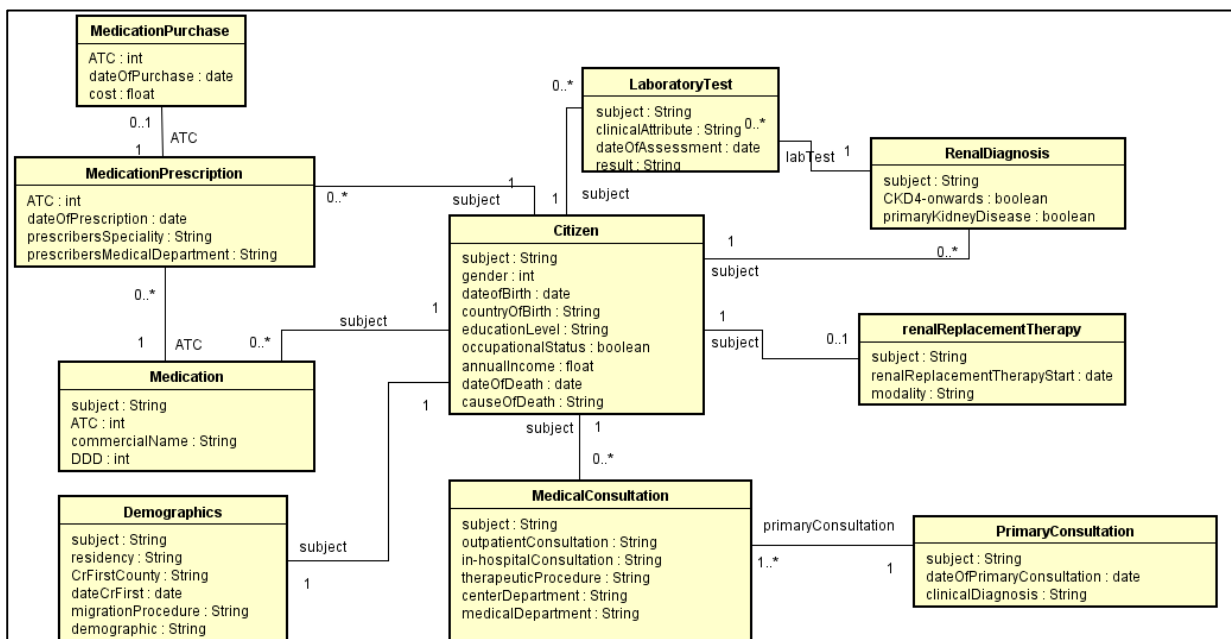
Συνολικά εκτελέστηκαν 24 πειράματα με διαφορετικά σύνολα δεδομένων, ποικίλων μεγεθών. Στο παρόν κεφάλαιο, θα αναλυθούν συνολικά δύο πειραματικές προσεγγίσεις, των οποίων τα αποτελέσματα αξίζουν να συζητηθούν και να αναλυθούν.

5.2 Αξιολόγηση Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων

5.2.1 1^ο Πείραμα: Εφαρμογή Τεχνικών Συντακτικής και Σηματολογικής Ομοιότητας συνδυαστικά με την Τεχνική Οντολογικής Ευθυγράμμισης

5.2.1.1 Περιγραφή Συνόλου Δεδομένων Πειράματος

Η πρώτη περίπτωση χρήσης εκμεταλλεύεται ένα σύνολο δεδομένων της υγειονομικής περιθαλψης δομημένο σε μορφή CSV, καλύπτοντας τα βήματα της προτεινόμενης προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων. Η Εικόνα 5-1 παρουσιάζει ένα εννοιολογικό διάγραμμα των οντοτήτων του συνόλου δεδομένων περίπτωσης χρήσης, αποκαλύπτοντας τις αλληλεπιδράσεις, καθώς και τα εσωτερικά χαρακτηριστικά κάθε οντότητας. Το σύνολο δεδομένων που χρησιμοποιήθηκε για την αξιολόγηση της αποτελεσματικότητας της προτεινόμενης προσέγγισης (Εικόνα 5-2) ήταν ένα υποσύνολο προσωπικών πληροφοριών ανώνυμων πολιτών, που προέρχονταν από το Ινστιτούτο Karolinska. Αναλυτικότερα, έχει το μέγεθος των 61 kilobytes (KB) και αποτελείται από 250 διαφορετικές περιπτώσεις προσωπικών πληροφοριών ορισμένων πολιτών, σχετικά με: (i) το προσωπικό αναγνωριστικό (subject), (ii) το φύλο (gender), (iii) την ημερομηνία γέννησης (dateOfBirth), (iv) την ημερομηνία θανάτου (dateOfDeath), και (v) την αιτία θανάτου (causeOfDeath).



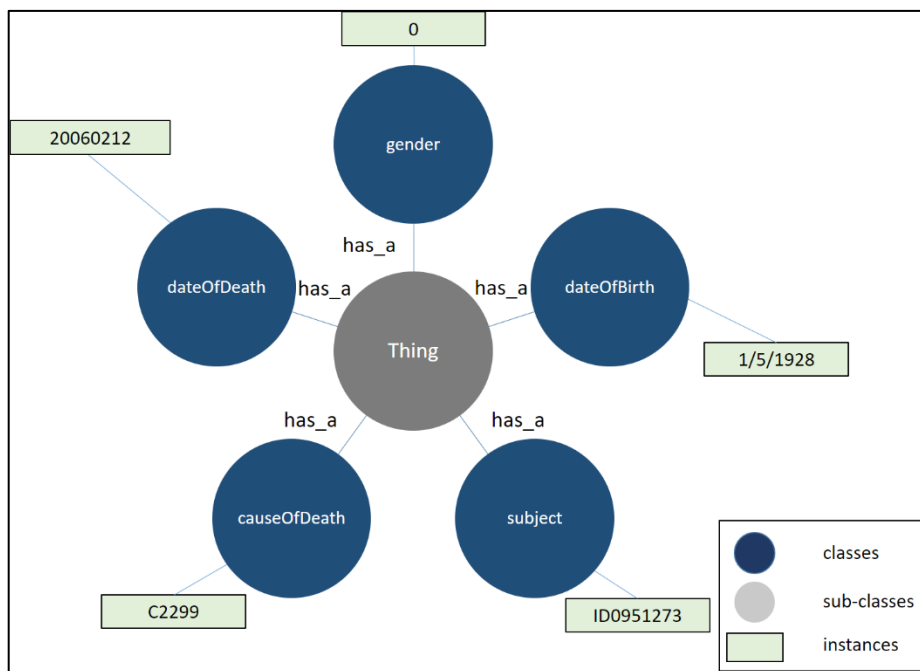
Εικόνα 5-1 – Εννοιολογικό διάγραμμα οντοτήτων συνόλου δεδομένων πρώτου πειράματος

1	subject;gender;dateOfBirth;dateOfDeath;causeOfDeath
2	ID0054619;1;1/6/1954;NA;NA
3	ID1012175;0;1/6/1968;NA;NA
4	ID0332172;0;1/11/1961;NA;NA
5	ID0186244;0;1/7/1963;NA;NA
6	ID0782623;1;1/5/1957;NA;NA
7	ID0944842;1;1/5/1961;NA;NA
8	ID0131218;1;1/7/1976;NA;NA
9	ID0131218;1;1/12/1951;NA;NA
10	ID0911095;1;1/11/1945;NA;NA
11	ID0951273;0;1/5/1928;20060212;C2299
12	ID0695153;1;1/4/1980;NA;NA
13	ID0695153;0;1/4/1977;NA;NA
14	ID0695153;1;1/4/1959;NA;NA
15	ID0695153;1;1/6/1951;NA;NA
16	ID1164801;0;1/1/1978;NA;NA
17	ID1409311;0;1/9/1960;NA;NA
18	ID0097871;0;1/8/1950;NA;NA
19	ID1095873;1;1/6/1924;NA;NA
20	ID0250381;0;1/10/1954;NA;NA

Εικόνα 5-2 – Δεδομένα περίπτωσης χρήσης για το πρώτο πείραμα

5.2.1.2 Αποτελέσματα Πειράματος

Επίπεδο Ontologies & Relationships: Στο εν λόγω βήμα, το ιεραρχικό δέντρο του συνόλου δεδομένων περίπτωσης χρήσης (Εικόνα 5-3) δημιουργήθηκε, όπου αντιπροσωπεύει τις έννοιες (κλάσεις), τις σχέσεις (ιδιότητες) και τα αξιώματά του (στιγμιότυπα). Επιπλέον, οι πόροι HL7 FHIR κατασκευάζονται επίσης ως οντολογίες, λαμβάνοντας ως είσοδο την XML αναπαράσταση κάθε ενός από τους διαφορετικούς πόρους HL7 FHIR.



Εικόνα 5-3 – Οντολογική αναπαράσταση δεδομένων περίπτωσης χρήσης για το πρώτο πείραμα

Επίπεδο Syntactic Similarity Identifier: Ο Πίνακας 5-1 δημιουργείται απεικονίζοντας τους πόρους HL7 FHIR με υψηλότερους βαθμούς συντακτικής ομοιότητας, σε σύγκριση με τα ονόματα των εννοιών του συνόλου δεδομένων.

Πίνακας 5-1 - Πόροι HL7 FHIR με υψηλότερους βαθμούς συντακτικής ομοιότητας

Use Case Dataset Attribute	HL7 FHIR Resources	Levenshtein Distance
subject	<i>Patient.contact</i>	0.267 (~27%)
gender	<i>Patient.gender</i>	0.643 (~65%)
dateOfBirth	<i>Patient.multipleBirth</i>	0.381 (~38%)
dateOfDeath	<i>Patient.contact.relationship</i>	0.250 (25%)
causeOfDeath	<i>Patient.contact.gender</i>	0.278 (~28%)

Η Εικόνα 5-4 απεικονίζει ένα παράδειγμα των αποτελεσμάτων του Syntactic Similarity Identifier, για το χαρακτηριστικό του φύλου (gender), για το οποίο υπολογίστηκε η απόσταση Levenshtein για τα διαφορετικά χαρακτηριστικά HL7 FHIR του πόρου του Ασθενούς (Patient resource) του HL7 FHIR.

```

193 0,176 <-- (entry.gender, #document.Patient)
194 0,333 <-- (entry.gender, Patient.identifier)
195 0,143 <-- (entry.gender, Patient.active)
196 0,167 <-- (entry.gender, Patient.name)
197 0,200 <-- (entry.gender, Patient.telecom)
198 0,643 <-- (entry.gender, Patient.gender)
199 0,235 <-- (entry.gender, Patient.birthDate)
200 0,313 <-- (entry.gender, Patient.deceased)
201 0,200 <-- (entry.gender, Patient.address)
202 0,190 <-- (entry.gender, Patient.maritalStatus)
203 0,238 <-- (entry.gender, Patient.multipleBirth)
204 0,077 <-- (entry.gender, Patient.photo)
205 0,200 <-- (entry.gender, Patient.contact)
206 0,250 <-- (entry.gender, contact.relationship)
207 0,167 <-- (entry.gender, contact.name)
208 0,267 <-- (entry.gender, contact.telecom)
209 0,267 <-- (entry.gender, contact.address)
210 0,571 <-- (entry.gender, contact.gender)
211 0,250 <-- (entry.gender, contact.organization)
212 0,286 <-- (entry.gender, contact.period)
213 0,143 <-- (entry.gender, Patient.animal)
214 0,286 <-- (entry.gender, animal.species)
215 0,167 <-- (entry.gender, animal.breed)
216 0,421 <-- (entry.gender, animal.genderStatus)
217 0,190 <-- (entry.gender, Patient.communication)
218 0,182 <-- (entry.gender, communication.language)
219 0,217 <-- (entry.gender, communication.preferred)
220 0,259 <-- (entry.gender, Patient.generalPractitioner)
221 0,214 <-- (entry.gender, Patient.managingOrganization)
222 0,083 <-- (entry.gender, Patient.link)
223 0,250 <-- (entry.gender, link.other)
224 0,167 <-- (entry.gender, link.type)

```

Εικόνα 5-4 - Παράδειγμα αποτελεσμάτων της τεχνικής συντακτικής ομοιότητας

Επίπεδο Syntactic Similarity Identifier: Ο Πίνακας 5-2 δημιουργείται απεικονίζοντας τους πόρους HL7 FHIR με υψηλότερους βαθμούς σημασιολογικής ομοιότητας, σε σύγκριση με τα ονόματα των εννοιών, των σχέσεων και των αξιωμάτων του συνόλου δεδομένων περίπτωσης χρήσης.

Πίνακας 5-2 - Πόροι HL7 FHIR με υψηλότερους βαθμούς σημασιολογικής ομοιότητας

Use Case Dataset Attribute	HL7 FHIR Resources	Cortical.io
subject	<i>Patient.identifier</i>	0.721 (~72%)
gender	<i>Patient.gender</i>	0.591 (~59%)
dateOfBirth	<i>Patient.birthDate</i>	0.872 (~87%)
dateOfDeath	<i>Patient.deceased</i>	0.793 (79%)
causeOfDeath	<i>Patient.contact.address</i>	0.209 (~21%)

Η Εικόνα 5-5 απεικονίζει ένα παράδειγμα των αποτελεσμάτων του Semantic Similarity Identifier για το χαρακτηριστικό του φύλου (gender), για το οποίο υπολογίστηκε η σημασιολογική απόσταση δακτυλικών αποτυπωμάτων για τα χαρακτηριστικά του πόρου HL7 FHIR του Ασθενούς (Patient resource).

```

197 0,156 <-- (entry.gender, Patient.identifier)
198 0,203 <-- (entry.gender, Patient.active)
199 0,176 <-- (entry.gender, Patient.name)
200 0,152 <-- (entry.gender, Patient.telecom)
201 0,591 <-- (entry.gender, Patient.gender)
202 0,132 <-- (entry.gender, Patient.birthDate)
203 0,187 <-- (entry.gender, Patient.deceased)
204 0,246 <-- (entry.gender, Patient.address)
205 0,117 <-- (entry.gender, Patient.maritalStatus)
206 0,117 <-- (entry.gender, Patient.multipleBirth)
207 0,185 <-- (entry.gender, Patient.photo)
208 0,196 <-- (entry.gender, Patient.contact)
209 0,247 <-- (entry.gender, contact.relationship)
210 0,222 <-- (entry.gender, contact.name)
211 0,198 <-- (entry.gender, contact.telecom)
212 0,278 <-- (entry.gender, contact.address)
213 0,555 <-- (entry.gender, contact.gender)
214 0,269 <-- (entry.gender, contact.organization)
215 0,235 <-- (entry.gender, contact.period)
216 0,147 <-- (entry.gender, Patient.animal)
217 0,122 <-- (entry.gender, animal.species)
218 0,140 <-- (entry.gender, animal.breed)
219 0,126 <-- (entry.gender, animal.genderStatus)
220 0,201 <-- (entry.gender, Patient.communication)
221 0,213 <-- (entry.gender, communication.language)
222 0,265 <-- (entry.gender, communication.preferred)
223 0,117 <-- (entry.gender, Patient.generalPractitioner)
224 0,117 <-- (entry.gender, Patient.managingOrganization)
225 0,192 <-- (entry.gender, Patient.link)
226 0,183 <-- (entry.gender, link.other)
227 0,170 <-- (entry.gender, link.type)

```

Εικόνα 5-5 - Παράδειγμα αποτελεσμάτων της τεχνικής συντακτικής ομοιότητας

Επίπεδο Overall Ontology Mapper: Ο Πίνακας 5-3 δημιουργείται με την απεικόνιση των μεγαλύτερων υπολογισμένων ζευγών των αποτελεσμάτων του Syntactic Similarity και του Semantic Similarity Identifier, παρέχοντας τα χαρακτηριστικά του συνόλου δεδομένων περίπτωσης χρήσης με υψηλότερες πιθανότητες ομοιότητας με τον αναγνωρισμένο πόρο HL7 FHIR.

Πίνακας 5-3 - Μεγαλύτερα υπολογισμένα ζεύγη των αποτελεσμάτων του Syntactic Similarity και του Semantic Similarity Identifier

Use Case Dataset Attribute	Syntactic Similarity Result	Similarity	Semantic Similarity Result	Similarity	Overall Mean
subject	<i>Patient. identifier</i>	0.167 (~17%)	<i>Patient. identifier</i>	0.721 (~72%)	~45%
gender	<i>Patient. gender</i>	0.643 (~65%)	<i>Patient. gender</i>	0.591 (~59%)	~62%
dateOfBirth	<i>Patient. birthDate</i>	0.176 (~18%)	<i>Patient. birthDate</i>	0.872 (~87%)	~53%
dateOfDeath	<i>Patient. deceased</i>	0.176 (~18%)	<i>Patient. deceased</i>	0.793 (79%)	~49%
causeOfDeath	<i>Patient. contact. gender</i>	0.278 (~28%)	<i>Patient. contact. address</i>	0.209 (~21%)	NO MATCHING

5.2.1.3 Συζήτηση Αποτελεσμάτων Πειράματος

Μία τεχνική αντιστοίχισης οντολογίας της υγειονομικής περίθαλψης προτάθηκε για τη μετατροπή των δεδομένων της υγειονομικής περίθαλψης σε μορφή HL7 FHIR, παράγοντας αποτελέσματα υψηλής ακρίβειας και αξιοπιστίας. Με περισσότερες λεπτομέρειες, με βάση αυτή την προσέγγιση οντολογικής αντιστοίχισης, μέσω των αποτελεσμάτων του Πίνακας 5-3 καθίσταται σαφές ότι η αντιστοίχιση των δύο διαφορετικών οντολογιών δεν είναι εύκολη διαδικασία. Αυτός είναι ο λόγος για τον οποίο στη συγκεκριμένη προσέγγιση απαιτούνται όχι μόνο σημασιολογικές τεχνικές αντιστοίχισης, αλλά και συντακτικές.

