



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πρόγραμμα Μεταπτυχιακών Σπουδών
«Προηγμένα Συστήματα Πληροφορικής – Ανάπτυξη Λογισμικού και Τεχνητής
Νοημοσύνης»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Διαδικτυακή εφαρμογή εμφάνισης και αξιολόγησης ειδήσεων με χρήση Angular και Python Web App platform for reading and reviewing news using Angular and Python
Όνοματεπώνυμο Φοιτητή	Κωνσταντίνος Ντζιαχρήστος
Πατρώνυμο	Βασίλειος
Αριθμός Μητρώου	ΜΠΣΠ 21039
Επιβλέπων	Ευθύμιος Αλέπης, Αναπληρωτής Καθηγητής

Ημερομηνία Παράδοσης **Νοέμβριος 2023**

Τριμελής Εξεταστική Επιτροπή

Ευθύμιος Αλέπης	Μαρία Βίρβου	Κωνσταντίνος Πατσάκης
Αναπληρωτής Καθηγητής	Καθηγήτρια	Αναπληρωτής Καθηγητής

Ευχαριστίες

Με την ολοκλήρωση της παρούσας διατριβής αισθάνομαι την ανάγκη να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή μου κ. Ευθύμιο Αλέπη για τη συνολική του βοήθεια καθ' όλη τη διάρκεια του μεταπτυχιακού προγράμματος σπουδών και κυρίως κατά τη διάρκεια της εκπόνησης της παρούσας εργασίας.

Τέλος οφείλω ένα μεγάλο ευχαριστώ στη σύζυγό μου Έφη και στη κόρη μου Άννα για τη συνεχή στήριξη και βοήθειά τους σε αυτή μου την προσπάθεια.

Περίληψη

Στις μέρες μας η πληροφορία παράγεται και καταναλώνεται μέσω του διαδικτύου με μια άνευ προηγουμένου ταχύτητα. Οι ειδησεογραφικοί ιστότοποι παράγουν εκατοντάδες και σε μερικές περιπτώσεις χιλιάδες άρθρα ημερησίως. Η εγκυρότητα των κειμένων που αναρτώνται καθημερινά είναι κάτι που δεν ελέγχεται μέχρι και σήμερα από κανένα θεσμό και με κανέναν τρόπο, με αποτέλεσμα να υπάρχουν περιπτώσεις όπου διαδίδονται ειδήσεις που άλλοτε έχουν αναπαραχθεί από λάθος πληροφορίες ενός ρεπορτάζ, άλλοτε «ντυμένες» ως χιουμοριστικές και άλλοτε για να εξυπηρετήσουν συμφέροντα. Δεν υπάρχει κανένα μέτρο προστασίας του αναγνώστη για την πληροφορία που καταναλώνει και το κατά πόσο κάτι θα γίνει πιστευτό, εάν είναι ψευδές, επαφίεται στην παιδεία του καθενός πράγμα που μπορεί να γίνει σε μερικές περιπτώσεις ακόμη και επικίνδυνο.

Με την παρούσα εργασία υλοποιείται μία Web App εφαρμογή όπου αντλούνται άρθρα από διάφορα ειδησεογραφικά πρακτορεία από το διαδίκτυο, αποθηκεύονται και εμφανίζονται σε μια σελίδα της εφαρμογής. Από εκεί ο χρήστης έχει τη δυνατότητα να αξιολογεί την είδηση που διαβάζει και αυτή η αξιολόγηση να είναι ορατή από όλους τους χρήστες. Η εφαρμογή έχει τη δυνατότητα να αξιολογεί τις αξιολογήσεις σε έναν πρώτο βαθμό, βάσει ορισμένων κριτηρίων που αντλούνται κατά τη χρήση της εφαρμογής και στη συνέχεια να απορρίπτει ή να αποδέχεται την εκάστοτε αξιολόγηση. Τα κριτήρια που έχουν υλοποιηθεί είναι η τοποθεσία της είδησης, δηλαδή ο τόπος στον οποίο έχει λάβει χώρα η είδηση, η τοποθεσία του χρήστη μέσω geolocation, ο χρόνος που χρειάστηκε ο χρήστης για την ανάγνωση του κειμένου και εάν έφτασε ως το τέλος του κειμένου. Βάσει των ενεργειών του χρήστη η εφαρμογή μπορεί να προσθέτει ή να αφαιρεί πόντους και προνόμια στους χρήστες. Έτσι κάθε χρήστης δημιουργεί ένα reputation που του δίνει διαφορετική βαρύτητα κάθε φορά που αξιολογεί κάτι. Η ιδέα της ποντιοδότησης των χρηστών είναι βασισμένη στην ιδέα της επιταχυμένης κοινότητας Stack Exchange, όπου οι χρήστες αξιολογούνται με βάση τις απαντήσεις που δίνουν σε ερωτήσεις άλλων χρηστών για διάφορα θέματα όπως τον προγραμματισμό στο Stack Overflow, τη φωτογραφία στο Photography και πολλών άλλων communities.

Για την υλοποίηση της πλατφόρμας χρησιμοποιήθηκε για το back end API και το scrap engine η γλώσσα προγραμματισμού Python 3.7 και το framework FastAPI. Για το Front End χρησιμοποιήθηκε η γλώσσα προγραμματισμού Typescript και το framework Angular 14. Τέλος για την αποθήκευση των δεδομένων χρησιμοποιήθηκε η βάση δεδομένων SQLite.

Abstract

In today's world, information is produced and consumed at an unprecedented speed through the internet. News websites generate hundreds, and in some cases, thousands of articles daily. The accuracy of the content posted daily remains largely unchecked, both by any authority and in any way, resulting in cases where news is disseminated, sometimes based on incorrect information from a report, sometimes disguised as humorous, and sometimes to serve specific interests. There is no protection measure for the information consumed by the reader, and whether something is believed, if it is false, depends on the individual's education, which can sometimes be dangerous.

This project aims to create a web application where articles are gathered from various news agencies on the internet, stored, and displayed on a page of the application. Users could evaluate the news they read, and this evaluation is visible to all users. The application can also evaluate the evaluations to some extent, based on certain criteria that are gathered during the use of the application, and then accept or reject each evaluation. The criteria implemented include the location of the news, the user's location through geolocation, the time it took the user to read the text, and whether they reached the end of the text. Based on user actions, the application can add or remove points and privileges from users. This way, each user creates a reputation that gives different weight to their evaluations each time they evaluate a new article. The idea of rewarding users is based on the concept of the Stack Exchange network, where users are evaluated based on the answers, they provide to questions from other users on various topics such as programming on Stack Overflow, photography on Photography, and many other communities.

For the implementation of the platform, Python 3.7 and the FastAPI framework were used for the back-end API and the scrap engine. TypeScript and the Angular 14 framework were used for the front-end. Finally, the SQLite database was used for data storage.

1 Περιεχόμενα

1	Περιεχόμενα.....	6
2	Εισαγωγή.....	7
3	Backend – Μηχανή άντλησης, επεξεργασίας και αποθήκευσης δεδομένων (Scrap Engine)	8
3.1	Article Scrap	8
3.2	Article Meta Data	9
3.3	Similarity Check	12
3.4	Summarize.....	12
3.5	Article Locations	13
4	WebApp Backend & Frontend	15
4.1	Model View Controller.....	15
4.2	Register	16
4.3	End Point – Login	16
4.4	End Point – Get New Articles	17
4.5	Feed.....	17
4.6	Εμφάνιση άρθρου.....	19
4.7	Εμφάνιση Αξιολογήσεων	21
4.8	Νέα αξιολόγηση	22
4.9	Reputation	23
5	Desing Patterns	25
5.1	Retry Pattern.....	25
5.2	Health End Point Monitoring Pattern	27
5.3	Observer Design Pattern	28
5.4	Singleton Design Patter	29
6	Συμπεράσματα	31
7	Disclaimer	32
8	ΠΑΡΑΘΕΣΗ Άρθρο 2 - Νόμος 2121/1993 - Αντικείμενο του δικαιώματος.....	33
9	Βιβλιογραφία.....	34

2 Εισαγωγή

Το πρώτο στάδιο στην ανάπτυξη της Web App εφαρμογής ήταν η υλοποίηση μιας Scrap Engine, μιας μηχανής άντλησης, επεξεργασίας και αποθήκευσης δεδομένων. Αρχικά μέσα στη Scrap Engine αναπτύχθηκε μια αράχνη που μπορεί να κατεβάζει από τρεις ειδησεογραφικούς ιστότοπους άρθρα. Επιπλέον η scrap μηχανή έχει τη δυνατότητα να εξάγει πληροφορίες που περιέχονται μέσα στα άρθρα όπως οι τοποθεσίες, τα ονοματεπώνυμα και οι οργανισμοί. Η ίδια μηχανή έχει τη δυνατότητα με χρήση του NLP (Natural Language Process) να βρίσκει εάν ένα άρθρο είναι παρόμοιο με κάποιο άλλο και να τα χαρακτηρίζει. Με χρήση του NLP μπορεί να δημιουργεί περιλήψεις στα κείμενα των άρθρων, εξάγοντας τις σημαντικότερες προτάσεις. Ακόμη με χρήση του Google API βρίσκει τα latitude και longitude όλων των τοποθεσιών που εμφανίζονται μέσα στα άρθρα και στους τίτλους τους. Τέλος αποθηκεύει τα άρθρα καθώς και όλες τις πληροφορίες που έχει εξάγει σε μια βάση δεδομένων.

Σε δεύτερο στάδιο υλοποιήθηκε το back-end API και το front-end Angular site όπου τα ειδησεογραφικά άρθρα εμφανίζονται σε μια ροή ειδήσεων. Ο χρήστης προκειμένου να χρησιμοποιήσει την εφαρμογή καλείται να κάνει εγγραφή και σύνδεση. Κατά την είσοδο του στην εφαρμογή μπορεί να διαβάσει τα άρθρα που υπάρχουν στη βάση δεδομένων και να τα αξιολογεί. Για κάθε κίνηση του χρήστη μέσα στην εφαρμογή το back-end προσθέτει ή αφαιρεί πόντους δημιουργώντας έτσι ένα reputation για κάθε χρήστη που τον ακολουθεί σε κάθε εμφάνιση του username του.

Στο τελευταίο μέρος της εργασίας αναφέρονται τα πρότυπα σχεδίασης (Design Patterns) που χρησιμοποιήθηκαν και τι σκοπούς εξυπηρετούν στην ανάπτυξη λογισμικού.

3 Backend – Μηχανή άντλησης, επεξεργασίας και αποθήκευσης δεδομένων (Scrap Engine)

3.1 Article Scrap

Για την άντληση των ειδησεογραφικών κειμένων δημιουργήθηκε ένα python script που χρησιμοποιεί τη βιβλιοθήκη BeautifulSoup 4. Η βιβλιοθήκη BeautifulSoup είναι μια διάσημη βιβλιοθήκη ανοιχτού κώδικα της python που χρησιμοποιείται για να εξάγει πληροφορίες από HTML σελίδες και XML αρχεία.[1] Το Script επισκέπτεται τις αρχικές σελίδες τριών μεγάλων ελληνικών ειδησεογραφικών ιστοτόπων και βρίσκει το άρθρο που είναι στην κορυφή της σελίδας. Από εκεί εξάγει το link του άρθρου το επισκέπτεται και με χρήση της βιβλιοθήκης requests της python κατεβάζει όλη την HTML. Για την ορθή λειτουργία του προγράμματος προστέθηκε ένας User Agent σε κάθε request. Τα User agents μεταμφιέζουν ένα πρόγραμμα σε χρήστη. Προσθέτοντας το στα headers των request προς τη σελίδα ο server της σελίδας θεωρεί πως το request έχει γίνει από χρήστη και επιστρέφει τη σελίδα, σε διαφορετική περίπτωση επιστρέφει μία άδεια σελίδα HTML.

```
HEADERS = {
    'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/39.0.2171.95 Safari/537.36'
```

Παράδειγμα User Agent

Πολλά site διαθέτουν διάφορα λογισμικά και υπηρεσίες προστασίας ενάντια σε προγράμματα «αράχνες».

Από τη σελίδα του άρθρου εξάγονται οι παρακάτω πληροφορίες και αποθηκεύονται στη βάση δεδομένων στον πίνακα Article:

- ο τίτλος του άρθρου
- το link του άρθρου
- ο συγγραφέας
- το κείμενο με HTML Tags (για την μετέπειτα καλύτερη εμφάνιση του στην εφαρμογή)
- το link της φωτογραφίας
- η ημέρα και η ώρα της δημοσίευσης

Πριν την αποθήκευση του άρθρου υπολογίζεται και αποθηκεύεται ο χρόνος που απαιτείται για να διαβαστεί το άρθρο με την παραδοχή πως ο αναγνώστης μπορεί να διαβάζει 200 λέξεις ανά λεπτό. Τέλος αποθηκεύεται σε ξεχωριστή στήλη του πίνακα το κείμενο χωρίς HTML Tags και Java scripts, για την περαιτέρω ανάλυσή του. Για τον καθαρισμό του κειμένου χρησιμοποιήθηκαν regular expressions.

Τα Regular Expressions είναι ένα είδος μοτίβου σε κείμενο που συναντάται σε πολλές σύγχρονες εφαρμογές και γλώσσες προγραμματισμού. Χρησιμοποιούνται για να επιβεβαιώσουμε εάν το input του χρήστη είναι αντίστοιχο του pattern (π.χ. email), για να βρούμε κείμενα μέσα σε μεγαλύτερα κείμενα καθώς και για να αντικαθίστούμε κείμενα με άλλα κείμενα ή να προχωράμε σε αναδιάταξη του κειμένου.[2]

```
re.sub(r"<script\b[^<]*(?:?!</script>-[\^<]*)*</script>", "", _a.body_html, flags=re.IGNORECASE)
```

Regular expression που βρίσκει και αφαιρεί java scripts από κείμενο

Ανάλυση regular expression:

- <script** # το regular expression <script βρίσκει κάθε κομμάτι κειμένου που ξεκινάει με <script.
- \b** # το regular expression \b ονομάζεται word boundary και διασφαλίζει πως το κείμενο <script θα ταιριάζει μόνο εάν είναι ολόκληρο και όχι ως μέρος άλλου κειμένου.
- [^<]*** # το regular expression [^<]* βρίσκει όλους τους χαρακτήρες εκτός του

	# χαρακτήρα <.
(?:	# το regular expression (?: δηλώνει πως ξεκινάει το γκρουπάρισμα.
(?!</script>)	# διασφαλίζει πως το περιεχόμενο δεν ταιριάζει με την ετικέτα κλεισίματος # </script>.
<[<]*	# το regular expression βρίσκει <[<]* όλους τους χαρακτήρες εκτός του # <. Δηλαδή Βρίσκει όλους τους χαρακτήρες από το <script> tag μέχρι και # το </script> tag.
)*	# το regular expression)* δηλώνει πως τελειώνει το γκρουπάρισμα.
</script>	# το regular expression </script> βρίσκει το tag </script> που με αυτό θα # πρέπει να τελειώνει το κείμενο.

```
re.sub('<[<]+>', '', html_text)
```

Regular expression που βρίσκει και αντικαθιστά HTML Tags σε κείμενο

Ανάλυση regular expression:

<	# το κομμάτι που θα ταυτιστεί με το regular expression θα πρέπει να # ξεκινάει με το σύμβολο <.
[^>]	# το regular expression [^>] ταιριάζει έναν ή περισσότερους χαρακτήρες # εκτός του > που είναι το tag που κλείνει τα HTML tags. Οι αγκύλες [] # δηλώνουν ένα σύνολο χαρακτήρων.
+	# δηλώνει πως μέσα στις αγκύλες θα πρέπει να υπάρχει τουλάχιστον # ένας χαρακτήρας.
>	# το κομμάτι που θα ταυτιστεί με το regex θα πρέπει να τελειώνει με τον # χαρακτήρα > που κλείνει κάθε HTML Tag.

id	body	site	title	original_link	author	date	image	body_html	reading_time
292	Η Ελλάδα κινείτ	https://...	«Hot dry windy», ο καιρός που ανησυχεί τους ...	https://...	Χρήστος Λογαράς	2023-07-15	https://...	<div class="post-body main-conte...	7
293	Οι ηθοποιοί το	https://...	ΗΠΑ: Το Χόλγουντ παραλύει	https://...	Newsroom	15/07/2023	https://...	<div class="content-col">...	4
294	«Βρόζα» η ...	http://...	Επιμένει ο καύσωνας: Με 44άρια η ερχόμενη ...	https://...	Σύνταξη IN	15/07/2023 15:50	https://...	<div ...	5
295	Σε εξέλιξη βρίο	https://...	Ποπάς και Τζουμάκας στην κούρσα διαδοχής το...	https://...		2023-07-15	https://...	<div class="post-body main-conte...	12
296	-Παπούς εναντ	https://...	Αποκάλυψη – Αυτή την Κυριακή στο Documenta: ...	https://...	Newsroom	15/07/2023	https://...	<div class="content-col">...	1
297	Ενέο με 11 βαθ	http://...	Στο μέγιστο επίπεδο επικινδυνότητας έξι πόλεις - ...	https://...	Σύνταξη IN	22/07/2023 12:42	https://...	<div ...	4
298	Το Έκτακτο ...	https://...	Νέο έκτακτο της ΕΜΥ: Ποιες περιοχές θα «ψήσει»...	https://...		2023-07-22	https://...	<div class="post-body main-conte...	2
299	Ανείπωτη ...	https://...	Κρίσιμη η κατάσταση στη Ρόδο – Εκκένωση δύο ...	https://...	Newsroom	22/07/2023	https://...	<div class="content-col">...	1
300	Pur και ...	https://...	Γιατί χάνονται τόσες ζωές σε θάλασσες και πισίνε...	https://...	Γιάννης Ζωιτός	2023-07-22	https://...	<div class="post-body main-conte...	11
301	Εκτός ελέγχου ε	http://...	Νέο μήνυμα από το 112 για εκκενώσεις χωριών ...	https://...	Σύνταξη IN	22/07/2023 12:51	https://...	<div ...	3

Παράδειγμα αποθήκευσης άρθρου στη βάση δεδομένων

3.2 Article Meta Data

Μετά την αποθήκευση του άρθρου η εφαρμογή εξάγει και αποθηκεύει κάποια επιπλέον στοιχεία, στον πίνακα Article Meta Data, που σχετίζονται με το άρθρο βάσει των ήδη αποθηκευμένων κειμένων.

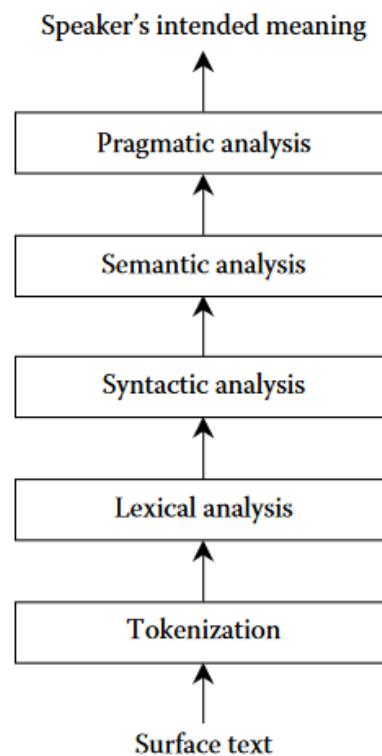
Με χρήση της βιβλιοθήκης Spacy και του εγκατεστημένου μοντέλου el_core_news_lg που εξειδικεύεται σε ελληνικά κείμενα, η εφαρμογή είναι σε θέση μέσω του Natural Language Process και συγκεκριμένα του Named-entity Recognition να εξάγει από τα κείμενα περιοχές – τοποθεσίες και ονοματεπώνυμα.