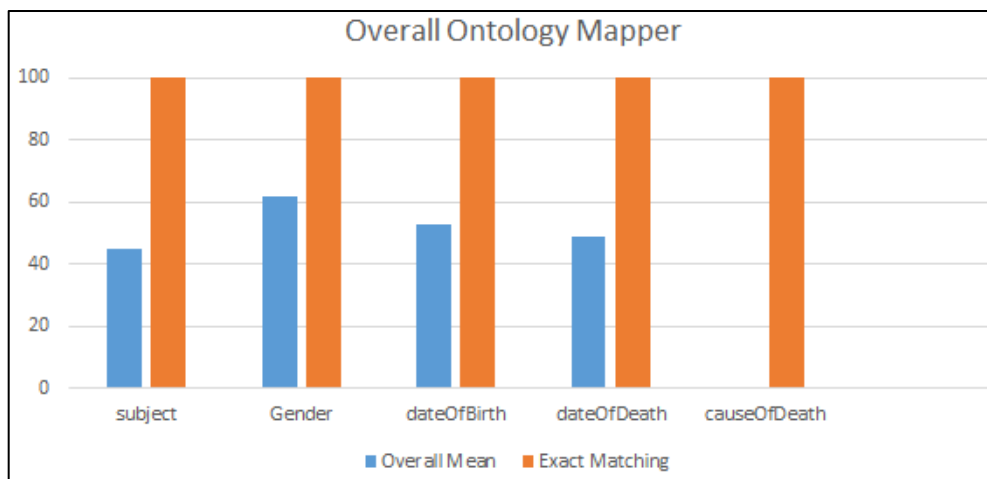
Πιο συγκεκριμένα, μπορεί να παρατηρηθεί ότι υπάρχουν αρκετές περιπτώσεις που μια οντολογία έχει ταιριάζει με ένα συγκεκριμένο HL7 FHIR Resource εξαιτίας των αποτελεσμάτων του Syntactic Similarity Identifier, ενώ η ίδια οντολογία έχει ταιριάζει με ένα διαφορετικό HL7 FHIR Resource εξαιτίας των αποτελεσμάτων του Semantic Similarity Identifier. Ένα συγκεκριμένο παράδειγμα αυτού του σεναρίου είναι η περίπτωση του attribute dateOfDeath που αντιστοιχούσε στον πόρο Patient.contact.relationship HL7 FHIR λόγω της συντακτικής του ομοιότητας, ενώ αντιστοιχούσε στον πόρο Patient.deceased HL7 FHIR βάσει της σημασιολογικής ομοιότητάς του. Ωστόσο, λόγω των αποτελεσμάτων του Overall Ontology Mapper, το χαρακτηριστικό dateOfDeath προσαρμόστηκε σωστά στον σωστό πόρο HL7 FHIR, με βάση το μέσο όρο των συνολικών αποτελεσμάτων. Συνεπώς, δεν είμαστε σε θέση να δημιουργήσουμε μοτίβα και κανόνες που να λένε ότι στην περίπτωση που υπάρχει συντακτική ομοιότητα, τότε θα υπάρχει επίσης σημασιολογική ομοιότητα μεταξύ των ίδιων χαρακτηριστικών και αντίστροφα.

Επιπλέον, το σύνολο δεδομένων της συγκεκριμένης περίπτωσης χρήσης μετασχηματίστηκε χειροκίνητα στο HL7 FHIR, προκειμένου να συγκριθούν τα αποτελέσματα της αναπτυγμένης τεχνικής με τα πραγματικά αποτελέσματα. Αυτός ήταν ο βασικός λόγος για τον οποίο επιλέχθηκε ένα μικρό δείγμα δεδομένων για την αξιολόγηση της τεχνικής, προκειμένου να παραχθεί χειροκίνητα πιο εύκολα το προαναφερθέν αποτέλεσμα. Για τα χειροκίνητα αποτελέσματα, υποτίθεται ότι ήταν υψηλής ακρίβειας, τηρώντας τις κατευθυντήριες γραμμές της μορφής HL7 FHIR, που θεωρούνται ως σημείο αναφοράς (τιμή αναφοράς) υψηλής ποιότητας και ακρίβειας. Ο Πίνακας 5-4 απεικονίζει τα αποτελέσματα του χειροκίνητου μετασχηματισμού, σε σύγκριση με τα αποτελέσματα της παρούσας προσέγγισης.

Πίνακας 5-4 - Αποτελέσματα του χειροκίνητου μετασχηματισμού

Use Case Dataset Attribute	Manual Results	Similarity	Automatic Results	Similarity
subject	<i>Patient. identifier</i>	100%	<i>Patient. identifier</i>	~45%
gender	<i>Patient. gender</i>	100%	<i>Patient. gender</i>	~62%
dateOfBirth	<i>Patient. birthDate</i>	100%	<i>Patient. birthDate</i>	~53%
dateOfDeath	<i>Patient. deceased</i>	100%	<i>Patient. deceased</i>	~49%
causeOfDeath	NO MATCHING	NO MATCHING	<i>Patient. contact.address</i>	NO MATCHING

Στην Εικόνα 5-6, απεικονίζεται ένα ραβδόγραμμα των προαναφερθέντων αποτελεσμάτων, που δείχνει τη διαφορά μεταξύ της υπολογισμένης αντιστοίχισης (Overall Mean) και της ακριβούς αντιστοίχισης (χειροκίνητα αποτελέσματα).



Εικόνα 5-6 – Απεικόνιση διαφοράς μεταξύ της υπολογισμένης αντιστοίχισης και της περίπτωσης ακριβούς αντιστοίχισης (χειροκίνητα αποτελέσματα)

Είναι σαφές ότι η προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων παρείχε αποτελέσματα ακρίβειας 100%. Ωστόσο, μπορεί να παρατηρηθεί διαφορά στο συνολικό ποσοστό, το οποίο όμως δεν επηρεάζει την τελική

αντιστοίχιση. Θα πρέπει να σημειωθεί ότι υπάρχουν περιπτώσεις όπου τα αποτελέσματα της προσέγγισης δεν ήταν αποτελεσματικά ή σωστά, όπως στην περίπτωση του χαρακτηριστικού causeOfDeath όπου τα αποτελέσματα του χειροκίνητου μετασχηματισμού δεν παρείχαν κανένα αντίστοιχο πόρο HL7 FHIR, ενώ το ίδιο αποτέλεσμα επίσης προήλθε από την παρούσα προσέγγιση. Παρ' όλα αυτά, σε αυτή την περίπτωση και τα δύο αποτελέσματα είναι σωστά, αφού στην τρέχουσα έκδοση του HL7 FHIR (v3.0.1) δεν υπάρχει πόρος που να περιγράφει ακόμη αυτό το ειδικό χαρακτηριστικό.

Θα πρέπει να σημειωθεί ότι αυτή η τεχνική εφαρμόζεται κατά τέτοιο τρόπο ώστε να είναι εφαρμόσιμος στις πιο συνηθισμένες μορφές δεδομένων (π.χ. αρχεία JSON, CSV, XML) από τη στιγμή που το αρχικό βήμα της τεχνικής είναι να μετασχηματίσει οποιοδήποτε είδος μορφής δεδομένων σε γλώσσα XML, μέσω του Μετασχηματιστή Δεδομένων. Προκειμένου να προσδιοριστεί η δυνατότητα εφαρμογής των τελευταίων, πραγματοποιήθηκαν επιπρόσθετες αξιολογήσεις σε δεδομένα μορφής JSON μικρότερων και μεγαλύτερων μεγεθών δεδομένων, οι οποίες επαλήθευσαν το εν λόγω αποτέλεσμα. Πιο συγκεκριμένα, έγινε το ίδιο πείραμα για ένα σύνολο δεδομένων 12 KB με λιγότερες μετρήσεις (54 μετρήσεις), ενώ επαναλήφθηκε και για σύνολο δεδομένων 2.032 KB, που περιέχει μεγαλύτερο αριθμό μετρήσεων (2.973 μετρήσεις). Ο λόγος που πραγματοποιήθηκαν αυτά τα πρόσθετα πειράματα ήταν να προσδιοριστεί ο τρόπος συμπεριφοράς της συνολικής προσέγγισης με βάση το μέγεθος του συνόλου δεδομένων, σύμφωνα με τη συνολική πρόοδο μετασχηματισμού.

Τεχνική Οντολογικής Ευθυγράμμισης

Σε τελικό στάδιο, αξίζει να σημειωθεί πως τα περισσότερα από τα αποτελέσματα φαίνεται πως δε ξεπέρασαν το όριο ομοιότητας του 90%, με αποτέλεσμα η προσέγγιση να τα θεωρήσει αναξιόπιστα, και έτσι να τα αγνοεί, σύμφωνα με τη λειτουργικότητα της Τεχνικής Ολικής Οντολογικής Αντιστοίχισης. Ωστόσο, για την αποφυγή επανάληψης του ίδιου πειράματος με σκοπό την εμφάνιση των αποτελεσμάτων της Τεχνικής Οντολογικής Ευθυγράμμισης, η τεχνική παραμετροποιήθηκε ώστε να μην αγνοήσει τα αποτελέσματα που δεν τηρούσαν το κριτήριο, προκειμένου να επαναληφθεί η ίδια διαδικασία εκτελώντας τα βήματα που περιγράφονται στο πλαίσιο της Οντολογικής Ευθυγράμμισης, ώστε εν τέλει να πραγματοποιηθεί η αντίστοιχη μετάφραση των δεδομένων υγείας σε HL7 FHIR. Έτσι, έχοντας τόσο τα αποτελέσματα που προέκυψαν από την Τεχνική Ολικής Οντολογικής Αντιστοίχισης, όσο και τα αποτελέσματα που προέκυψαν από την τεχνική Οντολογικής Ευθυγράμμισης, πραγματοποιήθηκε σύγκριση μεταξύ τους.

Στο πλαίσιο της Οντολογικής Ευθυγράμμισης, χρησιμοποιήθηκε η τεχνική του Alignment API για την εξεύρεση της καλύτερης αντιστοιχίας μεταξύ του συνόλου δεδομένων υγειονομικής περίθαλψης και των οντολογιών των πόρων HL7 FHIR. Προκειμένου να παρουσιαστεί η εφαρμογή του Alignment API, εφαρμόστηκε ένας συγκεκριμένος επεξεργαστής ο οποίος είναι ικανός να διαβάσει δύο οντολογίες OWL/RDF, να δημιουργήσει ένα αντικείμενο ευθυγράμμισης, να υπολογίσει την ευθυγράμμιση μεταξύ αυτών των οντολογιών και τελικά να απεικονίσει το αποτέλεσμα. Τα αποτελέσματα εκφράζονται στη μορφή ευθυγράμμισης (δηλ. RDF/XML), που περιέχει μεταδεδομένα σχετικά με την ευθυγράμμιση και το αντίστοιχο σύνολο αντιστοιχιών, για κάθε διαφορετική οντότητα των οντολογιών. Ωστόσο, δεδομένου ότι η προεπιλεγμένη μέθοδος που χρησιμοποιήθηκε για την ευθυγράμμιση των οντολογιών συγκρίνει μόνο τις ονομασίες των οντοτήτων για την εύρεση της αντιστοιχίας, αυτή η συγκεκριμένη μέθοδος διαμορφώθηκε για να εκτελεστεί με βάση τη συνάρτηση απόστασης της απόστασης Levenshtein - όπως στην περίπτωση του Syntactic Similarity Identifier. Σε αυτό το σενάριο, είναι σαφές ότι οι αντιστοιχίες περιέχουν παράγοντες ομοιότητας διαφορετικούς από το 1 (δηλαδή 100% ομοιότητα), ενώ υπάρχουν συμβολοσειρές που δεν είναι οι ίδιες και διατίθενται συμπληρωματικές αντιστοιχίες (Εικόνα 5-7). Ωστόσο, οι συνδυασμοί με τις μεγαλύτερες τιμές ομοιότητας διατηρήθηκαν, ενώ οι άλλοι απορρίφθηκαν.

```

1 <?xml version='1.0' encoding='utf-8' standalone='no'?>
2 <rdf:RDF xmlns='http://knowledgeweb.semanticweb.org/heterogeneity/alignment#'
3     xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
4     xmlns:xsd='http://www.w3.org/2001/XMLSchema#'
5     xmlns:align='http://knowledgeweb.semanticweb.org/heterogeneity/alignment#'>
6 <Alignment>
7   <xml>yes</xml>
8   <level>0</level>
9   <type>**</type>
10  <method>fr.inrialpes.exmo.align.impl.method.StringDistAlignment</method>
11  <time>309</time>
12  <ontol>
13    <Ontology rdf:about="http://alignapi.gforge.inria.fr/FHIR.owl">
14      <location>file:///user/FHIR.owl</location>
15      <formalism>
16        <Formalism align:name="OWL2.0" align:uri="http://www.w3.org/2002/07/owl#" />
17      </Formalism>
18    </Ontology>
19  </ontol>
20  <ontol2>
21    <Ontology rdf:about="http://alignapi.gforge.inria.fr/UCdata.owl">
22      <location>file:///user/UCdata.owl</location>
23      <formalism>
24        <Formalism align:name="OWL2.0" align:uri="http://www.w3.org/2002/07/owl#" />
25      </formalism>
26    </Ontology>
27  </ontol2>
28  <map>
29    <Cell>
30      <entity1 rdf:resource="http://alignapi.gforge.inria.fr/FHIR.owl#PatientGender"/>
31      <entity2 rdf:resource="http://alignapi.gforge.inria.fr/UCdata.owl#EntryGender"/>
32      <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.6142957142847141</measure>
33      <relation>=</relation>
34    </Cell>
35  </map>

```

Εικόνα 5-7 – Αποτελέσματα οντολογικής ευθυγράμμισης

Η αξιολόγηση αυτών των ευθυγραμμίσεων εκτελέστηκε και πάλι με βάση τη μη αυτόματη μετάφραση του συνόλου δεδομένων περίπτωσης χρήσης στο HL7 FHIR, προκειμένου να συγκριθούν τα αποτελέσματα του Alignment API, με τα πραγματικά αποτελέσματα, καθώς και με τα αποτελέσματα της προσέγγισης. Σε όλες τις περιπτώσεις, τα αποτελέσματα ήταν ακριβώς τα ίδια, αποδεικνύοντας την αποτελεσματικότητα τόσο της αναπτυγμένης τεχνικής, όσο και του Alignment API. Ωστόσο, απεικονίστηκαν μερικές μικρές διαφορές στον συνολικό χρόνο εκτέλεσης μεταξύ αυτών των δύο προσεγγίσεων, όπου το Alignment API είχε καλύτερο χρόνο εκτέλεσης. Λαμβάνοντας υπόψη τα συνολικά αποτελέσματα, αποδεικνύεται η αποτελεσματικότητα της αναπτυγμένης προσέγγισης, ανεξαρτήτως του αν τηρούνται τα όρια (δηλαδή 90%) που έχουν οριστεί από την Τεχνική Ολικής Οντολογικής Αντιστοίχισης. Κάτι τέτοιο σημαίνει πως το εν λόγω όριο, έχει τη δυνατότητα να μειωθεί πιθανότατα στο 40% λόγω της εγκυρότητας των αποτελεσμάτων του πειράματος.

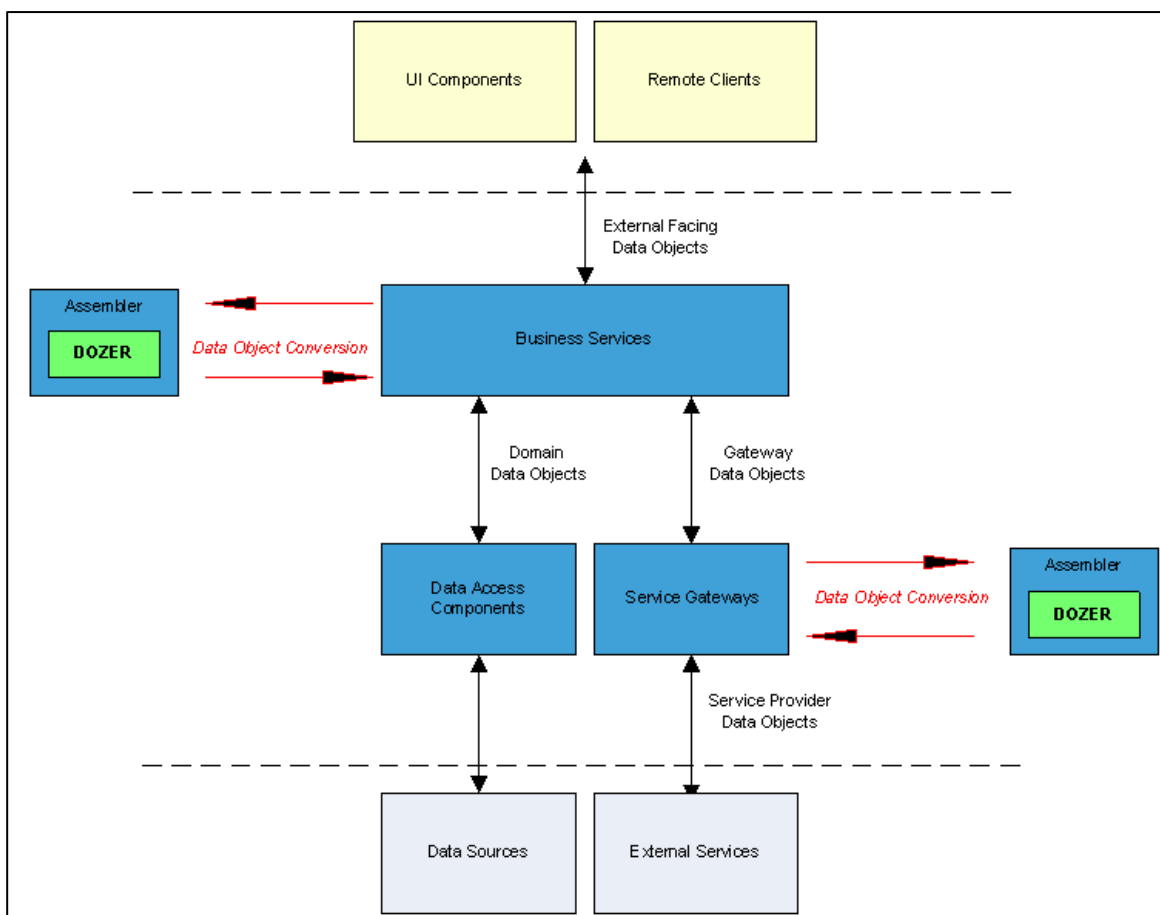
5.2.2 2^ο Πείραμα: Συγκριτική Αξιολόγηση Απόδοσης Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων

5.2.2.1 Περιγραφή Μηχανισμού Σύγκρισης

Το Apache Dozer [273] είναι ένας μηχανισμός αντιστοίχισης Java Bean [274] σε Java Bean ο οποίος είναι κατασκευασμένος για την αντιγραφή δεδομένων από το ένα αντικείμενο στο άλλο. Όπως περιγράφεται στο Dozer, αυτά τα Java Beans είναι διαφορετικών σύνθετων τύπων. Το Dozer υποστηρίζει την απλή αντιστοίχιση ιδιοτήτων, τη αντιστοίχιση σύνθετου τύπου, την αμφίδρομη αντιστοίχιση, τη σιωπηρή αντιστοίχιση και την αναδρομική αντιστοίχιση, από πλευράς χαρακτηριστικών αντιστοίχισης συλλογής που πρέπει να αντιστοιχηθούν στο επίπεδο στοιχείων. Το Dozer μπορεί να πραγματοποιεί αντιστοιχίσεις μεταξύ των ονομάτων χαρακτηριστικών και μπορεί να μετατρέπει με αυτόματο τρόπο διαφορετικούς τύπους δεδομένων. Επιτρέπει επίσης στους ενδιαφερόμενους να καθορίσουν προσαρμοσμένες μετατροπές δεδομένων χρησιμοποιώντας XML. Στις περισσότερες περιπτώσεις, η αντιστοίχιση μπορεί να εκτελεστεί αυτόματα από το Dozer χρησιμοποιώντας αντανάκλαση, αλλά παρέχει επίσης στον χρήστη τη δυνατότητα να εκτελεί προσαρμοσμένη αντιστοίχιση που μπορεί να περιγραφεί σε μορφή XML. Δεδομένου ότι η αντιστοίχιση είναι αμφίδρομη, πρέπει να οριστεί μόνο μία σχέση μεταξύ των κλάσεων.