Κατά το Natural Language Process οι προτάσεις ενός κειμένου αναλύονται πρώτα ως προς τη συντακτική τους δομή, προσφέροντας μια τάξη και μια δομή που είναι πιο κατάλληλη για ανάλυση ως προς τη σημασιολογία, δηλαδή την κυριολεκτική τους σημασία, και στη συνέχεια ακολουθεί ένα στάδιο πραγματικής ανάλυσης, κατά το οποίο καθορίζεται η σημασία της έκφρασης ή του κειμένου στο πλαίσιο του. Αυτό το τελευταίο στάδιο συνήθως θεωρείται ότι ασχολείται με την ομιλία, ενώ τα δύο προηγούμενα ασχολούνται γενικά με θέματα προτάσεων. Αυτή η προσπάθεια να γίνει αντιστοίχιση μεταξύ μιας διάκρισης στρωματοποίησης (συντακτική,

σημασιολογική και πραγματική χρήση) και μιας διάκρισης ως προς το επίπεδο λεπτομέρειας (πρόταση έναντι διαλόγου) μπορεί να προκαλέσει κάποια σύγχυση στη σκέψη σχετικά με τα θέματα που συνδέονται με την επεξεργασία της φυσικής γλώσσας. Είναι επίσης ευρέως αναγνωρισμένο ότι στην πραγματικότητα δεν είναι τόσο εύκολο να χωριστεί η επεξεργασία της γλώσσας σε καθαρά διακριτικά κουτιά που αντιστοιχούν σε κάθε στάδιο. Ωστόσο, μια τέτοια διάκριση χρησιμεύει στη διδασκαλία και αποτελεί επίσης τη βάση για αρχιτεκτονικά μοντέλα που καθιστούν την ανάλυση της φυσικής γλώσσας πιο εφικτή από πλευράς μηχανικής λογισμικού. [4]

Τα στάδια ανάλυσης στην επεξεργασία της φυσικής γλώσσας είναι τα εξής:

- **Tokenization:** Αυτό το στάδιο αντιστοιχεί στον διαχωρισμό του κειμένου σε λεκτικές μονάδες, όπως λέξεις ή σύμβολα. Σε αυτό το στάδιο, το κείμενο χωρίζεται σε "tokens" που μπορούν να είναι λέξεις, αριθμοί, σημεία στίξης κλπ.
- **Λεξική ανάλυση (Lexical Analysis):** Σε αυτό το στάδιο, κάθε token αναλύεται για να προσδιοριστεί η λέξη και η σημασία της. Αυτό συμπεριλαμβάνει την αναγνώριση της μορφολογίας, τον καθορισμό του καταλόγου των λέξεων, και τη σύνδεση με λεξικές βάσεις δεδομένων.
- **Συντακτική ανάλυση (Syntactic Analysis):** Σε αυτό το στάδιο, η δομή των προτάσεων αναλύεται στη βάση των συντακτικών κανόνων της γλώσσας. Αναλύονται οι σχέσεις μεταξύ λέξεων και οι συντακτικές δομές, όπως φράσεις και προτάσεις.
- **Σημασιολογική ανάλυση (Semantic Analysis):** Σε αυτό το στάδιο, γίνεται ανάλυση της σημασίας των προτάσεων, λαμβάνοντας υπόψη τον τρόπο με τον οποίο οι λέξεις συνδυάζονται και αλληλοεπιδρούν. Αυτό συμπεριλαμβάνει την αναγνώριση συνώνυμων, αντωνυμιών και άλλων σημασιολογικών πτυχών.
- **Πραγματική ανάλυση (Pragmatic Analysis):** Σε αυτό το στάδιο, γίνεται ανάλυση του νοήματος της εκφώνησης ή του κειμένου στο συγκεκριμένο πλαίσιο. Αυτό συμπεριλαμβάνει την κατανόηση της πραγματικής σημασίας, του σκοπού, της συμπεριφοράς και του περιβάλλοντος στο οποίο χρησιμοποιείται η φράση ή το κείμενο.



Στάδια Ανάλυσης του Natural Language Process [4]

Το Named-entity recognition (NER) είναι ένα βαθύ συνελκτικό εκπαιδευμένο νευρωνικό δίκτυο (deep convolutional pre trained neural network) που εξάγει οντότητες ονομάτων από μη δομημένο κείμενο και στη συνέχεια τις κατηγοριοποιεί σε προκαθορισμένες κατηγορίες π.χ. ονόματα, περιοχές, οργανισμοί, τιμές, ποσοστά κ.α. [3]

Accuracy Evaluation		
TOKEN_ACC	Tokenization	1.00
TOKEN_P		1.00
TOKEN_R		1.00
TOKEN_F		1.00
POS_ACC	Part-of-speech tags (coarse grained tags, Token.pos)	0.96
MORPH_ACC	Morphological analysis	0.91
MORPH_MICRO_P		0.96
MORPH_MICRO_R		0.96
MORPH_MICRO_F		0.96
SENTS_P	Sentence segmentation (precision)	0.92
SENTS_R	Sentence segmentation (recall)	0.93
SENTS_F	Sentence segmentation (F-score)	0.93
DEP_UAS	Unlabeled dependencies	0.88
DEP_LAS	Labeled dependencies	0.85
LEMMA_ACC	Lemmatization	0.90
TAG_ACC	Part-of-speech tags (fine grained tags, Token.tag)	0.93
ENTS_P	Named entities (precision)	0.75
ENTS_R	Named entities (recall)	0.79
ENTS_F	Named entities (F-score)	0.77

Πίνακας ακρίβειας για το μοντέλο `el_core_news_lg` [5]

```

Editable Code spaCy v3.0 - Python 3 - via Binder
import spacy

nlp = spacy.load("en_core_web_sm")
doc = nlp("Apple is looking at buying U.K. startup for $1 billion")

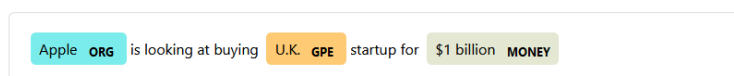
for ent in doc.ents:
    print(ent.text, ent.start_char, ent.end_char, ent.label_)

RUN

```

TEXT	START	END	LABEL	DESCRIPTION
Apple	0	5	ORG	Companies, agencies, institutions.
U.K.	27	31	GPE	Geopolitical entity, i.e. countries, cities, states.
\$1 billion	44	54	MONEY	Monetary values, including unit.

Using spaCy's built-in [displaCy visualizer](#), here's what our example sentence and its named entities look like:



Παράδειγμα χρήσης NER [3]

3.3 Similarity Check

Με την ολοκλήρωση της εισαγωγής ενός νέου άρθρου στη βάση δεδομένων ξεκινάει ο έλεγχος ομοιότητας με τα ήδη αποθηκευμένα άρθρα. Εκεί ελέγχεται αρχικά ο τίτλος του άρθρου σε σύγκριση με τους τίτλους των τελευταίων 30 εισαγμένων άρθρων.

Για την παραπάνω υλοποίηση χρησιμοποιήθηκε η βιβλιοθήκη της python Scikit Learn και συγκεκριμένα η κλάση `sklearn.feature_extraction.text.TfidfVectorizer`. Η κλάση αυτή μετατρέπει οποιασδήποτε μορφής κείμενο σε ένα matrix με TF-IDF χαρακτηριστικά (Term Frequency - Inverse Document Frequency). [6]

Τα TF-IDF χαρακτηριστικά είναι μια στατιστική απεικόνιση που αξιολογεί πόσο σχετική είναι μια λέξη σε ένα κείμενο. Αυτό πραγματοποιείται πολλαπλασιάζοντας δύο metrics: Το πόσες φορές εμφανίζεται μια λέξη σε ένα έγγραφο και η αντίστροφη συχνότητα της λέξης στο σύνολο των λέξεων. [7]

Με τη μέθοδο cosine similarity της ίδιας κλάσης υπολογίζεται ένα σκορ μεταξύ δύο κειμένων στο κατά πόσο είναι όμοια. Η παραπάνω μέθοδος χρησιμοποιεί τον παρακάτω τύπο: [7]

$$\text{similarity}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

Στη συνέχεια ελέγχονται τα Meta Data των άρθρων. Πραγματοποιείται έλεγχος ως προς τα πρόσωπα και τις περιοχές που εμφανίζονται στον τίτλο αλλά και στο κείμενο του άρθρου, με μεγαλύτερη βαρύτητα να δίνεται στα στοιχεία του τίτλου. Από τις συγκρίσεις προκύπτει μια βαθμολογία. Ένας βαθμός δίνεται για κάθε ίδιο μέρος και για κάθε ίδιο πρόσωπο που συναντώνται στα κυρίως μέρη του κειμένου. Πέντε βαθμοί δίνονται για κάθε ίδιο μέρος και για κάθε ίδιο πρόσωπο που συναντώνται στους τίτλους των δύο άρθρων.

Οι βαθμολογίες αξιολογούνται βάσει τιμών που ορίστηκαν έπειτα από εκτενείς δοκιμές πάνω στο μοντέλο. Για να χαρακτηριστεί ένα άρθρο όμοιο με κάποιο άλλο θα πρέπει:

- είτε να επιστρέψει η cosine similarity σκορ μεγαλύτερο του 0,2 και μικρότερο του 0,99
- είτε ο έλεγχος των meta data να δώσει σκορ μεγαλύτερο του 15

Η πληροφορία αποθηκεύεται στον πίνακα Similarity που έχει το id του νέου άρθρου καθώς και το id του άρθρου που βρέθηκε ως παρόμοιο. Σε περίπτωση που το άρθρο είναι μοναδικό δίνεται η τιμή -1.

Τέλος εάν το άρθρο που βρέθηκε να είναι παρόμοιο δεν είναι χαρακτηρισμένο ως μοναδικό τότε το άρθρο που μόλις εισάχθηκε συνδέεται ως παρόμοιο με το αρχικό άρθρο, πράγμα που διατηρεί την είδηση που βγήκε πρώτη ως μοναδική.

id	FK_article_id	unique	FK_similar_article	Title	Found Similar
9	13	1	-1	Δεκανίκιο, τροχονόμοι και χωροφύλακες	NULL
10	14	1	-1	Συνάντηση Ερντογάν-Άσαντ Πολιτικός θρίαμβος Πούτιν, ήττα των Αμερικανών	NULL
11	15	1	-1	Ο σκιπηλός στρατηγός	NULL
12	16	1	-1	Ο Ερντογάν... μαθαίνει τον αντίπαλό του - Πώς καταστά το Σύστημα για να παραμείνει στην ...	NULL
13	17	1	-1	Οι δύο τακτικές για τις εκλογές	NULL
14	18	1	-1	Σκηνικό εμφυλίου στη Βραζιλία - Εισβολή υποστηρικτών του Μπολσονάρο σε κυβερνητικά κτίρια	NULL
15	19	0	18	Χάος στη Βραζιλία: Εισβολή σε κτίρια από υποστηρικτές του Μπολσονάρο	Σκηνικό εμφυλίου στη Βραζιλία - Εισβολή υποστηρικτών του Μπολσονάρο σε κυβερνητικά ...
16	20	0	18	Βραζιλία: Χάος με οπαδούς του Μπολσονάρο να εισβάλουν σε κυβερνητικά κτίρια	Σκηνικό εμφυλίου στη Βραζιλία - Εισβολή υποστηρικτών του Μπολσονάρο σε κυβερνητικά ...
17	21	1	-1	«Μια προαναγγελθείσα επίθεση»: Πώς η ρητορική και μετά η σιωπή του Μπολσονάρο έστρωσε τ...	NULL
18	22	1	-1	Τα πάθη της Βουλής: Παικτες παγιδεύει η απογραφή, ποιοι αποχαιρετούν τα «άλλα»	NULL
19	23	1	-1	Η φωτογραφία της ημέρας - Πισωμένες από το χέρι Παπαχαράλημους, Δροσάκη και ...	NULL
20	24	0	21	Πώς ο Μπολσονάρο γίνεται και... αμερικάνικο πρόβλημα - Ποιοι πιέζουν τον Μπάιντεν	«Μια προαναγγελθείσα επίθεση»: Πώς η ρητορική και μετά η σιωπή του Μπολσονάρο ...

Πίνακας Similarity

3.4 Summarize

Επιπλέον η εφαρμογή δημιουργεί μια «περίληψη» του κειμένου του άρθρου, για την καλύτερη εμφάνιση του feed της εφαρμογής. Για την υλοποίηση της παραπάνω διαδικασίας χρησιμοποιήθηκε η βιβλιοθήκη της python nltk που ελέγχει όλες τις προτάσεις ενός κειμένου

και βρίσκει τις σημαντικότερες βάσει των λέξεων που περιέχουν. Στη συνέχεια επιστρέφει τις σημαντικότερες προτάσεις. [8]

FK_article_id	summary
Filter	Filter
74	Μέχρι στιγμής 45 χώρες έχουν προσφέρει βοήθεια στην Τουρκία.Εξάλλου, όπως μετέδωσε το συριακό κρατικό πρακτορείο ειδήσεων, στη Συρία έχουν σκοτωθεί τουλάχιστον 320 άνθρωποι λόγ...
75	pic.twitter.com/AuG7EVCm74— Seher Sultan (@sehersultan1912) February 6, 2023 Σοκ προκαλούν και άλλα βίντεο από τον ισχυρό σεισμό:Malatya'da meydana gelen son 7.7 siddetindeki ...
76	Σκληρές χάους στην Τουρκία μετά και τη νέα σεισμική δόνηση 7,5 Ρίχτερ ΤοΒΗΜΑ Team 06.02.2023, 12:40 Νέος, ισχυρότατος σεισμός, μεγέθους 7,5 βαθμών χτύπησε την Τουρκία, ...
77	Επτά μέλη της οικογένειάς μας είναι κάτω από τα ερείπια», δήλωσε στο Γαλλικό Πρακτορείο ο Μουχτίν Ορακτζέ, παρακολουθώντας την επιχείρηση έρευνας και διάσωσης μπροστά σε ...
78	Μετακινήθηκε η Αραβική τεκτονική πλάκα«Αυτό που ονομάζουμε Αραβική πλάκα μετακινήθηκε σε λίγα δευτερόλεπτα για πάνω από 3 μέτρα σε μήκος τουλάχιστον 150 χιλιομέτρων» υπογράμμισε
79	Ο κύριος σεισμός είχε μεγάλη διάρκεια, κράτησε σχεδόν ένα λεπτό, σύμφωνα με ανταποκριτή του πρακτορείου ειδήσεων Ρόιτερς.«Εκκλιση για διεθνή βοήθεια«Όλες οι ομάδες μας τήθηκαν σε ...
80	Περίπου 9.000 άνθρωποι συμμετείχαν στις επιχειρήσεις, ενώ μέχρι στιγμής 45 χώρες έχουν προσφέρει βοήθεια στην Τουρκία, σημείωσε ο ίδιος, κάνοντας λόγο για τη μεγαλύτερη καταστροφή
81	Στο δήμο Κορυδαλλού θα είναι όλα ανοιχτά εκτός από τα παρακάτω:Δημοτικά 9ο/ 10ο/ 5ο /12ο Νηπιαγωγείο 3ο/7ο /8ο / 18ο / 19ο /21ο Γυμνάσιο 4ο / 5ο Στο ΔΥΤΙΚΟ ΤΟΜΕΑ θα λειτουργήσει.
82	Gaziantep, Turkey τη pic.twitter.com/WYGcnuYFe3— Ahmet Duran Hergün (@AhmetDuranHergun) February 2, 2023 Πρόκειται για το σύνολο των μοναδικών ψηφιδωτών της ελληνιστικής πόλης
83	Mina Moustaka 08.02.2023, 16:54 Το δικαστήριο αποφάσισε να σθώσει ομόφωνα τον Πέτρο Φιλιππίδη για την κατηγορία του βιασμού, ενώ παράλληλα τον καταδίκασε, επίσης ομόφωνα,
84	Ομόφωνα ένοχο έκρινε το Μικτό Ορκωτό Δικαστήριο τον Πέτρο Φιλιππίδη για τις δύο απόπειρες βιασμού, ενώ τον σθώωσε ομόφωνα για την κατηγορία του τετελεσμένου βιασμού.Δίκη Φιλιππίδ.
85	Δεν υπήρχε βία στην υπόθεση Φιλιππίδη».Απευθυνόμενος μάλιστα στην έδρα είπε «τόσο μεγάλη αγωνία έχετε να πείτε ότι καταδικάσατε τον Πέτρο Φιλιππίδη με τόσο μεγάλη ποινή;» ...
86	ΤοΒΗΜΑ Team 08.02.2023, 20:03 Ποινή κάθειρξης έξι ετών για κάθε μία από τις δύο πράξεις της απόπειρας βιασμού για τις οποίες κρίθηκε ένοχος ο Πέτρος Φιλιππίδης πρότενε η ...
87	Απλά περιμένουμε την πρόσβαση για να το κάνουμε».Μητσοτάκης: Στο Ευρωπαϊκό Συμβούλιο η βοήθεια για την ΣυρίαΤο θέμα της διεθνούς βοήθειας προς την Συρία έθεσε εξ αρχής, μιλώντας σ...
88	Σεισμός στην Τουρκία: Γιατί τα κτίρια έπεσαν σαν τραπουλόχαρτα - Τι είναι το «rapack collapse» Τους 20.000 αγγίζει ο τραγικός απολογισμόςΤους 20.000 αγγίζει ο τραγικός απολογισμός του

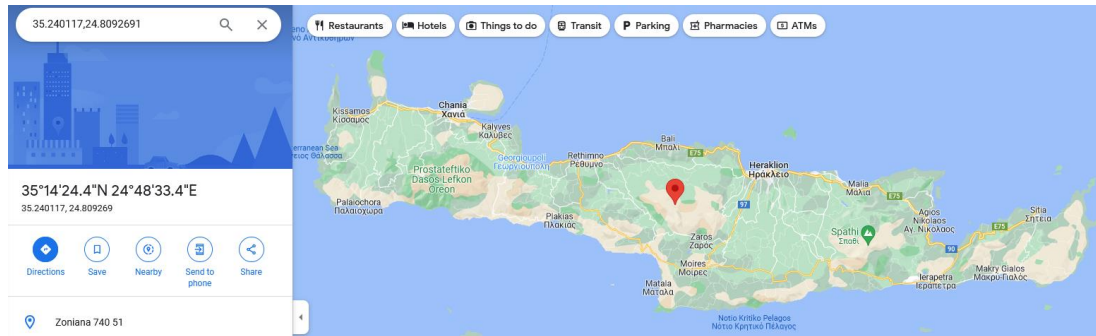
Πίνακας Summary

3.5 Article Locations

Για όλες τις τοποθεσίες που προστέθηκαν στον πίνακα Meta Data πραγματοποιείται ανεύρεση των συντεταγμένων Latitude and Longitude της περιοχής με κλήσεις στο API της Google.[8] Η αποθήκευση των συντεταγμένων βάσει του λεκτικού της περιοχής γίνεται για να μην καλεί η εφαρμογή συνεχώς το Google API για τις ίδιες περιοχές. Έτσι πριν την κλήση του API πραγματοποιείται έλεγχος στα ήδη αποθηκευμένα places και εάν η τοποθεσία υπάρχει τα latitude και longitude έρχονται από την βάση και όχι από το API.