Μεταξύ των διαφορών μεταξύ της προσέγγισης του Dozer και της παρούσας προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων είναι το γεγονός ότι στην περίπτωση του Dozer, οι αντιστοιχίσεις πρέπει να υπάρχουν στην εκκίνηση του Dozer και δεν μπορούν να τροποποιηθούν στη συνέχεια. Από την άλλη πλευρά, οι αντιστοιχίσεις στην περίπτωση της παρούσας προσέγγισης δεν έχουν τέτοιους περιορισμούς, ενώ εκτελούνται με δυναμικό τρόπο που βασίζεται όχι μόνο σε συγκεκριμένους κανόνες αντιστοίχισης αλλά και στη σημασιολογική σημασία των χαρακτηριστικών που θα αντιστοιχηθούν. Επιπλέον, η αντιστοίχιση στην προσέγγιση Dozer πραγματοποιείται μεταξύ αντικειμένων δεδομένων, πράγμα που σημαίνει ότι όλα τα χαρακτηριστικά που πρέπει να αντιστοιχηθούν θα πρέπει να μετατραπούν σε Java Beans, ενώ ο προγραμματιστής αναγκάζεται να αντιγράψει όλα τα ονόματα κλάσεων Bean. Στην περίπτωση της παρούσας προσέγγισης, τα χαρακτηριστικά δεν χρειάζονται ούτε μετασχηματισμό ούτε προ-επεξεργασία, πέραν του αυτοματοποιημένου μετασχηματισμού τους σε οντολογίες, ενώ οι τελευταίες εισάγονται όπως και στα άλλα στάδια της προσέγγισης και εκτελούν τα κατάλληλα επίπεδα της προσέγγισης, χωρίς επιπλέον μετασχηματισμό.

Η Εικόνα 5-8 απεικονίζει μερικές από τις κοινές περιοχές όπου το Dozer μπορεί να εισαχθεί σε μια αρχιτεκτονική. Συνήθως, το Dozer χρησιμοποιείται και εφαρμόζεται στα όρια (είσοδος / έξοδος δεδομένων), ενώ αντιστοιχεί τα αντικείμενα του τομέα σε εξωτερικές κλήσεις API και αντίστροφα, διασφαλίζοντας ότι τα αντικείμενα εσωτερικής περιοχής από τη βάση δεδομένων δεν διαρρέουν σε εξωτερικά επίπεδα παρουσίασης ή σε εξωτερικούς χρήστες.

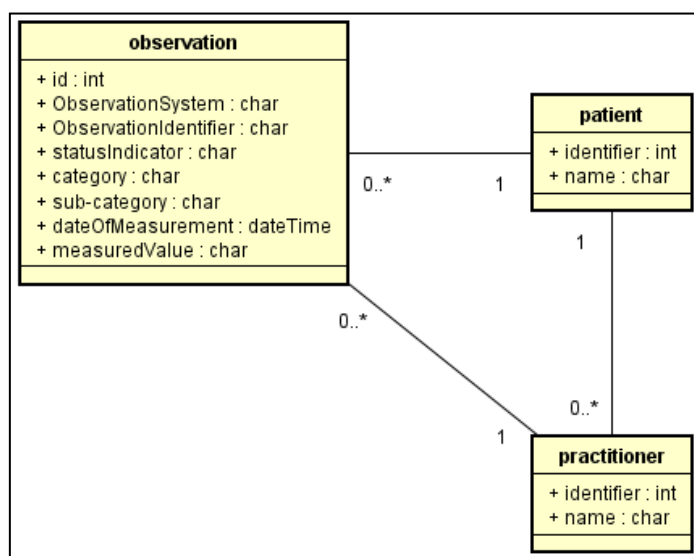


Εικόνα 5-8 – Περιοχές εφαρμογής Apache Dozer

5.2.2.2 Περιγραφή Συνόλου Δεδομένων Πειράματος

Στην περίπτωση του δεύτερου πειράματος, αρχικά ο Dozer έχει δοκιμαστεί μαζί με το σύνολο δεδομένων που περιγράφεται ακολούθως, ενώ τα αποτελέσματά του έχουν συγκριθεί με την παρούσα προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων. Το σύνολο δεδομένων περίπτωσης χρήσης μετασηματίστηκε και αναδιαρθρώθηκε από JSON μορφή σε HL7 FHIR, καλύπτοντας όλα τα προαναφερθέντα βήματα και τεχνικές της προσέγγισης. Το σύνολο δεδομένων που χρησιμοποιήθηκε για την αξιολόγηση της αποτελεσματικότητας της προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων ήταν ένα σύνολο δεδομένων των ζωτικών σημείων - καρδιακός

ρυθμός, ο οποίος μετρήθηκε σε συγκεκριμένη ημερομηνία από έναν ανώνυμο ασθενή, που προέρχεται από τον καταγραφέα δραστηριότητάς του (Misfit Vapor [275]) και αποθηκεύτηκε στην πλατφόρμα της BioAssist [276]. Συνοπτικά, το σύνολο δεδομένων αποτελείται από 238 διαφορετικές μετρήσεις του καρδιακού ρυθμού ενός συγκεκριμένου ασθενούς, που μετράται κάθε 5 λεπτά, την ημερομηνία 2017-02-20. Η Εικόνα 5-9 παρουσιάζει ένα εννοιολογικό διάγραμμα των οντοτήτων που έχουν εντοπιστεί στο σύνολο δεδομένων χρήσης, αποκαλύπτοντας τις σχέσεις που υπήρχαν μεταξύ τους, καθώς και τα εσωτερικά χαρακτηριστικά κάθε οντότητας.



Εικόνα 5-9 - Εννοιολογικό διάγραμμα οντοτήτων συνόλου δεδομένων πρώτου πειράματος

Η Εικόνα 5-10 απεικονίζει ένα στιγμιότυπο του συνόλου δεδομένων που αντιπροσωπεύει τη μέτρηση του καρδιακού ρυθμού του ανώνυμου ασθενούς, την ημερομηνία 2017-02-20T18: 27: 14.000Z, το οποίο είχε τιμή 81 bpm και μετρήθηκε από τον ίδιο τον ασθενή.

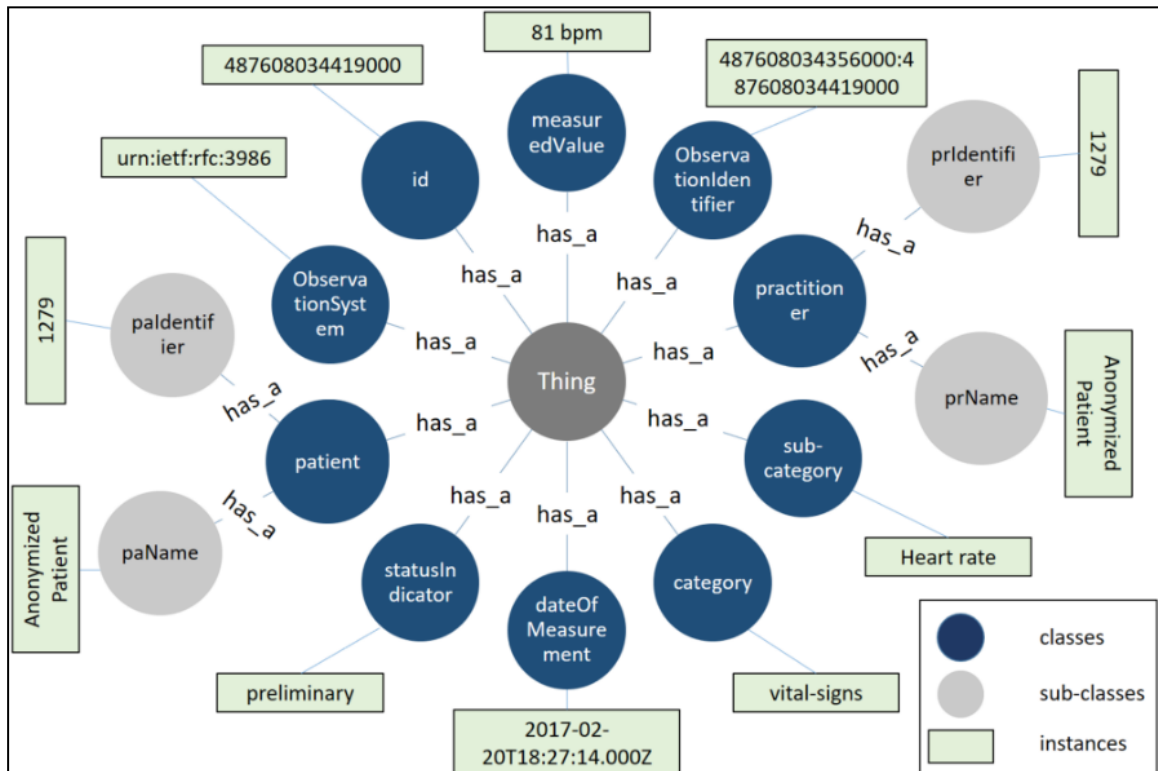
```

31 "observation": [
32   {
33     "id": 487608034419000,
34     "ObservationSystem": "urn:ietf:rfc:3986",
35     "ObservationIdentifier": "487608034356000:487608034419000",
36     "statusIndicator": "preliminary",
37     "category": "vital-signs",
38     "sub-category": "Heart rate",
39     "dateOfMeasurement": "2017-02-20T18:27:14.000Z",
40     "patient": {
41       "identifier": 1279,
42       "name": "Anonymized Patient"
43     },
44     "measuredValue": "81 bpm",
45     "practitioner": {
46       "identifier": 1279,
47       "name": "Anonymized Patient"
48     }
49   },
  
```

Εικόνα 5-10 - Δεδομένα περίπτωσης χρήσης για το δεύτερο πείραμα

5.2.2.3 Αποτελέσματα Πειράματος

Επίπεδο Ontologies & Relationships: Στο εν λόγω βήμα, το ιεραρχικό δέντρο του συνόλου δεδομένων περίπτωσης χρήσης (Εικόνα 5-11) δημιουργήθηκε, όπου αντιπροσωπεύει τις έννοιες (κλάσεις), τις σχέσεις (ιδιότητες) και τα αξιώματά του (στιγμιότυπα). Επιπλέον, οι πόροι HL7 FHIR κατασκευάζονται επίσης ως οντολογίες, λαμβάνοντας ως είσοδο την XML αναπαράσταση κάθε ενός από τους διαφορετικούς πόρους HL7 FHIR.



Εικόνα 5-11 - Οντολογική αναπαράσταση δεδομένων περίπτωσης χρήσης για το δεύτερο πείραμα

Επίπεδο Syntactic Similarity Identifier: Ο Πίνακας 5-5 δημιουργείται απεικονίζοντας τους πόρους HL7 FHIR με υψηλότερους βαθμούς συντακτικής ομοιότητας, σε σύγκριση με τα ονόματα των εννοιών του συνόλου δεδομένων περίπτωσης χρήσης.

Πίνακας 5-5 - Πόροι HL7 FHIR με υψηλότερους βαθμούς συντακτικής ομοιότητας

Use Case Dataset Attribute	HL7 FHIR Resources	Levenshtein Distance
id	<i>Patient.Identifier</i>	0.787 (~79%)
ObservationSystem	<i>Observation</i>	0.683 (~68%)
ObservationIdentifier	<i>Observation.id</i>	0.691 (~69%)
statusIndicator	<i>Patient.contact.relationship</i>	0.360 (36%)
category	<i>Observation.category</i>	0.678 (~68%)
sub-category	<i>Observation.category</i>	0.678 (~68%)
dateOfMeasurement	<i>DiagnosticReport.effectiveDateTime</i>	0.259 (~26%)
patient	<i>Patient</i>	1.0 (100%)
patient.identifier	<i>Patient.gender</i>	0.786 (~79%)
patient.name	<i>Patient.name</i>	1.0 (100%)
measuredValue	<i>Observation.component.value</i>	0.861 (~86%)
practitioner	<i>Observation.performer</i>	0.294 (~29%)
practitioner.identifier	<i>Observation.performer.Practitioner.identifier</i>	0.549 (~55%)
practitioner.name	<i>Observation.performer.Practitioner.name</i>	0.691 (~69%)

Επίπεδο Semantic Similarity Identifier: Ο Πίνακας 5-6 δημιουργείται απεικονίζοντας τους πόρους HL7 FHIR με υψηλότερους βαθμούς σημασιολογικής ομοιότητας, σε σύγκριση με τα ονόματα των εννοιών, των σχέσεων και των αξιωμάτων του συνόλου δεδομένων περίπτωσης χρήσης.

Πίνακας 5-6 - Πόροι HL7 FHIR με υψηλότερους βαθμούς σημασιολογικής ομοιότητας

Use Case Dataset Attribute	HL7 FHIR Resources	Cortical.io
id	<i>Observation.Identifier</i>	0.970 (97%)
ObservationSystem	<i>Observation.code</i>	0.963 (~96%)
ObservationIdentifier	<i>Observation.id</i>	0.891 (~89%)
statusIndicator	<i>Observation.status</i>	0.860 (86%)
category	<i>Observation.category</i>	0.994 (~99%)
sub-category	<i>Observation.code</i>	0.862 (~86%)
dateOfMeasurement	<i>Observation.effectiveDateTime</i>	0.759 (~76%)
patient	<i>Patient</i>	1.0 (100%)
patient.identifier	<i>Patient.identifier</i>	1.0 (100%)
patient.name	<i>Patient.name</i>	1.0 (100%)
measuredValue	<i>Observation.valueQuantity</i>	0.888 (~89%)
practitioner	<i>Observation.performer</i>	0.796 (~80%)
practitioner.identifier	<i>Observation.performer.Practitioner.identifier</i>	0.715 (~72%)
practitioner.name	<i>Observation.performer.Practitioner.name</i>	0.829 (~83%)

Επίπεδο Overall Ontology Mapper: Ο Πίνακας 5-7 δημιουργείται με την απεικόνιση των μεγαλύτερων υπολογισμένων ζευγών των αποτελεσμάτων του Syntactic Similarity και του Semantic Similarity Identifier, παρέχοντας τα χαρακτηριστικά του συνόλου δεδομένων περίπτωσης χρήσης με υψηλότερες πιθανότητες ομοιότητας με τον αναγνωρισμένο πόρο HL7 FHIR.

Πίνακας 5-7 - Μεγαλύτερα υπολογισμένα ζεύγη των αποτελεσμάτων του Syntactic Similarity και του Semantic Similarity Identifier

Use Case Dataset Attribute	Syntactic Similarity Result	Similarity	Semantic Similarity Result	Similarity	Overall Mean
id	<i>Observation. Identifier</i>	0.173 (~17%)	<i>Observation. Identifier</i>	0.970 (97%)	~57%
ObservationSystem	<i>Observation. code</i>	0.243 (~24%)	<i>Observation. code</i>	0.963 (~96%)	~60%
ObservationIdentifier	<i>Observation. id</i>	0.691 (~69%)	<i>Observation. id</i>	0.891 (~89%)	~79%
statusIndicator	<i>Observation. status</i>	0.206 (21%)	<i>Observation. status</i>	0.860 (86%)	~54%
category	<i>Observation. category</i>	0.678 (~68%)	<i>Observation. category</i>	0.994 (~99%)	~84%
sub-category	<i>Observation. code</i>	0.101 (~10%)	<i>Observation. code</i>	0.862 (~86%)	~48%
dateOf Measurement	<i>Observation. effectiveDate Time</i>	0.196 (~20%)	<i>Observation. effectiveDate Time</i>	0.759 (~76%)	~48%
patient	<i>Patient</i>	1.0 (100%)	<i>Patient</i>	1.0 (100%)	100%
patient.identifier	<i>Patient. identifier</i>	0.123 (~12%)	<i>Patient. identifier</i>	1.0 (100%)	~56%
patient.name	<i>Patient. name</i>	1.0 (100%)	<i>Patient. name</i>	1.0 (100%)	100%

measured Value	<i>Observation. value</i>	0.103	<i>Observation. value</i>	0.888	~50%
	<i>Quantity</i>	(~10%)	<i>Quantity</i>	(~89%)	
practitioner	<i>Observation. performer</i>	0.266	<i>Observation. performer</i>	0.796	~49%
	<i>Practitioner. performer.</i>	(~27%)	<i>Practitioner. performer.</i>	(~80%)	
practitioner. identifier	<i>Observation. performer.</i>	0.471	<i>Observation. performer.</i>	0.715	~60%
	<i>Practitioner. identifier</i>		<i>Practitioner. identifier</i>		
practitioner. name	<i>Observation. performer.</i>	0.206	<i>Observation. performer.</i>	0.829	~52%
	<i>Practitioner. name</i>		<i>Practitioner. name</i>		

Αποτελέσματα της Προσέγγισης Dozer

Τα αποτελέσματα της προσέγγισης Dozer είναι παρόμοια με τα αποτελέσματα του Πίνακα 5-7. Για το λόγο αυτό, αποφεύγεται η απεικόνιση της συνολικής διαδικασίας του μετασχηματισμού του Dozer.

5.2.2.4 Συζήτηση Αποτελεσμάτων Πειράματος

Για να συγκριθούν τα παραγόμενα αποτελέσματα με τις προσεγγίσεις μετασχηματισμού δεδομένων, πραγματοποιήθηκε ένας χειροκίνητος μετασχηματισμός του συνόλου δεδομένων των περιπτώσεων χρήσης, όπου οι διαφορετικές μεταβλητές μετασχηματίστηκαν χειροκίνητα σε μορφή HL7 FHIR. Ο Πίνακας 5-8 απεικονίζει τα αποτελέσματα του χειροκίνητου μετασχηματισμού, σε σύγκριση με τα αποτελέσματα της αναπτυγμένης τεχνικής. Είναι σαφές ότι η προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων παρείχε αποτελέσματα ακρίβειας 100%. Ωστόσο, παρατηρείται διαφορά στο συνολικό ποσοστό, το οποίο όμως δεν επηρεάζει την τελική αντιστοίχιση.