id	place	latitude	longitude
Filter	Filter	Filter	Filter
70	Αθήνας	37.9838096	23.7275388
71	Δημοκρατική	36.4412335	28.2176119
72	Βραζιλία».Η	-14.235004	-51.92528
73	Παναμά	8.537981	-80.782127
74	Καναδάς	56.130366	-106.346771
75	Μεξικό	23.634501	-102.552784
76	Μεξικού	23.634501	-102.552784
77	Πορτογαλία	39.399872	-8.224454
78	Πρεσπών	40.7637355	21.1414226
79	Μόσχα	55.755826	37.6173
80	Ελλάδα	39.074208	21.824312
81	Αθήνα	37.9838096	23.7275388
82	Ουάσιγκτον	38.9071923	-77.0368707
83	Κύπρο	35.126413	33.429859
84	Κρήτης	35.240117	24.8092691
85	Βόσπορο	41.2213125	29.1290157
86	Λευκωσία	35.1855659	33.3822764

Πίνακας Google Places



Παράδειγμα ορθής λειτουργίας γραμμής 84

4 WebApp Backend & Frontend

4.1 Model View Controller

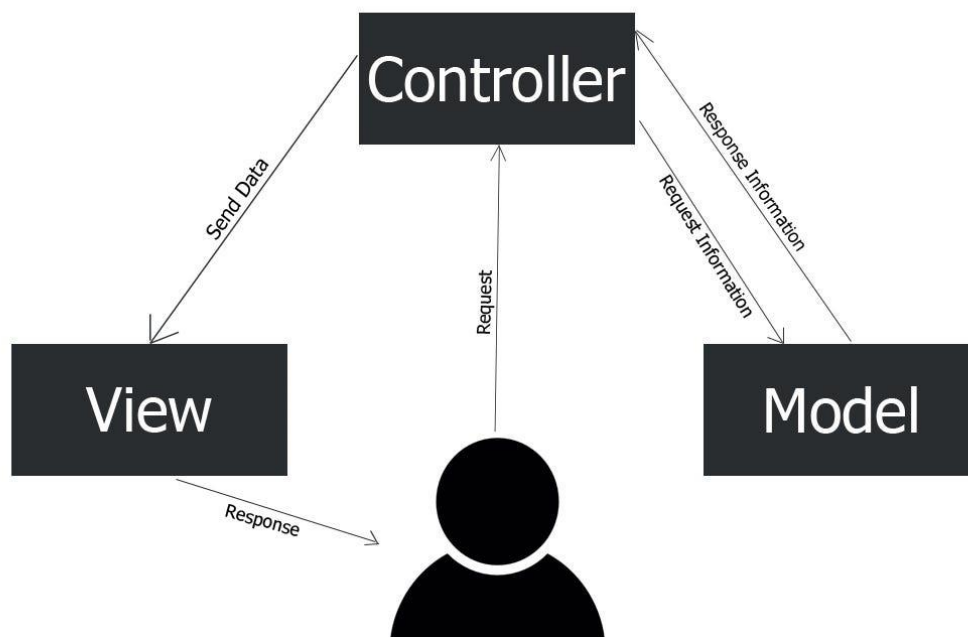
Το Back end Fast API καθώς και το Angular Front end έχουν δημιουργηθεί βάσει του αρχιτεκτονικού pattern για Web App εφαρμογές **Model View Controller (MVC)**.

Για την υλοποίηση του MVC pattern θα πρέπει η εφαρμογή να χωριστεί σε τουλάχιστον τρία κομμάτια:

- Στα Models που θα περιέχουν όλη την απεικόνιση των δεδομένων που χρησιμοποιεί ο χρήστης.
- Τα Views που χρησιμοποιούνται για να αποδώσουν κομμάτια των models και να τα εμφανίσουν στον χρήστη.
- Και τους Controllers που επεξεργάζονται εισερχόμενα αιτήματα, εκτελούν διεργασίες πάνω στα Models και επιλέγουν τα Views που θα εμφανιστούν στον χρήστη.

Κάθε κομμάτι του MVC pattern είναι πολύ καλά καθορισμένο και λειτουργεί αυτόνομα. Αυτό στη βιβλιογραφία αναφέρεται και ως «separation of concerns». Έτσι με τον ξεκάθαρο διαχωρισμό μεταξύ των παραπάνω κομματιών η εφαρμογή είναι ευκολότερα διαχειρίσιμη και επεκτάσιμη με την πάροδο του χρόνου, όσο μεγάλη και εάν γίνει. [10]

Model-View-Controller



Διάγραμμα MVC Pattern [15]

Με χρήση της βιβλιοθήκης sqlalchemy 0.0.8 και του ORM sqlalchemy 1.4 κάθε οντότητα της βάσης δεδομένων αναπαρίστανται ως κλάση στο back end της εφαρμογής στο αρχείο models.py. Επιπλέον ως views θεωρούνται όλα τα endpoints που παίρνουν data και επιστρέφουν data. Τέλος έχουν δημιουργηθεί controllers για κάθε διεργασία της εφαρμογής (π.χ. similarity_controller, auth_controller, reviews_controller κ.α.). Με την εκκίνηση του API δημιουργείται η βάση, όλοι οι πίνακες και όλες οι εξαρτήσεις εάν δεν υπάρχουν ήδη. [13][14]

Αντίστοιχα στο Front End κάθε Model αναπαρίστανται ως Typescript Interface ή Typescript Class. Ως Views θεωρούνται όλα τα components που περιέχουν HTML και CSS αρχεία, τα

οποία συνθέτουν όλες τις σελίδες της web εφαρμογής. Τέλος ως Controllers θεωρούνται όλα τα typescript αρχεία που περιέχουν λογική καθώς και τα services που καλούν, επεξεργάζονται και τροποποιούν τα data όπου και όποτε χρειάζεται.

4.2 Register

Για την αυθεντικοποίηση των χρηστών δημιουργήθηκε ένα endpoint εγγραφής νέων χρηστών όπου ο χρήστης καλείται να δώσει το email του, έναν κωδικό και το ονοματεπώνυμό του. Η εφαρμογή αποθηκεύει τους χρήστες στη βάση δεδομένων στον πίνακα User. Ο κωδικός γίνεται Hash με τον αλγόριθμο HS256.

user_id	first_name	last_name	email	password
16	Efi	Nanopoulou	efi2@gmail.com	\$2b\$12\$x/rFA7a7sL5XNrAuj7L6jOKZ9dDAJbdfLTL8V3IobA6w0p4FjkmBG
17	Κώστας	Ντζαχρήστος	kostas@gmail.com	\$2b\$12\$Xoaf6ec9.MRuqoF8b6cdTu94ZtIf4cO6ScHdqygcSHEzyVKFMk5zW
18	Maria	Papadopoulou	maria@hotmail.com	\$2b\$12\$PBmvf4h60MVgpnM/uzd/VOs34yr8oSN3jTnapwJTCeeKNvndppquKC

Πίνακας user



Εγγραφή νέου χρήστη

Email

Όνομα

Επώνυμο

Password

Verify Password

Εγγραφή

Σελίδα εγγραφής νέου χρήστη

4.3 End Point – Login

Για την είσοδο του χρήστη στην εφαρμογή δημιουργήθηκε το endpoint Login όπου ο χρήστης με το email του και τον κωδικό που έχει δώσει κατά την εγγραφή του γίνεται authenticate στο API και αποκτά πρόσβαση σε όλα τα endpoints.

Η εφαρμογή δημιουργεί ένα JWT (Json Web Token) το οποίο περιέχει το user id του χρήστη και το timestamp λήξης του TOKEN. Το JWT_SECRET_KEY για την αποκωδικοποίηση του TOKEN είναι αποθηκευμένο στο backend στο αρχείο creds.json.

Encoded PASTE A TOKEN HERE

```
eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJleHAiOjE2OTYxNDgyMDIsInN1YiI6IjE2In0.BLNEH1FhjEyTDRP-1KU0H2af4HaHb84Vkf6Q0x0fAN8
```

Decoded EDIT THE PAYLOAD AND SECRET

HEADER: ALGORITHM & TOKEN TYPE

```
{
  "alg": "HS256",
  "typ": "JWT"
}
```

PAYLOAD: DATA

```
{
  "exp": 1696148202,
  "sub": "16"
}
```

Αποκωδικοποίηση JWT

Το TOKEN αποθηκεύεται στο Front End στην περιοχή Cookies και χρησιμοποιείται μέσω ενός interceptor για κάθε κλήση στο API.