Πίνακας 5-8 – Αποτελέσματα του χειροκίνητου μετασχηματισμού

Use Case Dataset Attribute	Manual Results	Similarity	Automatic Results	Similarity	
ObservationSystem	<i>Observation. Identifier</i>	100%	<i>Observation. Identifier</i>	~57%	
	<i>Observation. code</i>	100%	<i>Observation. code</i>	~60%	
	Observation Identifier	<i>Observation. id</i>	100%	<i>Observation. id</i>	~79%
	statusIndicator	<i>Observation. status</i>	100%	<i>Observation. status</i>	~54%
	category	<i>Observation. category</i>	100%	<i>Observation. category</i>	~84%
	sub-category	<i>Observation. code</i>	100%	<i>Observation. code</i>	~48%
	dateOf Measurement	<i>Observation. effectiveDateTime</i>	100%	<i>Observation. effectiveDateTime</i>	~48%
	patient	<i>Patient</i>	100%	<i>Patient</i>	100%
	patient.identifier	<i>Patient.identifier</i>	100%	<i>Patient.identifier</i>	~56%
	patient.name	<i>Patient.name</i>	100%	<i>Patient.name</i>	100%
	measuredValue	<i>Observation. valueQuantity</i>	100%	<i>Observation. valueQuantity</i>	~50%
	practitioner	<i>Observation. performer</i>	100%	<i>Observation. performer</i>	~49%
		<i>Observation. performer.</i>		<i>Observation. performer.</i>	

practitioner. identifier	<i>Practitioner. identifier</i>	100%	<i>Practitioner. identifier</i>	~60%
practitioner. name	<i>Observation. performer. Practitioner.name</i>	100%	<i>Observation. performer. Practitioner.name</i>	~52%

Αναφορικά με τη σύγκριση της προσέγγισης και του Dozer, τα κριτήρια σύγκρισης θεωρήθηκαν ως:

(i) Η ακρίβεια των αποτελεσμάτων (Accuracy) (σε ποσοστό): Το κλάσμα των ανακτηθέντων δεδομένων που σχετίζεται με το ερώτημα, εμφανίζοντας τον αριθμό των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό όλων των αποτελεσμάτων που επιστράφηκαν, όπως αναφέρθηκε προηγουμένως στο (9).

$$accuracy = \frac{|{\text{relevant data}} \cap {\text{retrieved data}}|}{|{\text{retrieved data}}|} \quad (9)$$

(ii) Ανάκληση αποτελεσμάτων (Recall) (σε ποσοστό): Το κλάσμα των σχετικών δεδομένων που ανακτάται επιτυχώς, εμφανίζοντας τον αριθμό των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό των αποτελεσμάτων που πρέπει να επιστραφούν, όπως αναφέρθηκε προηγουμένως στο (10).

$$recall = \frac{|{\text{relevant data}} \cap {\text{retrieved data}}|}{|{\text{relevant data}}|} \quad (10)$$

(iii) Το μέτρο F (F): Ο αρμονικός μέσος όρος (δηλαδή ο συνδυασμός) ακρίβειας και ανάκλησης, όπως αναφέρθηκε προηγουμένως στο (13).

$$F = \frac{2 * precision * recall}{precision + recall} \quad (13)$$

(iv) Η διάρκεια (Latency) του μετασχηματισμού (σε sec).

(v) Τα σφάλματα (Errors) που συναντήθηκαν (σε αριθμούς).

(vi) Η συνολική πολυπλοκότητα (Complexity) (υψηλή (high), μέτρια (medium), χαμηλή (low)).

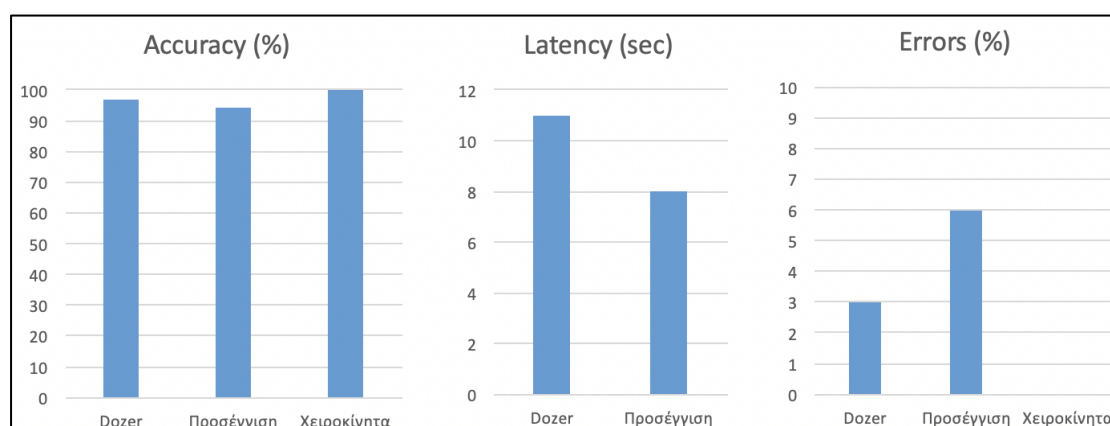
Θα πρέπει να διευκρινιστεί ότι οι μετρήσεις της ακρίβειας, της ανάκλησης και του F έχουν υπολογιστεί σε σχέση με την τα πρότυπα του UMLS [277]. Τα αποτελέσματα μετασχηματισμού του καταγεγραμμένου συνόλου δεδομένων συνοψίζονται στον Πίνακας 5-9. Για τα χειροκίνητα αποτελέσματα, θεωρείται

ότι έχουν υψηλή ακρίβεια, τηρώντας τις κατευθυντήριες γραμμές της μορφής HL7 FHIR, που θεωρείται σημείο αναφοράς υψηλής ποιότητας και ακρίβειας.

Πίνακας 5-9 - Κριτήρια και αποτελέσματα σύγκρισης προσεγγίσεων

Προσέγγιση	Κριτήρια					
	Accuracy (%)	Recall (%)	F (%)	Latency (sec)	Errors (%)	Complexity
Dozer	97	93	95	11	3	Low
Προσέγγιση	94	92	93	8	6	High
Χειροκίνητα	100	100	100	~7200	0	High

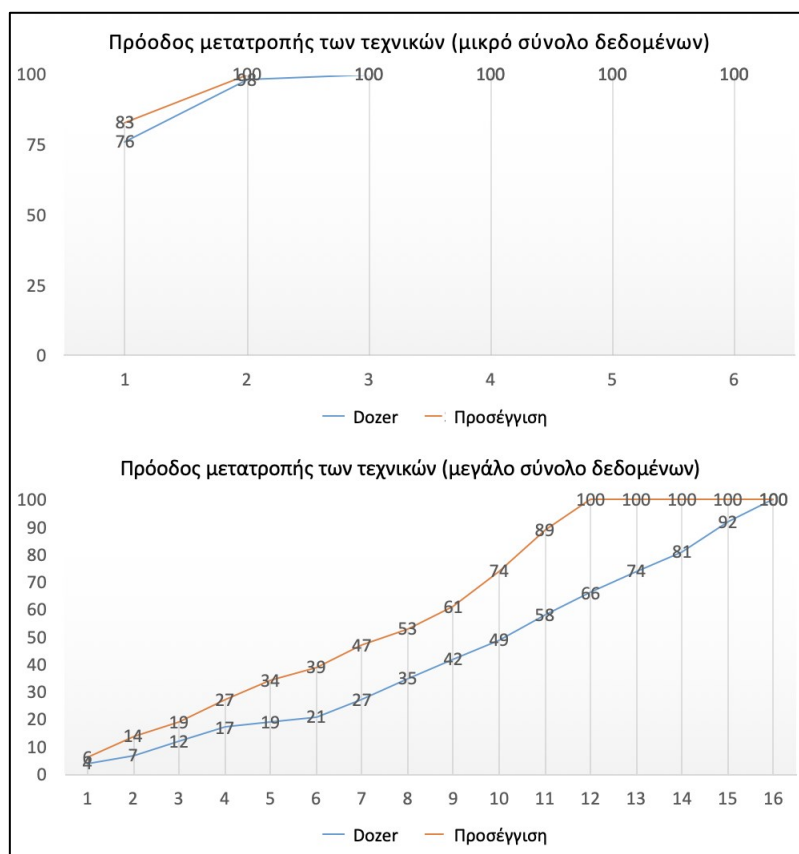
Στην Εικόνα 5-12 απεικονίζεται μια απεικόνιση των προαναφερθέντων αποτελεσμάτων μέσω ενός ραβδογράμματος, σύμφωνα με τα επιλεγμένα κριτήρια.



Εικόνα 5-12 - Απεικόνιση κριτηρίων και αποτελεσμάτων σύγκρισης

Πρέπει να αναφερθεί ότι όσο πιο κοντά το μέτρο F είναι στο 1, τόσο καλύτερα είναι τα αποτελέσματα μετασχηματισμού. Για το λόγο αυτό, όπως φαίνεται στον Πίνακα 5-9, το Μέτρο F του χειροκίνητου μετασχηματισμού είναι 1 (100%), επικυρώνοντας την υψηλή ακρίβεια των αποτελεσμάτων του χειροκίνητου μετασχηματισμού. Μέσω του Πίνακα 5-9 και της Εικόνα 5-12, είναι σαφές ότι, ανάλογα με τις απαιτήσεις των ενδιαφερομένων, οι δύο διαφορετικές προσεγγίσεις πληρούν διαφορετικά κριτήρια. Ωστόσο, απεικονίζεται ότι η εναλλαγή μεταξύ ακρίβειας και διάρκειας είναι τέτοια που επιτρέπει τη χρησιμοποίηση της προτεινόμενης προσέγγισης σε σενάρια όπου προκύπτουν ροές δεδομένων πραγματικού χρόνου για την υγειονομική περίθαλψη και επομένως η εκμετάλλευσή τους είναι κρίσιμη σε πραγματικό χρόνο. Αυτό φαίνεται από το γεγονός όπου ενώ η απόλυτη διαφορά της

μέτρησης F μεταξύ της προτεινόμενης προσέγγισης και της προσέγγισης του Dozer είναι 2%, η συνολική διαφορά χρόνου είναι 27%. Από τη στιγμή που παρατηρήθηκε η διαφορά μεταξύ των παραπάνω μετασχηματιστικών προσεγγίσεων όσον αφορά τον συνολικό χρόνο εκτέλεσης, θέλαμε να παρατηρήσουμε πώς συμπεριφέρονται αυτές οι προσεγγίσεις σε σύνολα δεδομένων διαφορετικού μεγέθους. Επομένως, για να δοθεί ένα επιπλέον κριτήριο σύγκρισης όσον αφορά τις δύο διαφορετικές προσεγγίσεις μετασχηματισμού, επαναλήφθηκε το ίδιο πείραμα με διαφορετικά σύνολα δεδομένων. Με περισσότερες λεπτομέρειες, το ίδιο πείραμα πραγματοποιήθηκε για ένα σύνολο δεδομένων 12 Kbytes (KB) με λιγότερες μετρήσεις (54 μετρήσεις - δηλαδή μικρό σύνολο δεδομένων), ενώ επαναλήφθηκε και για ένα σύνολο δεδομένων 2.032 KB, που περιέχει μεγαλύτερο αριθμό μετρήσεων (2.973 μετρήσεις - δηλαδή μεγάλο σύνολο δεδομένων). Ο λόγος που πραγματοποιήθηκαν αυτά τα πρόσθετα πειράματα ήταν να προσδιοριστεί ο τρόπος συμπεριφοράς αυτών των δύο προσεγγίσεων μετασχηματισμού σε συγκεκριμένες χρονικές σφραγίδες, με βάση το μέγεθος του συνόλου δεδομένων, σύμφωνα με τη συνολική πρόοδο αντιστοίχισης. Τα αποτελέσματα αυτών των πειραμάτων καταγράφονται στον Πίνακα 5-10 και Πίνακα 5-11, όπου απεικονίζουν το συνολικό χρόνο που απαιτείται και για τις δύο περιπτώσεις για να λάβει χώρα η αντιστοίχιση.



Εικόνα 5-13 – Απεικόνιση πρόοδου μετατροπής μικρού και μεγάλου συνόλου δεδομένων

Θα πρέπει να σημειωθεί ότι αυτά τα αποτελέσματα απεικονίζονται μέσω ενός γραφήματος στην Εικόνα 5-13. Μέσω του Πίνακας 5-10, του Πίνακας 5-11 και της Εικόνα 5-13, μπορούν να επιβεβαιωθούν τα αποτελέσματα του πειράματος, καθώς και τα συμπεράσματα που προκύπτουν από αυτά.

Περίπτωση του μικρού συνόλου δεδομένων

Μέσω της ανάλυσης του γραφήματος που απεικονίζει την πρόοδο αντιστοίχισης στο μικρό σύνολο δεδομένων (Πίνακας 5-10), δεν είναι δυνατόν να προσδιοριστεί η πρόοδος των δύο προσεγγίσεων, καθώς και οι δύο εκτελούνται σχεδόν με τον ίδιο τρόπο.

Πίνακας 5-10 - Σύγκριση αποτελεσμάτων περίπτωσης μικρού συνόλου δεδομένων

Χρονική σφραγίδα (sec)	Πρόοδος - μικρό σύνολο δεδομένων (%)	
	Dozer	Προσέγγιση
1	76	83
2	98	100
3	100	100
4	100	100
5	100	100
6	100	100

Περίπτωση του μεγάλου συνόλου δεδομένων

Μέσω της ανάλυσης του μεγάλου συνόλου δεδομένων (Πίνακας 5-11), καθίσταται σαφές ότι η προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων έχει μετασηματίσει σχεδόν το 39% του τρέχοντος συνόλου δεδομένων σε μορφή HL7 FHIR σε 6 δευτερόλεπτα, ενώ ταυτόχρονα η Dozer προσέγγιση μετασηματίσε το 21% του τρέχοντος συνόλου δεδομένων. Πρέπει να σημειωθεί ότι στην περίπτωση αυτή, η απόλυτη διαφορά της συνολικής πρόοδου μεταξύ της προτεινόμενης προσέγγισης και του Dozer σε 6 δευτερόλεπτα είναι 18%, πράγμα που σημαίνει ότι η προτεινόμενη προσέγγιση έχει αναλύσει και επεξεργαστεί σχεδόν το διπλάσιο ποσό δεδομένων. Παρ' όλα αυτά, η προσέγγιση του Dozer απαιτούσε συνολικά 16 δευτερόλεπτα για την εκτέλεση του μετασηματισμού, ενώ η προτεινόμενη προσέγγιση απαιτούσε 12 δευτερόλεπτα για να παρέχει τα ίδια αποτελέσματα. Όπως αναφέρθηκε

προηγουμένως, είναι αναμφισβήτητο ότι η προτεινόμενη προσέγγιση θα πρέπει να χρησιμοποιείται για ροές δεδομένων σε πραγματικό χρόνο και επομένως η εκμετάλλευσή τους είναι κρίσιμη σε πραγματικό χρόνο.

Πίνακας 5-11 - Σύγκριση αποτελεσμάτων περίπτωσης μεγάλου συνόλου δεδομένων

Χρονική σφραγίδα (sec)	Πρόοδος - μικρό σύνολο δεδομένων (%)	
	Dozer	Προσέγγιση
1	4	6
2	7	14
3	12	19
4	17	27
5	19	34
6	21	39
7	27	47
8	35	53
9	42	61
10	49	74
11	58	89
12	66	100
13	74	100
14	81	100
15	92	100
16	100	100

ΚΕΦΑΛΑΙΟ 6

Σύνοψη Διατριβής - Ανοιχτά Ερευνητικά Θέματα

Περιεχόμενα

6.1. Σύνοψη Διατριβής και Κυριότερων Συνεισφορών

6.1.1. Εφαρμογή και Χρήση της Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων

6.2. Ανοιχτά Ερευνητικά Θέματα και Μελλοντικά Σχέδια

Περίληψη

Στο ΚΕΦΑΛΑΙΟ 6, πραγματοποιείται η σύνοψη της Διδακτορικής Διατριβής, και των βασικών συνεισφορών της, ενώ περιγράφονται τα ανοιχτά ερευνητικά θέματα και οι μελλοντικοί στόχοι, όπως προέκυψαν από την παρούσα προσέγγιση της έρευνας, αλλά και από το χώρο της υγείας γενικότερα, αναφορικά τόσο με τον τομέα της διαλειτουργικότητας, αλλά και τον τομέα της ένταξης και χρήσης των οντολογιών στο χώρο της υγείας.

6.1 Σύνοψη Διατριβής και Κυριότερων Συνεισφορών

Την τελευταία δεκαετία, υπήρξε μια μετάβαση από έναν φτωχό σε δεδομένα σε έναν πλούσιο σε δεδομένα κόσμο. Μεγάλο μέρος αυτής της άνευ προηγουμένου αύξησης της παραγωγής δεδομένων μπορεί να αποδοθεί στην αφθονία χιλιάδων κινητών συσκευών, φορητών συσκευών και αισθητήρων που εμπίπτουν στην εποχή του διαδικτύου. Σύμφωνα με την Machina Research [278], ο συνολικός αριθμός των συνδεδεμένων συσκευών αναμένεται να είναι 27 δισ. μέχρι το 2024, ενώ σύμφωνα με την Cisco [279], μέχρι το 2020 θα ανήκουν σχεδόν 1.5 κινητές συσκευές ανά κάτοικο και περισσότερα από 601 εκατομμύρια φορητές συσκευές θα είναι σε χρήση. Αυτοί οι αριθμοί επιβεβαιώνουν μόνο ότι το IoT παραμένει μία από τις μεγαλύτερες τάσεις το 2019. Τέτοιο πράγμα έχει ως αποτέλεσμα μια μυριάδα ετερογενών συσκευών που θα συνδεθούν με τον κόσμο του διαδικτύου στο εγγύς μέλλον, παράγοντας δεδομένα διαφορετικών τύπων που μπορεί να έχουν συλλεχθεί σε διαφορετικά χρονοδιαγράμματα, από διαφορετικές συσκευές. Οι περισσότερες από αυτές τις συσκευές τυπικά χαρακτηρίζονται από υψηλό βαθμό ετερογένειας, όσον αφορά την ύπαρξη διαφορετικών προδιαγραφών, δυνατοτήτων ή προδιαγραφών δικτύου, οι οποίες πρέπει να είναι εύκολα διαχειρίσιμες από απλές πλατφόρμες και συστήματα IoT. Ως εκ τούτου, η ενσωμάτωση των συσκευών είναι ένα από τα πιο απαιτητικά ερευνητικά θέματα στην εποχή του IoT, αποκαλύπτοντας την ανάγκη ύπαρξης μιας διεθνούς διεπαφής που θα είναι σε θέση να διαχειριστεί όλες αυτές τις ετερογενείς συσκευές. Ειδικά για τις συσκευές Internet of Medical Things (IoMT) [280] που υιοθετούνται ευρέως και χρησιμοποιούνται στο Healthcare 4.0, η ανάγκη ενσωμάτωσης αυτών των συσκευών είναι ένα ζήτημα ζωτικής σημασίας. Για παράδειγμα, η τηλεπαρακολούθηση των ασθενών στο σπίτι απαιτεί μια μεγάλη ποικιλία αισθητήρων και συσκευών που συνήθως παρουσιάζουν ένα σύνολο πολύ συγκεκριμένων χαρακτηριστικών και έτσι είναι δύσκολο να ενσωματωθούν σε διαφορετικές πλατφόρμες για να προσφέρουν τα ιατρικά δεδομένα κάθε ασθενούς. Χωρίς την αποτελεσματική ενσωμάτωση των δεδομένων των ετερογενών ιατρικών συσκευών, δε θα συλλέγονται δεδομένα ασθενών και συνεπώς δεν είναι δυνατόν να γίνουν άλλες ενέργειες, καθώς δεν επιτυγχάνεται τελικά ολοκληρωμένη περίθαλψη.