Σύνδεση

Email *

Password *

Είσοδος

Σελίδα εισόδου χρήστη στην εφαρμογή

4.4 End Point – Get New Articles

Η μέθοδος `get_new_articles` εκτελείται αυτόματα κάθε φορά που το API ξεκινάει και στη συνέχεια με τη βοήθεια του `@repeat_every(seconds=N)` decorator εκτελείται ξανά κάθε N δευτερόλεπτα. Το συγκεκριμένο end point καλεί την αράχνη του Scrap Engine και κατεβάζει νέα άρθρα. [11]

4.5 Feed

Το End Point Feed δέχεται ως παραμέτρους την τιμή `limit` που είναι ένας ακέραιος αριθμός και την τιμή `user` που είναι το JWT Token του χρήστη από το Front End. Επιστρέφει τα τελευταία άρθρα με βάση την τιμή του `limit` (default = 10) για την εμφάνισή τους στο Front End. Επιστρέφει τα δεδομένα των Άρθρων που έχουν γίνει Join με τον πίνακα `Similarity` και τον πίνακα `Summary`.


```

"Article": {
},
"Similarity": {
  "FK_similar_article": 381,
  "FK_article_id": 384,
  "id": 286,
  "unique": false
},
"Summary": {
  "id": 289,
  "FK_article_id": 384,
  "summary": "Όσοο, νεότερες πληροφορίες αναφέρουν ότι οι πυροβόλας και οι καλάρτες καθάρων να απεγκλωβιστούν και είναι καλά στην υγεία τους.Να σημειωθεί ότι στη Μονή Υγείας αναπτύχθηκαν νωρίτερα πολλά σχήματα βαρέας τύπου σε μία προσπάθεια να προστατευθούν τη Μονή ενό δυνάμη της πυροβλατικής είχε παραταθεί περιμετρικά της Μάρου για τον λόγο.Εκκλιση από τον Ερυθρό ΣταυρόΕκκλιση προς τους πολίτες έκανε ο Ερυθρός Σταυρός Ρόδου, καθώς χρειάζονται φαρμακευτικά υλικά.Ειδικότερα χρειάζονται:Βαζελίνουχες και αποστειρωμένες γάζεςΑλοιφές εγκαυμάτωνΦυσιολογικό ορό 500mlΕλαστικούς επιδέσμοιΣτρώματα φυσικά σκελετοειδή Ισοτονικά φακελάκια, μπάρες θημηριακών, σπασμούς καφέδεςΚατά τη διάρκεια της Μονής, οι πολίτες θέλουν να τα προλάβουν στις αποθήκες του Ερυθρού στο παλιό κοιμητήριο Τηλ. επικοινωνίας: 6977919822εκ περιγραφής δικαστικής ραμματος από τον ΔΕΔΗΕ Ο ΔΕΔΗΕ, στο πλαίσιο της ολοκληρωμένης και συνεχούς επιτήρησης του κοινού, γνωστοποιεί ότι εξαιτίας της πυκνότητας που βρίσκεται σε εξέλιξη στο νησί κοντά στο Δίκτυο Υγείας Τάσης, τήθηκε εκτός ο ΘΗΣ (Θερμοηλεκτρικός Σταθμός) Ν. Ρόδου και κρίθηκε αναγκαίο να προγραμματισθούν για σήμερα, Τετάρτη 22 Ιουλίου 2023, έως το βράδυ, εκ περιτροπής διακοπές ηλεκτροδότησης περιορισμένης έκτασης και χρονικής διάρκειας έως 1 ώρα, με σκοπό την εξασφάλιση της ομαλής λειτουργίας του ηλεκτρικού συστήματος.Στο πλαίσιο της συντονισμένης προσπάθειας για την εύρυθμη λειτουργία του ηλεκτρικού συστήματος της Ρόδου, ο ΔΕΔΗΕ παρακαλεί θερμά τους καταναλωτές να περιορίσουν την άσκοπη κατανάλωση ηλεκτρικής ενέργειας, ειδικά για τις ώρες αιχμής 12:00 - 15:00 και 17:00 - 22:00.Τις συγκεκριμένες ώρες, συστήνεται:Να αποφεύγεται η χρήση των ηλεκτροβόρων συσκευών (πλυντήριο, φούρνος, θερμοσίφωνας κ.λπ.)"
}
    
```

Παράδειγμα Feed Response



Σελίδα Feed

Στη σελίδα ο χρήστης βλέπει τα τελευταία δέκα άρθρα. Εμφανίζονται ο τίτλος ,το link της πηγής, ο συντάκτης, εάν το κείμενο είναι μοναδικό ή όχι. Εάν δεν είναι μοναδικό υπάρχει link με το άρθρο που φαίνεται να είναι παρόμοιο και έχει αποθηκευτεί νωρίτερα. Επιπλέον εμφανίζεται το άθροισμα των αξιολογήσεων και ο αριθμός των αξιολογήσεων. Τέλος υπάρχει η φωτογραφία όπως έχει αντληθεί από την πηγή και οι παράγραφοι που το back end έχει χαρακτηρίσει ως σημαντικότερες και έρχονται ως «περιλήψη». Από το κουμπί  ο χρήστης μπορεί να ανοίξει τη σελίδα του άρθρου και να δει όλο το περιεχόμενο.

Στη σελίδα του άρθρου ο χρήστης έχει τη δυνατότητα να διαβάσει όλο το άρθρο, να ανοίξει και να δει τις αξιολογήσεις άλλων χρηστών και στο τέλος του άρθρου να βρει και παρόμοια άρθρα.

Τα παρόμοια άρθρα έρχονται με κλήση του endpoint similar articles που δέχεται ως όρισμα το άρθρο και επιστρέφει όλα τα παρόμοια άρθρα.

```

{
  "Article": {
    "title": "Δημοτική εκτίμηση ΟΗΕ: Ο απολογισμός των θυμάτων σε Τουρκία - Συρία θα διπλασιαστεί",
    "author": "Σύνταξη IN",
    "id": 93,
    "tags": null,
    "date": "12/02/2023 08:38",
    "reliability": null,
    "reading_time": null,
    "image": "https://www.in.gr",
    "original_link": "https://www.in.gr/2023/02/12/world/seismos-stin-tourkia-o-apologismos-ton-thymaton-se-tourkia-syria-the-diplasiastei-proeidopoi-i-o-ole/",
    "body": " Δεκάδες χιλιάδες μέλη σωστικών συνεργειών συνεχίζουν να ψάχνουν στα συντρίμια, εν μέσω του κρύου που χειροτερεύει την κατάσταση για τα εκατομμύρια ατομομαθητές.Ο απολογισμός των θυμάτων του καταστροφικού σεισμού – ο οποίος ξεπέρασε τους 28.000 νεκρούς – στην Τουρκία και τη Συρία, θα «διπλασιαστεί» και αποτελεί να αυξηθεί «ακόμα περισσότερο», προειδοποίησε ο Μάρτιν Γκρίφθς, κορυφαίο στέλεχος του Οργανισμού Ηνωμένων Εθνών αρμόδιος για τον συντονισμό των ανθρωπιστικών επιχειρήσεων.Ο κ. Γκρίφθς πήγε το Σάββατο στην πόλη Κακονάρμαρας, κοντά στο επίκεντρο του σεισμού 7,8 βαθμών της κλίμακας Ριχτήρ, ο οποίος σκότωσε τη ζωή σε 24.617 πολίτες στην Τουρκία και σε άλλους 3.547 στη Συρία, ή συνολικά σε 28.191 ανθρώπους, κατά τα πιο πρόσφατα, ακόμη προσωρινά δεδομένα από τις δύο χώρες. «ΣΙΑΣΤΕ ΕΠΙΘΗ» Συναίξει μόνο τις η ΦΡΑΚ τις έρευνες για τον πλήρη αριθμό. «Η έκκληση των συγγενών «θαύω ότι είναι δύσκολο να εκτελέσουμε με σεβασμό (τον απολογισμό των θυμάτων)», όπως σημείει ειγώρος ότι θα διπλασιαστεί, αν δεν αυξηθεί ακόμη περισσότερο», είπε ο κ. Γκρίφθς στο τηλεοπτικό δίκτυο Sky News.«δεν έχει αρχίσει ακόμη ολόκληρη η καταμέτρηση των νεκρών», πρόσθεσε.Δεκάδες χιλιάδες μέλη σωστικών συνεργειών συνεχίζουν να ψάχνουν στα συντρίμια, εν μέσω του κρύου που χειροτερεύει την κατάσταση για τα εκατομμύρια ατομομαθητές.Τα Ηνωμένα Έθνη προειδοποίησαν πως τουλάχιστον 870.000 άνθρωποι χρειάζονται επείγοντες επισιτιστική βοήθεια, ζωτά γεύματα, στην Τουρκία και στη Συρία. Υπολογίζεται πως ως και 5,3 εκατομμύρια άνθρωποι έμειναν άστεγοι μόνο στη Συρία.Σχεδόν 26 εκατομμύρια άνθρωποι εκλήθησαν από την καταστροφή, ανέθεσε χτες ο Παγκόσμιος Οργανισμός Υγείας, απευθύνοντας έκκληση για δωρεές 42,8 εκατ. δολαρίων ώστε να χρηματοδοτηθεί η προσπάθεια αντιμετώπισης των άμεσων αναγκών από υγειονομική άποψη.Κατά την υπηρεσία αντιμετώπισης καταστροφών της Τουρκίας (AFAD), πάνω από 32.000 μέλη τουρκικών σωστικών συνεργειών συμμετείχαν στις έρευνες, στα οποία προστίθενται 8.294 μέλη σωστικών συνεργειών και υπηρεσιών αφής από το εξωτερικό.«Γίνονται, οι άνθρωποι που έχουν επιφορτιστεί με τις επιχειρήσεις έρευνας και διάσωσης θα δώσουν τη θέση τους σε εργαζόμενους ανθρωπιστικών οργανώσεων το έργο των οποίων θα είναι να φροντίσουν τους επόμενους μήνες τον εξαιρετικά μεγάλο αριθμό των ανθρώπων που εκλήθησαν, τόνισε ακόμη χτες Σάββατο ο κ. Γκρίφθς σε βίντεο που μεταφορτώθηκε στον λογαριασμό του στο Twitter.«what I saw today in #Turkey was devastating.what were once homes, filled with families and memories, now lay contorted and tangled.Our thoughts are with those affected and we will continue to support in any way we can. pic.twitter.com/5HMDR10E9N» Μάρτιν Γκρίφθς (@UNReliefChief) February 11, 2023 Έκτακτη βοήθεια από ΠΟΥ ΠΟΥ ζητεί έκτακτη οικονομική βοήθεια, ύψους 42,8 εκατομμυρίων δολαρίων, για την κάλυψη των άμεσων υγειονομικών αναγκών.Μερίτερα την εβδομάδα αυτή, ο ΠΟΥ, ο οποίος αποδόμησε ήδη 16 εκατομμύρια δολάρια από το ταμείο έκτακτης ανάγκης που διαθέτει, είχε υπολογίσει σε 23 εκατομμύρια τον αριθμό των ανθρώπων που μπορεί να είναι εκτεθειμένοι μετά τους σεισμούς και την καταστροφή που προκάλυψαν.Στην Τουρκία, ο ΠΟΥ εκτιμά ότι περισσότερα από 4.000 κτίρια κατέρρευσαν κατά τη διάρκεια των σεισμών και ότι 15 νοσοκομεία υπέστησαν βλάβες ή μεγάλες ζημιές",
    "image": "https://www.in.gr/wp-content/uploads/2023/02/2023-02-11T231447Z_1238598244_RC22929XW69T_RTRMADP_5_TURKEY-QUAKE-768x512.jpg",
    "body_html": null
  },
  "Similarity": {
    "FK_article_id": 93,
  }
}

```

Παράδειγμα Similar Articles Response

Η λελογισμένη χρήση των ηλεκτρικών συσκευών κατά τη διάρκεια των ακραίων αυτών καιρικών συνθηκών, θα συμβάλλει καθοριστικά στην εξοικονόμηση ενέργειας, διασφαλίζοντας την υγεία, την ασφάλεια και την καλύτερη εξυπηρέτηση όλων των πολιτών. Ο ΔΕΔΔΗΕ εκφράζει τη λύπη του για την ταλαιπωρία των καταναλωτών και τους ευχαριστεί εκ των προτέρων για την ενεργό συμμετοχή τους στις προσπάθειες περιορισμού της κατανάλωσης ρεύματος.

Παρόμοια Άρθρα

Ανεξέλεγκτη η φωτιά στη Ρόδο: Καίγονται πεντάστερα ξενοδοχεία στο Κιοτάρι - Εκκενώσεις δια θαλάσσης

<http://www.in.gr>
Σύνταξη IN
22/07/2023 16:46
0 (0)
μοναδικό: false
Άρθρο: 301



[...] Και τα τρία ξενοδοχεία έχουν εκκενωθεί από τουρίστες και γίνονται προσπάθειες να περιοριστούν στο ελάχιστο οι ζημιές από την πυροσβεστική, τους εθελοντές και το

Αξιολόγηση

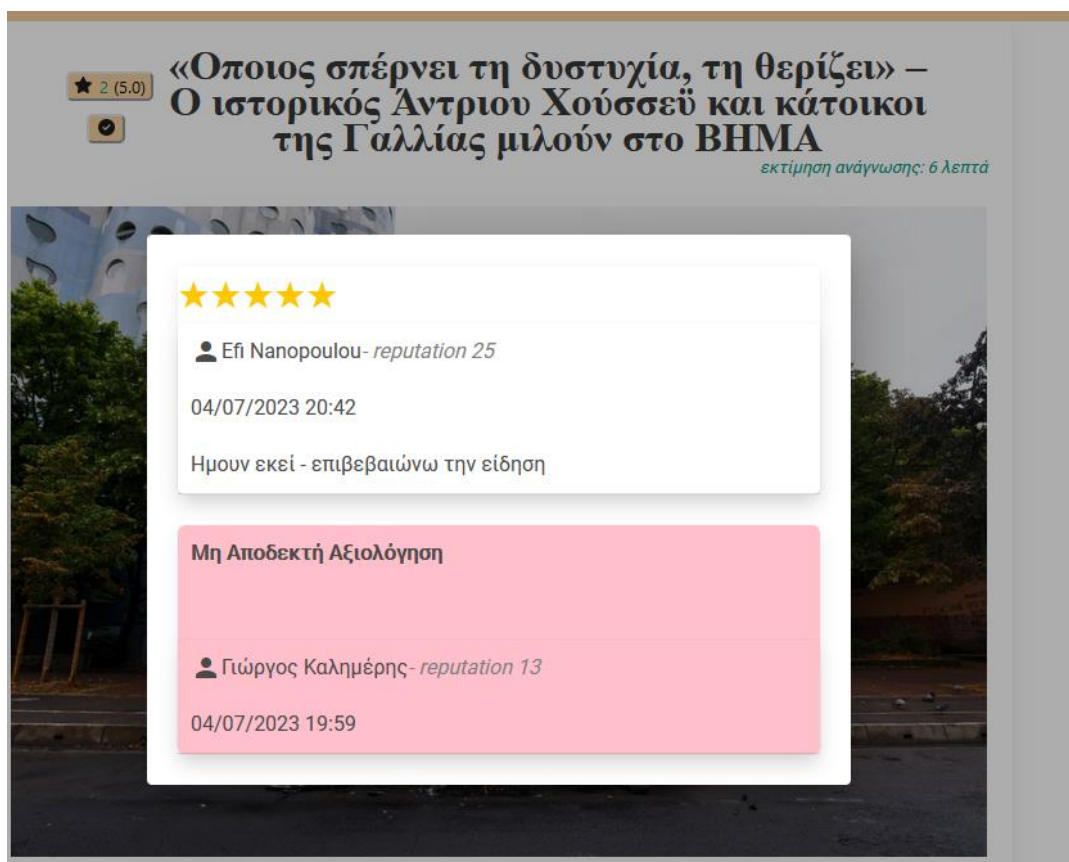
Παρόμοια άρθρα μέσα στη σελίδα του άρθρου

4.7 Εμφάνιση Αξιολογήσεων

Το End Point Article Reviews δέχεται ως παραμέτρους την τιμή article id που είναι ένας ακέραιος αριθμός και την τιμή user που είναι το JWT Token του χρήστη από το Front End. Επιστρέφει όλες τις αξιολογήσεις που έγιναν από όλους τους χρήστες για το συγκεκριμένο άρθρο.

```
{
  "id": 7,
  "FK_article_id": 288,
  "FK_user_id": 16,
  "grade": 5,
  "accepted": true,
  "created_at": "2023-07-04T20:42:47.203577",
  "comment": "Ημουν εκει - επιβεβαιώνω την ειδηση",
  "first_name": "Efi",
  "last_name": "Nanoroulou",
  "sum": 25
},
{
  "id": 6,
  "FK_article_id": 288,
  "FK_user_id": 10,
  "grade": 4,
  "accepted": false,
  "created_at": "2023-07-04T19:59:15.234306",
  "comment": "Δεν ισχύουν όσα γράφουν",
  "first_name": "Γιώργος",
  "last_name": "Καλημέρης",
  "sum": 13
}
```

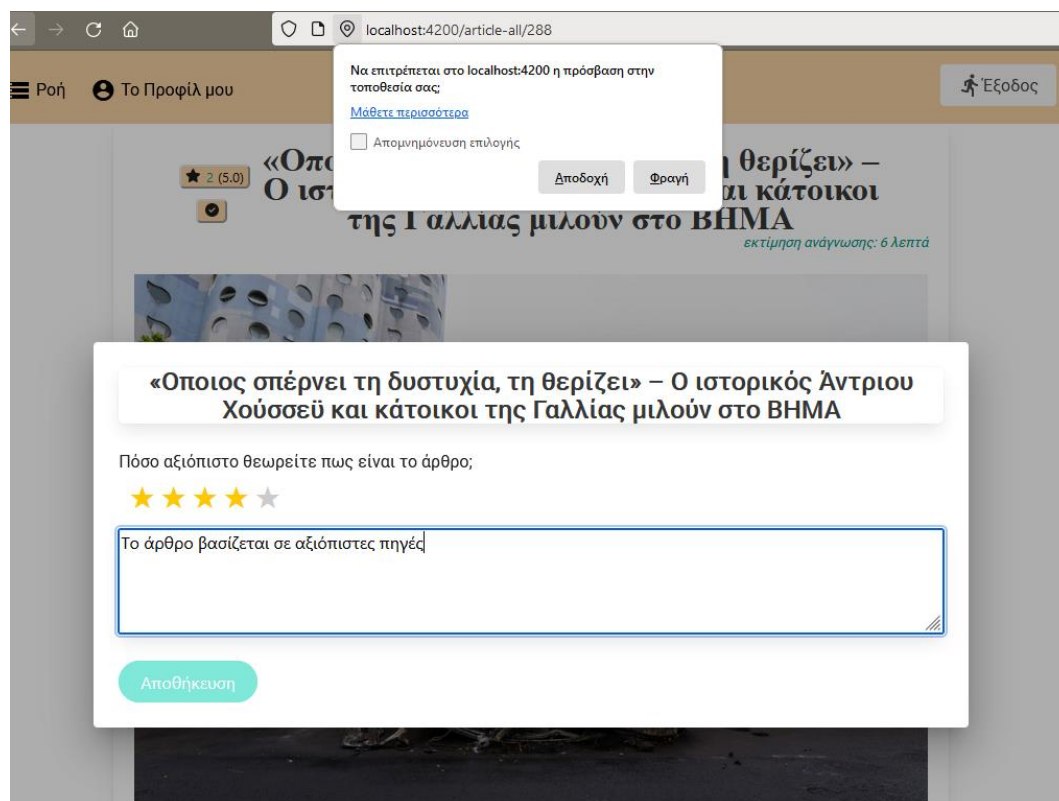
Παράδειγμα Reviews Response



Pop up σελίδα αξιολογήσεων

Δίπλα σε κάθε χρήστη υπάρχει το reputation του καθώς και το σχόλιο που ενδεχομένως να έχει γράψει. Μη αποδεκτές κριτικές εμφανίζονται με ροζ χρώμα, χωρίς όμως να φαίνεται η βαθμολογία, αλλά δείχνουν τον χρήστη και το reputation του.