Φαίνεται πως μεγάλα ποσά ετερογενών ιατρικών δεδομένων καθίστανται διαθέσιμα σε διάφορους οργανισμούς και αισθητήρες υγειονομικής περίθαλψης, αναμορφώνοντας έτσι πλήρως το Healthcare 4.0. Εκατοντάδες νοσοκομεία και συστήματα υγειονομικής περίθαλψης ασχολούνται καθημερινά με προκλήσεις στην εξαγωγή δεδομένων από περιορισμένες πηγές

δεδομένων, που επηρεάζουν τόσο την περίθαλψη των ασθενών όσο και την ιατρική έρευνα. Επί του παρόντος, η ταχέως αυξανόμενη διαθεσιμότητα των ιατρικών αρχείων ωθεί στην υιοθέτηση προσεγγίσεων που βασίζονται σε δεδομένα, προσφέροντας ευκαιρίες για αυτοματοποίηση των καθηκόντων που σχετίζονται με την υγειονομική περίθαλψη, παρέχοντας καλύτερη ανίχνευση ασθενειών, ακριβέστερη πρόγνωση, ταχύτερη κλινική έρευνα και καλύτερη προσαρμογή στη διαχείριση ασθενών. Παρόλα αυτά, οι οργανισμοί υγειονομικής περίθαλψης εξακολουθούν να αντιμετωπίζουν πολλές δυσκολίες στην εφαρμογή, τη συντήρηση και την αναβάθμιση των συστημάτων υγειονομικής περίθαλψης, συμπεριλαμβανομένων πολλών προκλήσεων στους τομείς της τεχνικής, της ασφάλειας και της ανθρώπινης αλληλεπίδρασης. Η ετερογένεια, η επικαιρότητα, η πολυπλοκότητα, ο θόρυβος και η ατέλεια συγκαταλέγονται στις πιο κοινές προκλήσεις της δημιουργίας αξίας από τα δεδομένα. Μεταξύ άλλων, αυτό που λείπει είναι ένα ολοκληρωμένο σύστημα ανταλλαγής δεδομένων το οποίο σύμφωνα με το [281] θα μπορούσε να μειώσει τις δαπάνες της υγειονομικής περίθαλψης κατά περισσότερο από 3 δισ. δολάρια κάθε χρόνο. Προκειμένου να ανταλλάσσονται δεδομένα με όσο το δυνατόν περισσότερους ενδιαφερόμενους, η διαλειτουργικότητα είναι ο μόνος τρόπος για να δοθεί η δυνατότητα στα συστήματα να αλληλοεπιδρούν μεταξύ τους, προκειμένου να αποκτήσουν την πλήρη εικόνα ενός ασθενούς και να θεωρηθούν αναγκαία στα ηλεκτρονικά συστήματα υγειονομικής περίθαλψης για την επίλυση των ετερογενών δεδομένων. Μια πρόσφατη μελέτη [282] εκτιμά ότι η παγκόσμια αγορά ενδιαμέσου λογισμικού για την υγειονομική περίθαλψη αναμένεται να φθάσει τα 3.07 δισ. δολάρια μέχρι το 2023 από 1.90 δισ. δολάρια το 2018, προκειμένου να ξεπεραστούν τα ζητήματα διαλειτουργικότητας της υγειονομικής περίθαλψης. Μεταξύ των διαφόρων υφιστάμενων ερευνών και προτύπων, ο οργανισμός HL7 παρέχει την ανάπτυξη και το πλαίσιο των προτύπων που χρησιμοποιούνται ευρέως στην ιατρική αγορά και την έρευνα, με το HL7 FHIR να είναι το πιο πρόσφατο πρότυπο που δημιουργήθηκε από την οργάνωση, για την ανταλλαγή κλινικών πληροφοριών.

Η παρούσα έρευνα αφορούσε σε αρχικό στάδιο τη μελέτη της διαλειτουργικότητας δεδομένων σε γενικό πλαίσιο, εκφράζοντας την ανάγκη της στο χώρο της υγείας. Ακολούθως αναφέρθηκαν οι σχετικές προκλήσεις, σε συνδυασμό με τους τομείς που συναντάται η έννοια της διαλειτουργικότητας των δεδομένων, επισημαίνοντας το όραμα του μέλλοντος για έναν κόσμο με διαλειτουργικές υπηρεσίες. Κατόπιν πραγματοποιήθηκε ανασκόπηση της σχετικής βιβλιογραφικής έρευνας που έλαβε χώρα σε τομείς που χρησιμοποιούνται στο χώρο της διαλειτουργικότητας, αλλά και της τελικής

προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων. Για το λόγο αυτό, έγινε αναφορά αρχικά στο χώρο της διαλειτουργικότητας δεδομένων, ενώ περιεγράφηκε το φαινόμενο των μεγάλων δεδομένων στην καθημερινότητά μας. Σε δεύτερο στάδιο, έγινε αναφορά στη διαλειτουργικότητα στο χώρο της υγείας, όπου επισημάνθηκε η ανάγκη ύπαρξής της, οι σχετικές κατηγορίες της, ορισμένα από τα πιο γνωστά πρότυπα υγείας για τη διαλειτουργική μεταφορά δεδομένων και κωδικοποιήσεων, ενώ έγινε εκτενής μελέτη σε αρκετές έρευνες. Παράλληλα, αναλύθηκαν οι οντολογίες, η ανάγκη τους και η χρήση τους στο χώρο της υγείας, ενώ επισημάνθηκαν οι τεχνικές οι οποίες στοχεύουν στη δημιουργία οντολογιών. Ακολούθως, έγινε αναφορά στον τομέα της επεξεργασίας φυσικής γλώσσας, αναφέροντας τη σημασία των τεχνικών συντακτικής και σημασιολογικής ομοιότητας, μεταξύ λεκτικών και κειμένων. Στη συνέχεια, αναλύθηκε η παρούσα προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων, όπου σε αρχικό στάδιο έγινε μια λεπτομερής αναφορά στις πιο κομβικές έρευνες που πραγματοποιήθηκαν στο πλαίσιο της έρευνας, σε συνδυασμό με τα αποτελέσματά τους. Ακολούθως, περιεγράφηκε κάθε τεχνική της παρούσας προσέγγισης, αναφορικά με τη δημιουργία των οντολογιών, την αποθήκευσή τους, τη συντακτική και σημασιολογική αντιστοιχισή τους, καθώς και την τελική απεικόνισή τους στο πρότυπο υγείας HL7 FHIR. Στο επόμενο στάδιο, έλαβε μέρος η αξιολόγηση της προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων, μέσω δύο διαφορετικών περιπτώσεων χρήσης. Η πρώτη περίπτωση χρήσης σχετίστηκε με ένα σύνολο δεδομένων υγείας από ανώνυμους ασθενείς ενός Ινστιτούτου Υγείας, ενώ η δεύτερη περίπτωση χρήσης σχετίστηκε πάλι με ένα σύνολο δεδομένων από έναν ανώνυμο ασθενή ο οποίος προσέφερε τα δεδομένα του, όπως προήλθαν από μια φορητή συσκευή. Παράλληλα με τη δεύτερη περίπτωση, πραγματοποιήθηκε και αναλύθηκε μία συγκριτική αξιολόγηση της παρούσας προσέγγισης με μία άλλη προσέγγιση μετασχηματισμού και αντιστοιχισής δεδομένων. Σε τελικό στάδιο, πραγματοποιήθηκε η σύνοψη της Διδακτορικής Διατριβής, και των βασικών συνεισφορών της, ενώ περιεγράφηκαν τα ανοιχτά ερευνητικά θέματα και οι μελλοντικοί στόχοι όπως προέκυψαν από την παρούσα προσέγγιση της έρευνας. Ο στόχος της έρευνας, όπως έχει αναφερθεί, ήταν η δημιουργία μιας προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων σε μορφή ενός επιπέδου μετασχηματισμού ανομοιογενών δεδομένων υγείας μέσω ενός αυτοματοποιημένου τρόπου ο οποίος μετασχηματίζει τόσο σε σημασιολογικό όσο και σε συντακτικό επίπεδο, δεδομένα υγείας διαφορετικής αναπαράστασης και μορφής, προκειμένου να ταυτοποιηθούν, να αντιστοιχηθούν σε μία ενιαία γλώσσα (HL7 FHIR), και στο τέλος να συγχωνευτούν σε ένα κοινό επίπεδο.

6.1.1 Εφαρμογή και Χρήση της Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων

Αξιζει να αναφερθεί πως η συγκεκριμένη προσέγγιση αποτελεί μια γενικευμένη ολιστική λύση διαλειτουργικότητας δεδομένων, η οποία χάρις στην ευελιξία της και την ελαστικότητά της, μπορεί να εφαρμοστεί σε πολλαπλούς τομείς, καλύπτοντας πολλαπλές ανάγκες. Ειδικότερα:

- Στο χώρο της **Υγειονομική περίθαλψης** η παρούσα προσέγγιση μπορεί να χρησιμοποιηθεί για τη βελτιστοποίηση και την τυποποίηση της ποιότητας της ιατρικής περίθαλψης. Οι πολλαπλές πηγές πληροφοριών, όπως δεδομένα από εργαστήρια, κλινικές, φαρμακεία, νοσοκομεία και φορείς παροχής πρωτοβάθμιας περίθαλψης, που χρησιμοποιούν πολλαπλά συστήματα για την καταγραφή αυτών των πληροφοριών, μπορούν να μετασχηματιστούν αυτοματοποιημένα σε μία κοινή γλώσσα και δομή, διαθέτοντας τους ίδιους κανόνες αναπαράστασης. Ως εκ τούτου, θα επιτυγχάνεται η σχετική ανταλλαγή EHR και η παροχή ολοκληρωμένης υγειονομικής περίθαλψης στους ασθενείς.
- Στο χώρο των **Τηλεπικοινωνιών και των Δικτύων** ήδη έχει γίνει μία πρώτη προσπάθεια στο πλαίσιο της παρούσας έρευνας για την ένταξη της προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων. Ειδικότερα, λαμβάνοντας υπόψιν τις ανάγκες και τα οφέλη των δικτύων 5^{ης} γενιάς (5th Generation - 5G), στην έρευνα [283] δημιουργήθηκε μια πλατφόρμα, όπου μέσω Wireless-Fidelity (Wi-Fi) συνδέονται συσκευές IoMT που μπορούν να εξυπηρετήσουν διαφορετικά σενάρια εφαρμογών (π.χ. παρακολούθηση ζωτικών σημείων, παρακολούθηση από το σπίτι). Μέσω του προτεινόμενου μηχανισμού λήψης δεδομένων, αποκτώνται τα δεδομένα των συσκευών IoMT, καθώς και οι προδιαγραφές δικτύου τους, προκειμένου να διαχωριστούν σε διαφορετικά τμήματα δικτύου 5G (5G network slices) μέσω του μηχανισμού διαχείρισης τμημάτων, το καθένα από το οποίο εξυπηρετεί τις απαιτήσεις υπολογισμού διαφόρων ιατρικών σεναρίων. Τα τρία ιατρικά σενάρια που εξετάζονται για την κατανομή των συσκευών IoMT σε τρία διαφορετικά τμήματα, τα οποία εμπεριέχουν διαφορετικές απαιτήσεις υπολογισμού είναι: α) Μαζικό σενάριο IoMT το οποίο περιλαμβάνει απαιτήσεις για στιγμιαία καταγραφή, συλλογή και λήψη ιατρικών δεδομένων ασθενών, σωματική δραστηριότητα και ζωτικά σημεία, β) Ενισχυμένο σενάριο κινητής ευρυζωνικότητας που απαιτεί βελτιωμένα ποσοστά δεδομένων και υψηλή συνδεσιμότητα για απομακρυσμένη ιατρική εξέταση και γ) Σενάριο υπηρεσιών κρίσιμης σημασίας που απαιτεί ενισχυμένη αξιοπιστία δικτύου για τους παρόχους υγειονομικής

περίθαλψης σε κρίσιμες περιπτώσεις. Στη συνέχεια, εκτελείται η παρούσα προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων σε κάθε διαφορετικό τμήμα δικτύου για την αντιμετώπιση της ετερογένειας δεδομένων και παρέχει τα δεδομένα που ανήκουν στα ίδια τμήματα σε ένα κοινό και ερμηνεύσιμο πρότυπο για περαιτέρω ανάλυση. Σαν αποτέλεσμα, μπορεί να εμφανιστεί ξεκάθαρα η εφαρμογή της παρούσας προσέγγισης στο χώρο των Τηλεπικοινωνιών και των Δικτύων, η οποία μπορεί να επεκταθεί και για άλλα ιατρικά σενάρια, με πολλαπλές απαιτήσεις δικτύου.

- Στο χώρο της **Αναγνώρισης Συσκευών** ήδη έχει γίνει μία πρώτη προσπάθεια στο πλαίσιο της παρούσας έρευνας για την ένταξη της προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων. Ειδικότερα, στην έρευνα [284] έχει εφαρμοστεί η τεχνική συντακτικής και σημασιολογικής αντιστοίχισης της Υπηρεσιών Διαλειτουργικότητας Δεδομένων, στα API των εταιριών που κατασκευάζουν φορητές συσκευές και συσκευές παρακολούθησης ζωτικών σημάτων (π.χ. Garmin, Fitbit), προκειμένου να εντοπιστεί η ομοιότητα μεταξύ τους και εν τέλει να γίνει κατανοητή η φύση του κάθε ξεχωριστού API, τόσο σε συντακτικό όσο και σε σημασιολογικό επίπεδο. Με τον τρόπο αυτό, γίνεται αφενός η αναγνώριση άγνωστων συσκευών λόγω της μερικής σύγκρισής τους με ήδη αναγνωρισμένες συσκευές, ενώ παράλληλα διευκολύνεται η εύρεση των API μεθόδων που είναι υπεύθυνες για τη λήψη των δεδομένων από κάθε συσκευή. Αποτέλεσμα όλης αυτής της τεχνικής, είναι η εφαρμογή μέρους της προσέγγισης Υπηρεσιών Διαλειτουργικότητας Δεδομένων, σε ένα διαφορετικό χώρο και τομέα.
- Στον υπόλοιπο ερευνητικό χώρο, όπως της **Αναζήτησης, της Κυβέρνησης, ή των Ηλεκτρονικών Προμηθειών** η παρούσα προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων μπορεί να εφαρμοστεί τόσο για την αναζήτηση πληροφοριών όσο και για την ανταλλαγή κυβερνητικών εγγράφων και δεδομένων ή την πραγματοποίηση ψηφιακών και ηλεκτρονικών προμηθειών. Ειδικότερα, λαμβάνοντας υπόψιν πως η προσέγγιση αυτή χρησιμοποιεί οντολογίες που προέρχονται αφενός από το σύνολο των εκάστοτε δεδομένων, αλλά και από το σύνολο του προτύπου υγείας HL7 FHIR για να γίνει η σχετική σύγκριση, είναι ιδιαίτερα εύκολο η ίδια σύγκριση να πραγματοποιηθεί με οποιοδήποτε οντολογίες που παρέχονται από οποιοδήποτε άλλο πρότυπο, διαφορετικού χώρου από αυτόν της υγείας. Για παράδειγμα, στο χώρο των Ηλεκτρονικών Προμηθειών μπορεί να χρησιμοποιηθεί το πρότυπο ISO 15000 σαν πρότυπο σύγκρισης και αντιστοίχισης, το οποίο εφόσον μετασχηματιστεί σε

οντολογική μορφή να αντιστοιχηθεί με τα τρέχοντα δεδομένα (στην οντολογική τους μορφή) τα οποία πρέπει να μεταφερθούν και να μετασχηματιστούν. Ομοίως μπορεί να επαναληφθεί και για οποιονδήποτε άλλο τομέα, όπου απλά θα προσαρμόζεται κάθε φορά το πρότυπο αντιστοίχισης και τελικού μετασχηματισμού. Για αυτό άλλωστε επιλέχτηκε και η χρήση των οντολογιών, ώστε να καταλήξουμε σε μία γενικευμένη ολιστική προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων.

6.2 Ανοιχτά Ερευνητικά Θέματα και Μελλοντικά Σχέδια

Ένα διαλειτουργικό οικοσύστημα πληροφορικής για την υγεία κάνει τα σωστά δεδομένα διαθέσιμα στους κατάλληλους ανθρώπους την κατάλληλη στιγμή σε όλα τα προϊόντα και τους οργανισμούς κατά τρόπο που να μπορεί να βασιστεί και να χρησιμοποιηθεί με αξιοπιστία από τους παραλήπτες. Μέχρι το 2024 [285], τα άτομα, οι πάροχοι υπηρεσιών υγείας, οι κοινότητες και οι ερευνητές θα πρέπει να διαθέτουν μια σειρά από διαλειτουργικά προϊόντα και υπηρεσίες πληροφορικής για την υγεία, που επιτρέπουν στο σύστημα υγειονομικής περίθαλψης να μαθαίνει συνεχώς και να προωθεί το στόχο της βελτίωσης της υγειονομικής περίθαλψης. Αυτό το «εκπαιδευτικό σύστημα υγείας» θα πρέπει επίσης να επιτρέπει χαμηλότερα κόστη υγειονομικής περίθαλψης, βελτιωμένη υγεία του πληθυσμού, πραγματική ενδυνάμωση των καταναλωτών και προώθηση της καινοτομίας. Για παράδειγμα, όλα τα άτομα, οι οικογένειές τους και οι πάροχοι υπηρεσιών υγείας θα πρέπει να μπορούν να στέλνουν, να λαμβάνουν, να βρίσκουν και να χρησιμοποιούν πληροφορίες για την υγεία με κατάλληλο, ασφαλές, έγκαιρο και αξιόπιστο τρόπο. Τα άτομα πρέπει να μπορούν να μοιράζονται με ασφάλεια τα δεδομένα τους, να ενημερώνονται με τους παρόχους περίθαλψης και να χρησιμοποιούν τις πληροφορίες για να στηρίξουν τη δική τους υγεία και ευεξία μέσω ενημερωμένων από κοινού λήψης αποφάσεων. Ένα διαλειτουργικό οικοσύστημα πληροφορικής για την υγεία θα πρέπει να υποστηρίζει τις κρίσιμες λειτουργίες δημόσιας υγείας, όπως η επιτήρηση της ασθένειας σε πραγματικό χρόνο και η αντιμετώπιση καταστροφών, καθώς και η συγκέντρωση δεδομένων για έρευνα με στόχο την υψηλότερη ποιότητα φροντίδας, και όχι απαραίτητα μεγαλύτερη ποσότητα φροντίδας.

Καθώς εργαζόμαστε προς αυτό το όραμα για το μελλοντικό διαλειτουργικό οικοσύστημα ΤΠΕ για την υγεία, θα σχεδιάσουμε και θα εκτελέσουμε το έργο μας για να ευθυγραμμιστούμε με ένα σύνολο κατευθυντήριων αρχών:

- **Οικοδόμηση της υπάρχουσας υποδομής πληροφορικής για την υγεία.** Έχουν πραγματοποιηθεί σημαντικές επενδύσεις στις ΤΠΕ για την Υγεία μέσω του συστήματος παροχής φροντίδας και σε άλλους συναφείς τομείς που πρέπει να ανταλλάσσουν πληροφορίες με τα άτομα και τους παρόχους περίθαλψης. Στο μέτρο του δυνατού, θα ενθαρρύνουμε τους ενδιαφερόμενους να χτίσουν από την υπάρχουσα υποδομή πληροφορικής για την υγεία, αυξάνοντας τη διαλειτουργικότητα και τη λειτουργικότητα, όπως απαιτείται.

- **Ένα μέγεθος δεν ταιριάζει σε όλα.** Η διαλειτουργικότητα απαιτεί τεχνική και πολιτική συμμόρφωση μεταξύ δικτύων, τεχνικών συστημάτων και των συνιστωσών τους. Απαιτεί επίσης αλλαγή συμπεριφοράς και πολιτισμού εκ μέρους των χρηστών. Θα προσπαθήσουμε για τη διαλειτουργικότητα της γραμμής βάσης σε όλη την υποδομή πληροφορικής για την υγεία, επιτρέποντας ταυτόχρονα στους καινοτόμους και τους τεχνολόγους να διαφοροποιούν την εμπειρία των χρηστών (αίσθηση και λειτουργία των εργαλείων) προκειμένου να ανταποκρίνονται καλύτερα στις ανάγκες του χρήστη, στις προσωπικές τους προτιμήσεις και σε άλλους παράγοντες.
- **Ενδυνάμωση ατόμων.** Οι πολίτες υιοθετούν ταχέως την τεχνολογία για τη διαχείριση πολλών πτυχών της ζωής τους, συμπεριλαμβανομένης της υγείας και της ευεξίας. Ωστόσο, πολλά από αυτά τα εργαλεία δεν ενσωματώνουν ακόμη πληροφορίες από το σύστημα παροχής υγειονομικής περίθαλψης. Οι πληροφορίες για την υγεία από το σύστημα παροχής φροντίδας θα πρέπει να είναι εύκολα προσβάσιμες από τους ιδιώτες και να τους δίνουν τη δυνατότητα να γίνουν πιο δραστήριοι στην υγεία τους, όπως άλλοι τύποι δεδομένων τους εξουσιοδοτούν σε άλλες πτυχές της ζωής τους.
- **Επένδυση στην αγορά.** Η ζήτηση διαλειτουργικότητας από χρήστες πληροφορικής για την υγεία είναι ένας ισχυρός οδηγός για την προώθηση του οράματός μας. Καθώς η μεταρρύθμιση πληρωμών και παράδοσης φροντίδας αυξάνει τη ζήτηση για διαλειτουργικότητα, θα συνεργαστούμε και θα υποστηρίξουμε αυτές τις προσπάθειες.
- **Απλοποίηση.** Όπου είναι δυνατόν, θα πρέπει πρώτα να εφαρμοστούν απλούστερες λύσεις, με πρόβλεψη για πιο σύνθετες μεθόδους στο μέλλον.
- **Διατήρηση της διαμόρφωσης.** Τα πολύπλοκα συστήματα είναι πιο ανθεκτικά στην αλλαγή όταν χωρίζονται σε ανεξάρτητα στοιχεία που μπορούν να συνδεθούν μεταξύ τους. Επειδή η ιατρική και η τεχνολογία θα αλλάξουν με την πάροδο του χρόνου, πρέπει να διατηρήσουμε τις ικανότητες των συστημάτων να εξελιχθούν και να επωφεληθούν από τις καλύτερες τεχνολογικές και υγειονομικές υπηρεσίες. Η διαμόρφωση δημιουργεί ευελιξία που επιτρέπει την καινοτομία και την υιοθέτηση νέων, αποδοτικότερων προσεγγίσεων με την πάροδο του χρόνου χωρίς να αναθεωρούνται ολόκληρα συστήματα.

- **Εξέταση του τρέχοντος περιβάλλοντος και υποστήριξη πολλαπλών επιπέδων προόδου.** Η κάθε κλινική πρακτική δε θα ενσωματώσει την τεχνολογία πληροφοριών για την υγεία στην εργασία της τα επόμενα 10 χρόνια, ενώ κάθε πρακτική δε θα υιοθετήσει την τεχνολογία της υγείας στο ίδιο επίπεδο εξειδίκευσης. Συνεπώς, πρέπει να υπολογίσουμε μια σειρά δυνατοτήτων μεταξύ των πηγών πληροφοριών και των χρηστών πληροφοριών, συμπεριλαμβανομένων των χρηστών EHR και των μη ηλεκτρονικών συστημάτων, καθώς προωθούμε τη διαλειτουργικότητα. Τα άτομα και οι φροντιστές έχουν συνεχή ανάγκη να βρουν, να αποστέλλουν, να λαμβάνουν και να χρησιμοποιούν τις δικές τους πληροφορίες για την υγεία τόσο εντός όσο και εκτός του συστήματος παροχής φροντίδας και η διαλειτουργική υποδομή πρέπει να το επιτρέπει αυτό.
- **Εστίαση στην αξία.** Θα προσπαθήσουμε να διασφαλίσουμε ότι οι προσπάθειές μας για τη διαλειτουργικότητα θα έχουν τη μεγαλύτερη αξία για τα άτομα και τους φορείς παροχής φροντίδας. Η βελτίωση της υγείας, η υγειονομική περίθαλψη και το χαμηλότερο κόστος θα πρέπει να είναι μετρήσιμα με την πάροδο του χρόνου και τουλάχιστον να αντισταθμίζουν τις επενδύσεις των οικονομικών πόρων.
- **Προστασία της ιδιωτικής ζωής και της ασφάλειας σε όλες τις πτυχές της διαλειτουργικότητας.** Είναι σημαντικό να διατηρηθεί η εμπιστοσύνη του κοινού και ότι οι πληροφορίες για την υγεία είναι ασφαλείς. Για να δημιουργήσουμε και να διατηρήσουμε καλύτερα αυτή την εμπιστοσύνη, θα προσπαθήσουμε να διασφαλίσουμε την ύπαρξη κατάλληλων, ισχυρών και αποτελεσματικών διασφαλίσεων για τις πληροφορίες για την υγεία καθώς αυξάνεται η διαλειτουργικότητα σε ολόκληρο τον κλάδο. Θα υποστηρίξουμε επίσης μεγαλύτερη διαφάνεια για τα άτομα όσον αφορά τις επιχειρηματικές πρακτικές των φορέων που χρησιμοποιούν τα δεδομένα τους, ιδίως εκείνων που δεν καλύπτονται από τους κανόνες της HIPAA.

Παράλληλα, στο εγγύς μέλλον χρειάζεται να εργαστούμε για την αξιολόγηση της προτεινόμενης προσέγγισης με διαφορετικές γλώσσες αντιστοίχισης, καθώς η επίσημη αναπαράσταση αυτών των αντιστοιχιών θα είναι χρήσιμη για την επέκταση της προσέγγισής μας, παρέχοντας βελτιωμένα αποτελέσματα, με υψηλότερα επίπεδα ακρίβειας. Επιπλέον, θα συνεχίσουμε να αξιολογούμε το προτεινόμενο πλαίσιο με πολλαπλά σύνολα δεδομένων EHR, διαφόρων ιατρικών προτύπων και μορφών, συμπεριλαμβανομένων μορφών άγνωστου χαρακτήρα. Είναι επίσης μέσα στα σχέδιά μας να ενσωματωθεί η προτεινόμενη προσέγγιση για την παροχή αποτελεσμάτων σε πραγματικό χρόνο μέσω μιας

γενικής και επεκτάσιμης αρχιτεκτονικής ικανής να ασχολείται με πολλαπλά δεδομένα που σχετίζονται με την υγειονομική περίθαλψη, αντιμετωπίζοντας τους περιορισμούς ασφάλειας και προστασίας της ιδιωτικής ζωής [286]. Επιπλέον, οι εφαρμοσμένοι αλγόριθμοι για την αντιστοίχιση των δομών των δεδομένων της υγειονομικής περίθαλψης στους πόρους HL7 FHIR θα ενημερωθούν για την επίτευξη των μέγιστων επιπέδων ακρίβειας (δηλαδή ~ 100%), δεδομένου ότι πιθανά σφάλματα αντιστοίχισης των δεδομένων θα μπορούσαν να δημιουργήσουν σοβαρά ζητήματα περίθαλψης. Παρόλα αυτά, η γενική προσέγγιση φαίνεται πολλά υποσχόμενη για τον τομέα της υγειονομικής περίθαλψης, καθώς θα μπορούσε να εφαρμοστεί σε σενάρια πραγματικού κόσμου για τη μετατροπή ετερογενών δεδομένων διαφορετικής φύσης και μορφής σε ένα παγκοσμίως αναγνωρισμένο ιατρικό πρότυπο, όπως το HL7 FHIR. Αυτός είναι ο λόγος για τον οποίο στο μελλοντικό μας έργο θα μετρήσουμε και θα υπολογίσουμε τις επιχειρηματικές, στρατηγικές και οικονομικές επιπτώσεις της αναπτυγμένης προσέγγισης όσον αφορά την εφαρμογή της σε περιπτώσεις πραγματικής χρήσης.

Αναφορικά με το χώρο των οντολογιών, η συγγραφή μίας αυτοματοποιημένης τεχνικής αντιστοίχισης οντολογιών εξακολουθεί να αποτελεί δύσκολο ερευνητικό έργο, λαμβάνοντας υπόψη τους πιθανούς περιορισμούς των οντολογιών. Τέτοιοι περιορισμοί θα μπορούσαν να είναι τα περιορισμένα σύνολα κατασκευών ιδιοκτησίας και η πρόληψη των οντολογιών να προσθέσουν δεδομένα που δεν συνάδουν με τους προκαθορισμένους περιορισμούς τους, τα οποία όμως δεν είναι πάντοτε ευεργετικά, καθώς ενδέχεται να υπάρχουν περιπτώσεις όπου δεδομένα που εισάγονται από νέα πηγή να είναι διαρθρωτικά ασυμβίβαστα με τους περιορισμούς. Σε αυτό το πλαίσιο, θα εργαστούμε για την αξιολόγηση της προτεινόμενης προσέγγισης με διαφορετικές τεχνικές αντιστοίχισης οντολογιών, λαμβάνοντας υπόψη την πρωτοβουλία ΟΑΕΙ. Τα επόμενα σχέδια θα είναι η αξιολόγηση της προτεινόμενης τεχνικής με το Alignment API χρησιμοποιώντας διαφορετικούς αλγόριθμους απόστασης (π.χ. Metric String for Ontology Alignment (SMOA) [287]), ενώ ταυτόχρονα θα αξιολογήσουμε και θα συγκρίνουμε την ισχύουσα τεχνική με πρόσθετες προσεγγίσεις οντολογικής αντιστοίχισης [288][289], το σύστημα αντιστοίχισης Falcon [170], το σύστημα αντιστοίχισης RiMOM [290], το FOAM [291]) το Protégé PROMPT [292], ή το COMA++ [293]. Ως εκ τούτου, θα εντοπιστούν οι κύριες διαφορές, καθώς και τα πλεονεκτήματα και τα μειονεκτήματα της αναπτυγμένης προσέγγισης, και θα είναι ευκολότερο να εκτιμηθεί υπό ποιες συνθήκες και λόγους, κάθε οντολογική προσέγγιση ταιριάζει καλύτερα. Ως αποτέλεσμα, θα μπορέσουμε να εντοπίσουμε τα οφέλη κάθε προσέγγισης και είτε να επικαιροποιήσουμε την παρούσα προσέγγισή μας

με τα νέα μας ευρήματα είτε να επαναχρησιμοποιήσουμε μια υπάρχουσα προσέγγιση οντολογικής αντιστοίχισης προς τη διαλειτουργικότητα της υγειονομικής περιθαλψης.

Στο πλαίσιο των μελλοντικών μας στόχων, είναι να εκτελέσουμε επίσης πρόσθετες δοκιμές με την παρούσα προσέγγιση, προσαρμόζοντας τους αλγορίθμους της συντακτικής και της σημασιολογικής ομοιότητας, ώστε να καταλήξουμε στους αλγορίθμους που θα προσφέρουν πιο αξιόπιστα αποτελέσματα, με λιγότερο εσφαλμένες αντιστοιχίσεις. Αυτές οι μετρήσεις ομοιότητας θα μπορούσαν να είναι η απόσταση N-gram [294] για τον υπολογισμό του αριθμού κοινών/διακριτών n-grams μεταξύ των δοσμένων συμβολοσειρών, του SMOA [287] ή το UMLS Metathesaurus [165] για τον προσδιορισμό της συντακτικής και σημασιολογικής ομοιότητας.

Επιπλέον, θα διεξαχθούν επιπρόσθετα πειράματα για την σωστή προσαρμογή των βαρών του μέσου όρου που υπολογίζεται από τον Overall Ontology Mapper, μεταξύ των αποτελεσμάτων της συντακτικής ομοιότητας και των αποτελεσμάτων της σημασιολογικής ομοιότητας. Σε αυτό το πλαίσιο, θα μελετηθούν και θα αξιολογηθούν επιπλέον στρατηγικές συσχέτισης δεδομένων, συμπεριλαμβανομένου του Ordered Weighted Average [73], δεδομένου ότι μέχρι τώρα διαφορετικά βάρη μπορούν να προσαρμοστούν μόνο χειροκίνητα στην περίπτωση που είτε τα αποτελέσματα του Syntactic ή του Semantic Similarity Identifier δεν παρέχουν αξιόπιστα αποτελέσματα.

Σε τελικό στάδιο, θα ενημερώσουμε περαιτέρω την προσέγγιση Υπηρεσιών Διαλειτουργικότητας Δεδομένων για να επωφεληθούμε από τις κατασκευασμένες οντολογίες επαναλαμβάνοντας όχι μόνο τις οντολογίες των πόρων HL7 FHIR αλλά και τις οντολογίες των συνόλων δεδομένων για την υγεία, προκειμένου να ανανεώνουμε ήδη ορισμένες οντολογίες που έχουν ομοιότητες μεταξύ τους, με πρόσθετες τριάδες πληροφοριών, αποφεύγοντας έτσι την εκ νέου δημιουργία παρόμοιων οντολογιών.

Βιβλιογραφία

- [1] The world of IoT is shrinking, <https://www.lantronix.com/blog/embedded-world-the-world-of-iot-is-shrinking/>
- [2] Digitally transform healthcare, <https://www.gartner.com/en/industries/healthcare-providers>
- [3] What is “data interoperability”, <http://datainteroperability.org/>
- [4] The four definitions of interoperability, <https://healthcareit.me/2015/04/07/the-four-definitions-of-interoperability/>
- [5] Data interoperability, <https://www.iso.org/standard/63598.html>
- [6] What is interoperability, <http://ec.europa.eu/idabc/en/document/5313/5883.html>
- [7] Interoperability, <https://en.wikipedia.org/wiki/Interoperability>
- [8] What is an Electronic Health Record, <https://www.healthit.gov/faq/what-electronic-health-record-ehr>
- [9] Personal Health Records, <https://www.hcinnovationgroup.com/home/article/13012287/personal-health-records>
- [10] The global market for wearable medical devices was valued at \$4.8 billion in 2015, <https://www.prnewswire.com/news-releases/the-global-market-for-wearable-medical-devices-was-valued-at-48-billion-in-2015-300582717.html>
- [11] How Big Data can help save \$400 billion in healthcare costs, <https://www.cio.com/article/2993986/how-big-data-can-help-save-400-billion-in-healthcare-costs.html>
- [12] Healthcare 4.0, <http://www.uco.es/congresos/cbms2019/healthcare-4-0/>
- [13] Pandey, S. C. (2016, October). Data mining techniques for medical data: a review. In 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs) (pp. 972-982). IEEE.
- [14] ITU Telecommunication Standardization Sector, <https://www.itu.int/en/ITU-T/Pages/default.aspx>
- [15] Java Introduction, https://www.w3schools.com/java/java_intro.asp
- [16] JVM Architecture, <https://www.javatpoint.com/internal-details-of-jvm>
- [17] 3 BARRIERS TO INTEROPERABILITY, AND 3 REASONS IT MATTERS, <https://bridgeconnector.co/3-reasons-interoperability-challenges-still-exist-in-healthcare-and-3-reasons-you-should-care/>
- [18] HIMSS, <https://www.himss.org/>
- [19] CHIME, <https://chimecentral.org/>
- [20] Health Innovation Alliance, <https://health-innovation.org/>
- [21] Health IT Now Pushes for Information Blocking Regulation, Says Administration “Must Uphold its End of the Bargain”, <https://www.hcinnovationgroup.com/interoperability-hie/news/13030705/health-it-now-pushes-for-information-blocking-regulation-says-administration-must-uphold-its-end-of-the-bargain>
- [22] HL7 FHIR, <https://www.hl7.org/fhir/>
- [23] HL7 v2, https://www.hl7.org/implement/standards/product_brief.cfm?product_id=185

- [24] HL7 v3, https://www.hl7.org/implement/standards/product_brief.cfm?product_id=186
- [25] HL7 CDA, https://www.hl7.org/implement/standards/product_brief.cfm?product_id=7
- [26] INTEROPERABILITY: WHY IS IT SO HARD?, <https://medicalinteroperability.org/interoperability-why-is-it-so-hard/>
- [27] Big Data Growth, <https://blogs.thomsonreuters.com/answeron/big-data-business-opportunities-infographics/>
- [28] AI & Big Data, <https://www.forbes.com/consent/?toURL=https://www.forbes.com/ai-big-data/>
- [29] What is NoSQL, <https://www.mongodb.com/nosql-explained>
- [30] MongoDB, <https://www.mongodb.com/>
- [31] What is MapReduce, <https://www.ibm.com/analytics/hadoop/mapreduce>
- [32] Hosseini, M., & Dixon, B. E. (2016). Syntactic interoperability and the role of standards. In Health Information Exchange (pp. 123-136). Academic Press.
- [33] Jabbar, S., Ullah, F., Khalid, S., Khan, M., & Han, K. (2017). Semantic interoperability in heterogeneous IoT infrastructure for healthcare. *Wireless Communications and Mobile Computing*, 2017.
- [34] Soursos, S., Žarko, I. P., Zwickl, P., Gojmerac, I., Bianchi, G., & Carrozzo, G. (2016, June). Towards the cross-domain interoperability of IoT platforms. In 2016 European conference on networks and communications (EuCNC) (pp. 398-402). IEEE.
- [35] New European Interoperability Framework, https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf
- [36] Silberer, C., & Lapata, M. (2012, July). Grounded models of semantic representation. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (pp. 1423-1433). Association for Computational Linguistics.
- [37] The New European Interoperability Framework, https://ec.europa.eu/isa2/eif_en
- [38] Levels of Information Systems Interoperability (LISI), <http://web.cse.msstate.edu/~hamilton/C4ISR/LISI.pdf>
- [39] Moxley, F. I., Simon, L., & Wells, E. J. (2000). Laying the Foundation for Coalition Interoperability through NATO's C3 Technical Architecture. DEFENSE INFORMATION SYSTEMS AGENCY FALLS CHURCH VA.
- [40] Taylor, S. J., Mustafee, N., Strassburger, S., Turner, S. J., Low, M. Y., & Ladbrook, J. (2007, December). The SISO CSPI PDG standard for commercial off-the-shelf simulation package interoperability reference models. In Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come (pp. 594-602). IEEE Press.
- [41] Clark, T., & Jones, R. (1999, June). Organisational interoperability maturity model for C2. In Proceedings of the 1999 Command and Control Research and Technology Symposium (pp. 1-13).
- [42] ISO, <https://www.iso.org/home.html>
- [43] ISO 15000, <https://www.iso.org/standard/61433.html>
- [44] ISO/IEC 11179-1, <https://www.iso.org/standard/61932.html>
- [45] CEN, <https://www.cen.eu/Pages/default.aspx>
- [46] UN/CEFACT, <https://www.unece.org/cefact.html>
- [47] UBL, https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl
- [48] CEN-BII, <https://eespa.eu/glossary/cen-bii/>

- [49] XML Naming and Design Rules Technical Specification, <https://www.unece.org/fileadmin/DAM/cefact/xml/UNCEFACT%2BXML%2BND R%2BV3p0.pdf>
- [50] UN/CEFACT Modeling Methodology, http://www.unece.org/fileadmin/DAM/cefact/umm/old_umm_files/ch1_intro.pdf
- [51] Unified modelling language, <https://www.uml.org/>
- [52] Core Component Technical Specification, https://www.unece.org/cefact/codesfortrade/ccts_index.html
- [53] World Wide Web Consortium, <https://www.w3.org/>
- [54] URI schemes, <https://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml>
- [55] HTTP, <https://www.w3.org/Protocols/HTTP/1.1/rfc2616bis/draft-lafon-rfc2616bis-03.html>
- [56] XML schema, <https://www.w3.org/standards/xml/schema>
- [57] XSL, <https://www.w3.org/Style/XSL/>
- [58] XML and XPath, https://www.w3schools.com/xml/xml_xpath.asp
- [59] SOAP, https://docs.oracle.com/cd/A97335_02/integrate.102/a90297/overview.htm
- [60] SMTP, <https://techterms.com/definition/smtp>
- [61] WSDL, <https://www.w3.org/TR/2001/NOTE-wsdl-20010315>
- [62] IDABC, <http://ec.europa.eu/idabc/en/home.html>
- [63] ISA2, https://ec.europa.eu/isa2/home_en
- [64] OASIS, https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=oic
- [65] Oasis Cover Pages, <https://www.oasis-open.org/aggregator/sources/1>
- [66] Oasis xml.org, <https://www.oasis-open.org/news/pr/oasis-launches-xml-org>
- [67] Oasis eGovernment, <https://www.oasis-open.org/committees/egov/>
- [68] IDA and the EU's e-government policy, <http://ec.europa.eu/idabc/en/document/1920.html>
- [69] Understanding Transmission Control Protocol/Internet Protocol (TCP/IP), <https://www.lifewire.com/transmission-control-protocol-and-internet-protocol-816255>
- [70] Open Database Connectivity, <https://www.simba.com/resources/odbc/>
- [71] Common Object Request Broker Architecture (CORBA), https://www.ibm.com/support/knowledgecenter/en/SSMKHH_9.0.0/com.ibm.etools.mft.doc/bc22400_htm
- [72] IDL, <https://www.omg.org/spec/IDL/About-IDL/>
- [73] SGML, <https://www.loc.gov/preservation/digital/formats/fdd/fdd000465.shtml>
- [74] Arayici, Y., Fernando, T., Munoz, V., & Bassanino, M. (2018). Interoperability specification development for integrated BIM use in performance based design. *Automation in Construction*, 85, 167-181.
- [75] Arunkumar, G., & Venkataraman, N. (2015). A novel approach to address interoperability concern in cloud computing. *Procedia Computer Science*, 50, 554-559.
- [76] Mecca, G., Santomauro, M., Santoro, D., & Veltri, E. (2016). On federated single sign-on in e-government interoperability frameworks. *International Journal of Electronic Governance*, 8(1), 6-21.
- [77] What is Interoperability?, <https://www.himss.org/library/interoperability-standards/what-is-interoperability>
- [78] Healthcare interoperability, analytics get top billing in health IT, <https://searchhealthit.techtarget.com/opinion/Healthcare-interoperability-analytics-get-top-billing-in-health-IT>

- [79] Four challenges achieving healthcare interoperability, <https://www.tigerconnect.com/blog/four-challenges-achieving-healthcare-interoperability/>
- [80] The 5 key benefits of healthcare interoperability, <https://www.beckershospitalreview.com/healthcare-information-technology/the-5-key-benefits-of-healthcare-interoperability.html>
- [81] Whitepaper: Why clinical interoperability requires a single record, not a single system, <https://www.medicaleconomics.com/article/whitepaper-why-clinical-interoperability-requires-single-record-not-single-system>
- [82] 6 barriers to healthcare interoperability, according to ONC, <https://www.beckershospitalreview.com/ehrs/6-barriers-to-healthcare-interoperability-according-to-onc.html>
- [83] The challenges and opportunities with interoperability, <https://www.hcinnovationgroup.com/policy-value-based-care/article/13009318/the-challenges-and-opportunities-with-interoperability>
- [84] DirectTrust, <https://www.directtrust.org/>
- [85] Poor Data Sharing Leads to Delays in Care, Global Survey Finds, <https://www.hcinnovationgroup.com/interoperability-hie/news/13030463/poor-data-sharing-leads-to-delays-in-care-global-survey-finds>
- [86] CEN/TC 251 - Health informatics, https://standards.cen.eu/dyn/www/f?p=204:7:0:::FSP_ORG_ID:6232&cs=18CA078392807EDD402B798AAEF1644E1
- [87] openEHR, <https://www.openehr.org/>
- [88] What is Healthcare Terminology, https://www.healthcatalyst.com/what_is_healthcare_terminology/
- [89] ICD-10, <https://icd.who.int/browse10/2010/en>
- [90] DICOM, <https://www.dicomstandard.org/>
- [91] SNOMED, <http://www.snomed.org/>
- [92] LOINC, <https://loinc.org/>
- [93] World Health Organization, <https://www.who.int/>
- [94] Gonçalves, M. J. A., Rocha, Á., & Cota, M. P. (2015). Interoperability Framework for Competences and Learning Outcomes. *J. UCS*, 21(8), 1042-1060.
- [95] Allen, W. E. (2009). Terminologia anatomica: international anatomical terminology and terminologia histologica: international terms for human cytology and histology. *Journal of anatomy*, 215(2), 221.
- [96] IHE, <https://www.ihe.net/>
- [97] NEMA, <https://www.nema.org/pages/default.aspx>
- [98] Regenstrief Institute, <https://www.regenstrief.org/>
- [99] Personal Connected Health Alliance, <https://www.pchalliance.org/>
- [100] WONCA, <https://www.globalfamilydoctor.com/>
- [101] American College of Radiology, <https://www.acr.org/>
- [102] SNOMED RT, <https://confluence.ihtsdotools.org/display/DOCGLOSS/reference+terminology>
- [103] College of American Pathologists, <https://www.cap.org/>
- [104] CTV3, <https://confluence.ihtsdotools.org/display/DOCGLOSS/NHS+Clinical+Terms+Version+3>
- [105] What is Rest?, <https://restfulapi.net/>

- [106] Introduction to HTML, https://www.w3schools.com/html/html_intro.asp
- [107] CSS Introduction, https://www.w3schools.com/css/css_intro.asp
- [108] JSON, <https://www.json.org/>
- [109] Atom, <https://atom.io/>
- [110] HTTPS, <https://searchsoftwarequality.techtarget.com/definition/S-HTTP>
- [111] OAuth, <https://oauth.net/>
- [112] Jian, W. S., Hsu, C. Y., Hao, T. H., Wen, H. C., Hsu, M. H., Lee, Y. L., ... & Chang, P. (2007). Building a portable data and information interoperability infrastructure—framework for a standard Taiwan Electronic Medical Record Template. *Computer methods and programs in biomedicine*, 88(2), 102-111.
- [113] XSLT Introduction, https://www.w3schools.com/xml/xsl_intro.asp
- [114] Web Ontology Language, <https://www.w3.org/OWL/>
- [115] Dogac, A., Laleci, G. B., Kirbas, S., Kabak, Y., Sinir, S. S., Yildiz, A., & Gurcan, Y. (2006). Artemis: Deploying semantically enriched Web services in the healthcare domain. *Information Systems*, 31(4-5), 321-339.
- [116] Kilic, O., & Dogac, A. (2009). Achieving clinical statement interoperability using R-MIM and archetype-based semantic transformations. *IEEE Transactions on Information Technology in Biomedicine*, 13(4), 467-477.
- [117] ISO 13606, <https://www.iso.org/standard/40784.html>
- [118] Chen, R., Klein, G. O., Sundvall, E., Karlsson, D., & Åhlfeldt, H. (2009). Archetype-based conversion of EHR content models: pilot experience with a regional EHR system. *BMC medical informatics and decision making*, 9(1), 33.
- [119] ISO 21090, <https://www.iso.org/standard/35646.html>
- [120] Goossen, W., Goossen-Baremans, A., & Van Der Zel, M. (2010). Detailed clinical models: a review. *Healthcare informatics research*, 16(4), 201-214.
- [121] Kim, W., Lim, S., Ahn, J., Nah, J., Kim, N. Integration of IEEE 1451 and HL7 exchanging information for patients' sensor data. *Journal of medical systems*, 34(6), pp. 1033-1041 (2010)
- [122] Lasierra, N. et al.: A three stage ontology-driven solution to provide personalized care to chronic patients at home." *Journal of biomedical informatics* 46.3, pp. 516-529 (2013)
- [123] Kim, JH., Lee, DS., Chung, KY.: Context-aware based item recommendation for personalized service. *Proc Int Conf Inf Sci Appl IEEE Comp Soc* pp. 595-600 (2011)
- [124] Bouamrane, M. M., Tao, C., Sarkar, I. N.: Managing interoperability and complexity in health systems. *Methods of information in medicine*, 54(1), pp. 1-4 (2005)
- [125] Bahga, A., Madiseti, V. K.: A cloud-based approach for interoperable electronic health records (EHRs). *IEEE Journal of Biomedical and Health Informatics*, 17(5), pp. 894-906 (2013)
- [126] Mezghani, E., Exposito, E., Drira, K., Da Silveira, M., Pruski, C.: A semantic big data platform for integrating heterogeneous wearable data in healthcare. *Journal of medical systems*, 39(12), p. 185 (2015)
- [127] Batra, S., & Sachdeva, S. Data Quality for Semantic Interoperable Electronic Health Records. in *Databases (QDB'16)*, 21.
- [128] ADL, <https://specifications.openehr.org/releases/AM/latest/ADL2.html>
- [129] Detro, S. P., Morozov, D., Lezoche, M., Panetto, H., Santos, E. P., & Zdravkovic, M. (2016, February). Enhancing semantic interoperability in healthcare using semantic process mining. In *6th International Conference on Information Society and Technology, ICIST 2016 (Vol. 1, pp. 80-85)*.

- [130] Blackman-Lees, S. (2018, January). Towards a conceptual framework for persistent use: A technical plan to achieve semantic interoperability within electronic health record systems. In Proceedings of the 51st Hawaii International Conference on System Sciences.
- [131] Lassila, O., & Swick, R. R. (1998). Resource description framework (RDF) model and syntax specification.
- [132] Nee, O., Hein, A., Gorath, T., Hülsmann, N., Laleci, G. B., Yuksel, M., ... & Fruntelata, A. (2008). SAPHIRE: intelligent healthcare monitoring based on semantic interoperability platform: pilot applications. *IET communications*, 2(2), 192-201.
- [133] Ryan, A., & Eklund, P. (2008). A framework for semantic interoperability in healthcare: A service oriented architecture based on health informatics standards. *Studies in health technology and informatics*, 136, 759.
- [134] Schmidt, M. T., Hutchison, B., Lambros, P., & Phippen, R. (2005). The enterprise service bus: making service-oriented architecture real. *IBM Systems Journal*, 44(4), 781-797.
- [135] Jini Technology, <https://river.apache.org/release-doc/current/specs/html/jini-spec.html>
- [136] Schulz, S., & Martínez-Costa, C. (2013). How ontologies can improve semantic interoperability in health care. In *Process Support and Knowledge Representation in Health Care* (pp. 1-10). Springer, Cham.
- [137] Horrocks, I., & Patel-Schneider, P. F. (2003, October). Reducing OWL entailment to description logic satisfiability. In *International semantic web conference* (pp. 17-29). Springer, Berlin, Heidelberg.
- [138] Marcos, C., González-Ferrer, A., Peleg, M., & Cavero, C. (2015). Solving the interoperability challenge of a distributed complex patient guidance system: a data integrator based on HL7's Virtual Medical Record standard. *Journal of the American Medical Informatics Association*, 22(3), 587-599.
- [139] Oracle Data Integrator, <https://www.oracle.com/middleware/technologies/data-integrator.html>
- [140] Virtual Medical Record, https://www.hl7.org/implement/standards/product_brief.cfm?product_id=338
- [141] Martínez-Costa, C., Menárguez-Tortosa, M., & Fernández-Breis, J. T. (2010). An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. *Journal of biomedical informatics*, 43(5), 736-746.
- [142] McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. *W3C recommendation*, 10(10), 2004.
- [143] OWL 2 Web Ontology Language, <https://www.w3.org/TR/owl2-syntax/>
- [144] Semantic Web, <https://www.w3.org/standards/semanticweb/>
- [145] A Simple Explanation Of 'The Internet Of Things', <https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#31d4beb71d09>
- [146] A. Yaser, and Bish, "Overcoming the semantic and other barriers to GIS interoperability", *International Journal of Geographical Information Science*, 12(4), 1998, pp 229- 314.
- [147] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-220.
- [148] World Wide Web, <https://webfoundation.org/>
- [149] OWL-Lite, <https://stinson.ca/>
- [150] OWL Full, <https://www.w3.org/TR/owl-ref/>

- [151] What is RDF triplestore?, <https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore/>
- [152] RDF Schema, <https://www.w3.org/TR/rdf-schema/>
- [153] What is a URL, <https://www.lifewire.com/what-is-a-url-2626035>
- [154] Ferdinand, M., Zirpins, C., & Trastour, D. (2004, July). Lifting XML schema to OWL. In International conference on web engineering (pp. 354-358). Springer, Berlin, Heidelberg.
- [155] Bohring, H., & Auer, S. (2005). Mapping XML to OWL Ontologies. *Leipziger Informatik-Tage*, 72, 147-156.
- [156] Ermilov, T., Heino, N., Tramp, S., & Auer, S. (2011, May). Ontowiki mobile-knowledge management in your pocket. In *Extended Semantic Web Conference* (pp. 185-199). Springer, Berlin, Heidelberg.
- [157] Ghawi, R., & Cullot, N. (2009, August). Building ontologies from XML data sources. In *2009 20th International Workshop on Database and Expert Systems Application* (pp. 480-484). IEEE.
- [158] Bedini, I., Matheus, C., Patel-Schneider, P. F., Boran, A., & Nguyen, B. (2011, September). Transforming XML schema to OWL using patterns. In *2011 IEEE Fifth International Conference on Semantic Computing* (pp. 102-109). IEEE.
- [159] Xu, J., & Li, W. (2007, December). Using relational database to build OWL ontology from XML data sources. In *2007 International Conference on Computational Intelligence and Security Workshops (CISW 2007)* (pp. 124-127). IEEE.
- [160] P. Wennerberg, "Aligning medical domain ontologies for clinical query extraction", 12th conference of the EU chapter of the association for computational linguistics: student research workshop. Association for Comp. Linguistics, 2009.
- [161] V. Kumar, "Ontology based public healthcare system in Internet of Things (IoT)", *Procedia Computer Science*, vol. 50, pp. 99-102, 2015.
- [162] J. Kim, and K. Y. Chung, "Ontology-based healthcare context information model to implement ubiquitous environment", *Multimedia Tools and Applications*, vol. 71(2), pp. 873-888, 2014.
- [163] M. Azarm, and L. Peyton, "An ontology for a patient-centric healthcare interoperability framework", *IEEE/ACM International Workshop on Software Engineering in Healthcare Systems (SEHS)*, pp. 34-41, 2018.
- [164] H. Liyanage, P. Krause, and S. de Lusignan, "Using ontologies to improve semantic interoperability in health data". *Journal of innovation in health informatics*, vol. 22(2), pp. 309-315, 2015.
- [165] McInnes, B. T., Pedersen, T., & Pakhomov, S. V. (2009). UMLS-Interface and UMLS-Similarity: open source software for measuring paths and semantic similarity. In *AMIA Annual Symposium Proceedings (Vol. 2009, p. 431)*. American Medical Informatics Association.
- [166] Unified Medical Language System (UMLS), <https://www.nlm.nih.gov/research/umls/>
- [167] Mihalcea, R., Corley, C., & Strapparava, C. (2006, July). Corpus-based and knowledge-based measures of text semantic similarity. In *AAAI (Vol. 6, pp. 775-780)*.
- [168] Jiang, J. J., & Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. *arXiv preprint cmp-lg/9709008*.
- [169] Khan, W. A., Khattak, A. M., Lee, S., Hussain, M., Amin, B., & Latif, K. (2012, February). Achieving interoperability among healthcare standards: building semantic mappings

- at models level. In Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication (p. 101). ACM.
- [170] Hu, W., & Qu, Y. (2008). Falcon-AO: A practical ontology matching system. *Journal of web semantics*, 6(3), 237-239.
- [171] Sunna, W., & Cruz, I. (2007, November). Using the agreement maker to align ontologies for the OAEI campaign 2007. In Proceedings of the 2nd International Conference on Ontology Matching-Volume 304 (pp. 133-138). CEUR-WS. org.
- [172] S. Sonsilphong et al., "A semantic interoperability approach to health-care data: Resolving data-level conflicts", *Expert Systems*, vol. 33(6), pp. 531-547, 2016.
- [173] M. Mony, et al., "Semantic search based on ontology alignment for information retrieval", *Journal of Com. App.*, vol. 107(10), 2014.
- [174] A., Sunitha, and G. S., Babu. "Ontology-Driven Knowledge-Based Health-Care System, An Emerging Area - Challenges And Opportunities - Indian Scenario", In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-8, pp. 239-246, November 2014
- [175] A. H., Nezhadi, B., Shadgar, and A., Osareh, "Ontology alignment using machine learning techniques", *Int. Journal of Computer Science & Information Technology*, vol. 3(2), pp. 139-150, 2011
- [176] Ontology Alignment Evaluation Initiative, <http://oaei.ontologymatching.org/>
- [177] Collobert, R., & Weston, J. (2008, July). A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning (pp. 160-167). ACM.
- [178] Zhao, Z., & Liu, H. (2007, June). Spectral feature selection for supervised and unsupervised learning. In Proceedings of the 24th international conference on Machine learning (pp. 1151-1157). ACM.
- [179] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- [180] Dahlgren, K., & Stabler, E. (1998). U.S. Patent No. 5,794,050. Washington, DC: U.S. Patent and Trademark Office.
- [181] Atserias, J., Casas, B., Comelles, E., González, M., Padró, L., & Padró, M. (2006, May). FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. In *LREC (Vol. 6, pp. 48-55)*.
- [182] Evolution of Natural Language Processing, https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html
- [183] Yujian, L., & Bo, L. (2007). A normalized Levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, 29(6), 1091-1095.
- [184] Gragera, A., & Suppakitpaisarn, V. (2016, March). Semimetric properties of sørensen-dice and tversky indexes. In *International Workshop on Algorithms and Computation* (pp. 339-350). Springer, Cham.
- [185] Nguyen, H. V., & Bai, L. (2010, November). Cosine similarity metric learning for face verification. In *Asian conference on computer vision* (pp. 709-720). Springer, Berlin, Heidelberg.
- [186] Melter, R. A. (1987). Some characterizations of city block distance. *Pattern recognition letters*, 6(4), 235-240.
- [187] Dreßler, K., & Ngonga Ngomo, A. C. (2017). On the efficient execution of bounded jaro-winkler distances. *Semantic Web*, 8(2), 185-196.
- [188] Real, R., & Vargas, J. M. (1996). The probabilistic basis of Jaccard's index of similarity. *Systematic biology*, 45(3), 380-385.

- [189] Kunimoto, R., Vogt, M., & Bajorath, J. (2016). Maximum common substructure-based Tversky index: an asymmetric hybrid similarity measure. *Journal of computer-aided molecular design*, 30(7), 523-531.
- [190] Mizuno, S., Yamaguchi, T., Fukushima, A., Matsuyama, Y., & Ohashi, Y. (2005). Overlap coefficient for assessing the similarity of pharmacokinetic data between ethnically different populations. *Clinical Trials*, 2(2), 174-181.
- [191] Nelsen, R. B. (2001). Kendall tau metric. *Encyclopaedia of mathematics*, 3, 226-227.
- [192] Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, pp. 133-142).
- [193] Su, Z., Ahn, B. R., Eom, K. Y., Kang, M. K., Kim, J. P., Kim, M. K.: Plagiarism detection using the Levenshtein distance and Smith-Waterman algorithm, *Innovative Computing Information and Control (ICICIC)*, International Conference on IEEE, pp. 569-569 (2008).
- [194] Ligowski, L., & Rudnicki, W. (2009, May). An efficient implementation of Smith Waterman algorithm on GPU using CUDA, for massively parallel scanning of sequence databases. In *2009 IEEE International Symposium on Parallel & Distributed Processing* (pp. 1-8). IEEE.
- [195] Paul, S., Maji, P.: City block distance and rough-fuzzy clustering for identification of co-expressed micrnas, *Molecular BioSystems*, vol. 10(6), pp. 1509-1523 (2014).
- [196] Blondal, T., Nielsen, S. J., Baker, A., Andreasen, D., Mouritzen, P., Teilum, M. W., & Dahlsveen, I. K. (2013). Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods*, 59(1), S1-S6.
- [197] Jain, A., Jain, A., Chauhan, N., Singh, V., Thakur, N.: Information retrieval using cosine and jaccard similarity measures in vector space model, *International Journal of Computer Applications*, vol. 164(6) (2017).
- [198] Buck, C., Koehn, P.: Quick and reliable document alignment via tf/idf-weighted cosine distance, In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, vol. 2, pp. 672-678 (2016).
- [199] Sari, A. P., Saptono, R., Suryani, E.: The Implementation of Jaro-Winkler Distance and Naive Bayes Classifier for Identification System of Pests and Diseases on Paddy, *ITSMART: Jurnal Teknologi dan Informasi*, vol. 7(1), pp. 1-7 (2018).
- [200] Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).
- [201] Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E. G., & Milios, E. E. (2005, November). Semantic similarity methods in wordNet and their application to information retrieval on the web. In *Proceedings of the 7th annual ACM international workshop on Web information and data management* (pp. 10-16). ACM.
- [202] Wordnet, <https://wordnet.princeton.edu/>
- [203] MeSH, <https://meshb.nlm.nih.gov/>
- [204] Rong, X. (2014). word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*.
- [205] Kovalev, V., Kalinovskiy, A., & Kovalev, S. (2016). Deep learning with theano, torch, caffe, tensorflow, and deeplearning4j: Which one is the best in speed and accuracy?.
- [206] SCALA, <https://www.scala-lang.org/>
- [207] Apache Spark, <https://spark.apache.org/>
- [208] Aarts, E., & Korst, J. (1988). Simulated annealing and Boltzmann machines.
- [209] Continuous Bag of Words, <https://iksinc.online/tag/continuous-bag-of-words-cbow/>

- [210] How does Word2Vec's Skip-Gram work?, <https://becominghuman.ai/how-does-word2vecs-skip-gram-work-f92e0525def4>
- [211] Cortical.io, <https://www.cortical.io/>
- [212] Webber, F. D. S. (2015). Semantic folding theory and its application in semantic fingerprinting. arXiv preprint arXiv:1511.08855.
- [213] Hinton, G. E., & Ghahramani, Z. (1997). Generative models for discovering sparse distributed representations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1358), 1177-1190.
- [214] "Gartner: Top 10 Strategic Technology Trends For 2013", <http://www.forbes.com/sites/eric savitz/2012/10/23/gartner-top-10-strategic-technology-trends-for-2013>
- [215] "5 IoT Statistics You Need to Know in 2019", <https://www.vxchnge.com/blog/iot-statistics>
- [216] ITU-T Y.2060, "Overview of Internet of Things", June 2012
- [217] Ian G. Smith, "The Internet of Things 2012 - New Horizons-Cluster Book", 2012.
- [218] Van Deursen, A., & Klint, P. (2002). Domain-specific language design requires feature descriptions. *Journal of Computing and Information Technology*, 10(1), 1-17.
- [219] Atkinson, C., Kennel, B., & Goß, B. (2010, October). The level-agnostic modeling language. In *International Conference on Software Language Engineering* (pp. 266-275). Springer, Berlin, Heidelberg.
- [220] A. Abraham, "Artificial Neural Networks. Handbook of Measuring System Design, J. Wiley & Sons, Ltd, 2005
- [221] Luis J., Mena, et al. "Mobile personal health system for ambulatory blood pressure monitoring." *Computational and mathematical methods in medicine*, 2013.
- [222] "The wearables report 2016: Reviewing a fast-changing market" <https://www.fbicgroup.com/sites/default/files/The%20Wearables%20Report%202016%20by%20FBIC%20Global%20Retail%20and%20Technology%20June%2021%202016.pdf>
- [223] "SNOMED - IHTSDO" <http://www.ihtsdo.org/about-us/history/snomed/>
- [224] "IDC" <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>
- [225] "EU citizens' digital health literacy" http://ec.europa.eu/public_opinion/flash/fl_404_en.pdf
- [226] C., Domingo. "Managing healthcare through social networks." *Computer* 43.7, 2010, pp. 20-25.
- [227] J. M., Hook, et al. "The value of healthcare information exchange and interoperability in New York state." *AMIA Annual Symposium Proceedings*. Vol. 2006. American Medical Informatics Assoc., 2006.
- [228] What is the HITECH ACT, <https://compliance-group.com/what-is-the-hitech-act/>
- [229] Patient Protection and Affordable Care Act, <https://www.healthcare.gov/glossary/patient-protection-and-affordable-care-act/>
- [230] Social, IOM Capturing. "Behavioral Domains in Electronic Health Records, Phase 1." Edited by Records CotRSaBDaMfEH. Washington DC: Institute of Medicine, 2014.
- [231] V. Stroetman, et al. "Semantic interoperability for better health and safer healthcare [34 pages].", The European Commission, 2009, pp. 1-34.
- [232] A., Celesti, M., Fazio, A., Romano, A., & M., Villari. "A hospital cloud-based archival information system for the efficient management of HL7 big data". *MIPRO*, 39th Intern. Convention, IEEE, 2016, pp. 406-411.

- [233] "OpenRefine" <http://openrefine.org>
- [234] "SVMs" <http://www.support-vector-machines.org/>
- [235] "Protégé Framework" <http://protege.stanford.edu/>
- [236] GraphDB: "<http://graphdb.ontotext.com/>"
- [237] I., Shvorob. "New Approach for Saving Semistructured Medical Data." *Advances in Intel. Systems and Computing*. Springer, 2017, pp. 29-40.
- [238] Kiourtis, A., Mavrogiorgou, A., Kyriazis, D., Maglogiannis, I., & Themistocleous, M. (2016, March). Towards data interoperability: turning domain specific knowledge to agnostic across the data lifecycle. In *2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*(pp. 109-114). IEEE.
- [239] A., Ryan. "Towards semantic interoperability in healthcare: ontology mapping from SNOMED-CT to HL7 version 3." *2nd Australasian workshop on Advances in ontologies*, Vol. 72, 2006.
- [240] Gartner Press Release, <http://www.gartner.com/newsroom/id/3598917>
- [241] Kwon, J., Kim, D. H., Park, W., Kim, L.: A wearable device for emotional recognition using facial expression and physiological response. In: *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pp. 5765-5768. IEEE Press (2016)
- [242] Dionisi, A., Marioli, D., Sardini, E., Serpelloni, M.: Autonomous wearable system for vital signs measurement with energy-harvesting module. *IEEE Transactions on Instrumentation and Measurement*, 65(6), pp. 1423-1434 (2016)
- [243] Koshti, M., Ganorkar, S., Chiari, L.: IoT Based Health Monitoring System by Using Raspberry Pi and ECG Signaly. *International Journal of Innovative Research in Science, Engineering and Technology*, 5(5) (2016)
- [244] Kalra, M., Lal, N.: Data mining of heterogeneous data with research challenges. In *Colossal Data Analysis and Networking (CDAN)*, pp. 1-6. IEEE (2016)
- [245] Medical errors drop with improved communication during hospital shift changes, <https://source.wustl.edu/2014/11/medical-errors-drop-with-improved-communication-during-hospital-shift-changes/>
- [246] EUROPEAN CITIZENS' DIGITAL HEALTH LITERACY, http://ec.europa.eu/public_opinion/flash/fl_404_en.pdf
- [247] Domingo, M.C.: Managing Healthcare through Social Networks, *IEEE Computer*, vol. 43, no. 7, pp. 20-25 (2010)
- [248] Mead, CN.: Data interchange standards in healthcare IT-computable semantic interoperability: now possible but still difficult, do we really need a better mousetrap? *Journal of Healthcare Information Management*, 20(1), pp. 71-78 (2006)
- [249] Cuadrado, J. S., Molina, J. G., Tortosa, M. M.: RubyTL: a practical, extensible transformation language. In *European Conference on Model Driven Architecture-Foundations and Applications*. Springer Berlin Heidelberg. pp.158-172 (2006)
- [250] MOFScript, <https://eclipse.org/gmt/mofscript/>
- [251] Kendall, E., Dutra, M., Jacobs, J., Schwab, S.: The Ontology Definition Metamodel (ODM) and Motivation for a Semantic Meta-Object Facility (SMOF).
- [252] Iyengar, S. S. (2005). U.S. Patent No. 6,874,146. Washington, DC: U.S. Patent and Trademark Office.
- [253] Fernández-Breis, J. T., Martiinez-Bejar, R.: A cooperative framework for integrating ontologies. *International J of Human-Computer Studies*, 56(6), pp. 665-720 (2002)
- [254] SMT007 MAGAZINE (March 2018), <http://iconnect007.uberflip.com/i/948150-smt-mar2018/0?m4=>

- [255] Malpractice Report Finds Hospital Miscommunication Costs \$1.7 Billion, <https://www.managedcaremag.com/news/malpractice-report-finds-hospital-miscommunication-costs-17-billion>
- [256] Healthcare Interoperability Can Reduce Cost and Improve Care, <https://www.isharemedical.com/healthcare-interoperability-reduce-cost/>.
- [257] A. Parand, et al. "The role of hospital managers in quality and patient safety: a systematic review." *BMJ open*, vol. 4(9), 2014.
- [258] S. Schulz, and C. Martínez-Costa. "How ontologies can improve semantic interoperability in health care." *Process Support and Knowledge Representation in Health Care*. Springer, pp. 1-10, 2013.
- [259] C. Chudzian, "Ontology creation process in knowledge management support system for a research institute", *Journal of Telecommunications and Information Technology*, pp. 47-53, 2008.
- [260] FHIR data module, <https://www.hl7.org/fhir/linked-data-module.html>
- [261] Staworko, S., Boneva, I., Labra Gayo, J. E., Hym, S., Prud'hommeaux, E. G., & Solbrig, H. (2015). Complexity and Expressiveness of ShEx for RDF. In 18th International Conference on Database Theory (ICDT 2015). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- [262] Trang, <http://www.thaiopensource.com/relaxng/trang.html>
- [263] XML Schema Object Model (XSOM), <https://directory.fsf.org/wiki/XSOM>
- [264] Jung, <http://jung.sourceforge.net/>
- [265] M., Hacherouf, S. N. Bahloul, and C. Cruz, "Transforming XML documents to OWL ontologies: A survey." *Journal of Information Science*, vol. 41(2), pp. 242-259, 2015.
- [266] What is Open Data?, <http://opendatahandbook.org/guide/en/what-is-open-data/>
- [267] DBPedia, <https://wiki.dbpedia.org/>
- [268] GeoNames, <https://www.geonames.org/>
- [269] Seber, G. A., & Lee, A. J. (2012). *Linear regression analysis* (Vol. 329). John Wiley & Sons.
- [270] Karolinska Institute, <https://ki.se/>
- [271] David, J., Euzenat, J., Scharffe, F., & Trojahn dos Santos, C. (2011). The alignment API 4.0. *Semantic web*, 2(1), 3-10.
- [272] Netbeans IDE, <https://netbeans.org/>
- [273] Apache Camel: Dozer, <https://camel.apache.org/dozer.html>
- [274] Java Bean, <https://www.javatpoint.com/java-bean>
- [275] Misfit Vapor, <https://misfit.com/misfit-vapor>.
- [276] BioAssist, <https://bioassist.gr/>.
- [277] UMLS-based reference alignment, http://www.cs.ox.ac.uk/isg/projects/SEALS/oaei/2017/oaei2017_umls_reference.html
- [278] Global M2M market, <https://machinaresearch.com/news/global-m2m-market-to-grow-to-27-billion-devices-generating-usd16-trillion-revenue-in-2024/>
- [279] Cisco, Cisco visual networking index: Global mobile data traffic forecast update, White Paper (2015).
- [280] Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. *Healthcare informatics research*, 22(3), 156-163.
- [281] Health data exchange could save Medicare, <https://healthitanalytics.com/news/health-data-exchange-could-save-medicare-3.12-billion-a-year>

- [282] Growing Investments to overcome Healthcare Interoperability Issues, <https://www.prnewswire.com/news-releases/global-healthcare-middleware-market-to-2023-growing-investments-to-overcome-healthcare-interoperability-issues-300659900.html>
- [283] Mavrogiorgou, A., Kiourtis, A., Touloupou, M., Kapassa, E., & Kyriazis, D. (2019). Internet of Medical Things (IoMT): Acquiring and Transforming Data into HL7 FHIR through 5G Network Slicing. *Emerging Science Journal*, 3(2), 64-77.
- [284] Mavrogiorgou, A., Kiourtis, A., Symvoulidis C., & Kyriazis, D. (2018, November). Identification of medical IoT devices APIs through ontology mapping techniques. In 5th EAI International Conference on IoT Technologies for HealthCare (HealthyIoT) (in press). Springer.
- [285] Global Healthcare Interoperability Solutions Market Forecast to 2024: Projected to Grow at a CAGR of 12.6% During the Forecast Period to Reach \$4.2B, <https://www.businesswire.com/news/home/20190327005479/en/Global-Healthcare-Interoperability-Solutions-Market-Forecast-2024>
- [286] Kiourtis, A., Mavrogiorgou, A., & Kyriazis, D. (2018, August). Towards a Secure Semantic Knowledge of Healthcare Data Through Structural Ontological Transformations. In *Joint Conference on Knowledge-Based Software Engineering* (pp. 178-188). Springer, Cham.
- [287] G., Stoilos, G., Stamou, and S., Kollias, "A string metric for ontology alignment. In *International Semantic Web Conference*", pp. 624-637, 2005.
- [288] J., Euzenat, and P., Shvaiko, "Ontology matching", vol. 18. Heidelberg: Springer, 2007.
- [289] Ontology Matching, <http://ontologymatching.org/>
- [290] J., Li, J., Tang, Y., Li, and Q., Luo. "RiMOM: A dynamic multistrategy ontology alignment framework". *IEEE Transactions on Knowledge and data Engineering*, vol. 21(8), pp. 1218-1232, 2009.
- [291] E., Prestat, M. M., David, J., Hultman, N., Taş, R., Lamendella, J., Dvornik, and E., Holman, "FOAM (functional ontology assignments for metagenomes): a hidden Markov model (HMM) database with environmental focus". *Nucleic acids research*, vol. 42(19), pp. 145-145, 2014.
- [292] M., Taboada, D., Martinez, and J., Mira, "Experiences in reusing knowledge sources using Protégé and PROMPT". *International Journal of Human-Computer Studies*, vol. 62(5), pp. 597-618, 2005.
- [293] D., Aumueller, H. H., Do, S., Massmann, and E., Rahm, "Schema and ontology matching with COMA++". In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pp. 906-908, 2005.
- [294] G., Kondrak, "N-gram similarity and distance". In *International symposium on string processing and information retrieval*, pp. 115-126, 2005.
- [295] Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on systems, Man, and Cybernetics*, 18(1), 183-190.