4.8 Νέα αξιολόγηση



Pop up νέα αξιολόγηση

Το κουμπί Αποθήκευση δεν ενεργοποιείται εάν ο χρήστης δεν επαληθεύσει την τοποθεσία του και συμπληρώσει βαθμολογία.

Το End Point New Review δέχεται ως παραμέτρους ένα αντικείμενο τύπου New Review καθώς και τον χρήστη user. Η βασική λειτουργία του endpoint είναι ο έλεγχος και η αποθήκευση νέων αξιολογήσεων. Το αντικείμενο τύπου New Review έρχεται με data συμπληρωμένα από το Front End και περιέχει τις παρακάτω πληροφορίες:

- Το id του άρθρου που θα αποθηκευτεί η αξιολόγηση
- Τον βαθμό της αξιολόγησης που έβαλε ο χρήστης (ένα ως πέντε)
- Τον χρόνο σε λεπτά που αφιέρωσε ο χρήστης να διαβάσει το άρθρο (είναι ο χρόνος από την εισαγωγή του χρήστη στη σελίδα του άρθρου μέχρι τη στιγμή που θα πατήσει το κουμπί προσθήκη αξιολόγησης)
- Τον χρόνο που απαιτεί το άρθρο για να διαβαστεί και έχει υπολογιστεί κατά την αποθήκευση του άρθρου στη βάση δεδομένων
- Τα Latitude και Longitude του χρήστη τη στιγμή που πατάει το κουμπί προσθήκη αξιολόγησης στο Front End
- Το σχόλιο του χρήστη που θα είναι ορατό σε όλους τους χρήστες εφόσον εγκριθεί η αξιολόγησή του από το back end
- Η τιμή reach_bottom που είναι True, False και δηλώνει εάν ο χρήστης scrollare μέχρι το τέλος του άρθρου

Πριν την καταχώρηση της αξιολόγησης του χρήστη πραγματοποιείται έλεγχος της αξιολόγησης στον Review Controller. Αρχικά ελέγχεται η απόσταση του χρήστη από την απόσταση των περιοχών που έχουν εξαχθεί από το άρθρο. (π.χ. Εάν ένα άρθρο μιλάει για την Κρήτη, ένας χρήστης που θα αξιολογήσει την είδηση από την Κρήτη θα συγκεντρώσει περισσότερους πόντους σε σχέση με κάποιον που προχώρησε σε αξιολόγηση της είδησης από τη Θεσσαλονίκη ή ακόμα πιο μακριά π.χ. τον Καναδά). Στη συνέχεια ελέγχεται ο χρόνος που αφιέρωσε ο χρήστης στο να διαβάσει το κείμενο σε σχέση με τον προ-υπολογισμένο χρόνο που θεωρείται πως χρειάζεται για να διαβαστεί το κείμενο. Τέλος ελέγχεται εάν έφτασε μέχρι το τέλος του κειμένου. Εάν περάσει όλους τους ελέγχους, δηλαδή – το geolocation του χρήστη είναι σχετικά κοντά στο geolocation του άρθρου και έχει διαβάσει το άρθρο στα προκαθορισμένα απαιτούμενα χρονικά περιθώρια (με ένα εύρος 20 δευτερολέπτων περισσότερα ή λιγότερα) και έχει φτάσει μέχρι το τέλος του κειμένου τότε η αξιολόγηση εγκρίνεται και αποθηκεύεται στον πίνακα των αξιολογήσεων. Σε περίπτωση που δεν περάσει τους ελέγχους αποθηκεύεται στην βάση δεδομένων αλλά επισημαίνεται στο πεδίο accepted, με False, ως μη αποδεκτή αξιολόγηση.

id	FK_article_id	FK_user_id	grade	time_spent_reading	time_article_needs	user_latitude	user_longitude	reach_bottom	accepted	created_at	comment
4	280	16	3	0.334	300.0	38.0751226	23.8116458	0	0	2023-07-02 14:04:20.496159	Δεν ισχύει
5	286	16	1	2.912	420.0	38.0751162	23.8116391	0	0	2023-07-03 19:13:16.632472	Δεν είναι σωστό
6	288	10	4	7.391	360.0	38.0751259	23.8116525	0	0	2023-07-04 19:59:15.234306	Δεν ισχύουν όσον γράφουν
7	288	16	5	11.48	360.0	38.0751241	23.8116666	1	1	2023-07-04 20:42:47.203577	Ημουν εκεί - επιβεβαιώνα την είδηση
8	293	16	4	105.993	240.0	38.0751325	23.8116261	1	1	2023-07-15 09:50:48.389416	

Πίνακας Review

4.9 Reputation

Ο κάθε χρήστης ανάλογα με τις αξιολογήσεις και τα άρθρα που διαβάζει (Transaction) λαμβάνει ή χάνει πόντους αξιολόγησης που του δημιουργούν το δικό του reputation.

Κάθε Transaction προσθέτει ή αφαιρεί πόντους στον χρήστη:


- +2 πόντοι για κάθε ανάγνωση άρθρου
- +5 πόντοι για κάθε αξιολόγηση που έγινε δεκτή από το Backend
- -10 πόντοι για κάθε αξιολόγηση που απορρίφθηκε από το Backend
- -10 πόντοι εάν προσπαθήσει να αξιολογήσει ξανά το ίδιο άρθρο

To end point user transactions δέχεται ως όρισμα τον χρήστη από το JWT Token και επιστρέφει όλα τα transactions του συνδεδεμένου χρήστη.

```
{
  "FK_user_id": 10,
  "FK_article_id": 16,
  "created_at": "2023-09-24T10:19:24.230207",
  "article_title": "0 Ερντογάν... μαθαίνει τον αντίπαλό του - Πώς καταπατά το Σύνταγμα για να παραμείνει στην εξουσία",
  "review_accepted": null,
  "transaction_points": 2,
  "id": 48,
  "FK_review_id": null,
  "review_grade": null
},
{
  "FK_user_id": 10,
  "FK_article_id": 314,
  "created_at": "2023-09-24T10:19:21.377304",
  "article_title": "Στα ύψη η συμμετοχή στις εκλογές του ΣΥΡΙΖΑ - Ψηφίζουν πάνω από 250 μέλη το λεπτό",
  "review_accepted": null,
  "transaction_points": 2,
  "id": 47,
  "FK_review_id": null,
  "review_grade": null
},
{
```

Παράδειγμα User transactions response

☰ Ροή
👤 Το Προφίλ μου
🏠 Έξοδος



Efi Nanoroulou
Reputation: 25

ιστορικό:

22/07/2023 15:49	Ανάγνωση άρθρου
+2	Επιμένει ο καύσωνας: Με 44άρια η ερχόμενη βδομάδα - Ποιες περιοχές θα «καούν»
22/07/2023 14:40	Αξιολόγηση άρθρου
+2	Φωτιά στη Ρόδο: Ανεξέλεγκτη η φωτιά – Η πυρκαγιά έφτασε σε ξενοδοχεία – Εκκενώσεις και μέσω θάλασσας
22/07/2023 14:38	Ανάγνωση άρθρου
+2	Φωτιά στη Ρόδο: Ανεξέλεγκτη η φωτιά – Η πυρκαγιά έφτασε σε ξενοδοχεία – Εκκενώσεις και μέσω θάλασσας
22/07/2023 11:29	Αξιολόγηση άρθρου
+5	Γιατί χάνονται τόσες ζωές σε θάλασσες και πισίνες - Οι αιτίες, τα λάθη - Ειδικό εξηγούν στο ΒΗΜΑ
22/07/2023 11:28	Ανάγνωση άρθρου
+2	Γιατί χάνονται τόσες ζωές σε θάλασσες και πισίνες - Οι αιτίες, τα λάθη - Ειδικό εξηγούν στο ΒΗΜΑ
22/07/2023 10:48	Ανάγνωση άρθρου
+2	Κρίσιμη η κατάσταση στη Ρόδο – Εκκένωση δύο ακόμα χωριών
15/07/2023 10:04	Αξιολόγηση άρθρου
-10	«Hot dry windy», ο καιρός που ανησυχεί τους επιστήμονες - Οι περιοχές με τις υψηλότερες θερμοκρασίες

Σελίδα Το προφίλ του χρήστη με τα reputation logs

5 Desing Patterns

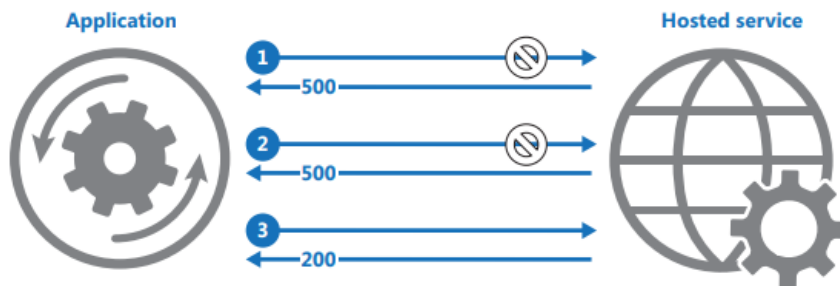
Αρκετά σημεία της υλοποίησης σχεδιάστηκαν και αναπτύχθηκαν πάνω σε πρότυπα σχεδιασμού (design patterns). Τα πρότυπα σχεδιασμού είναι τυπικές λύσεις που έχουν δουλευτεί από χιλιάδες προγραμματιστές πάνω σε κοινά προβλήματα που εμφανίζονται συχνά στην ανάπτυξη λογισμικού. Δεν αφορούν κώδικα αλλά τη συνταγή για το πως να δομήσεις τον κώδικά σου σε ένα συγκεκριμένο πρόβλημα. [17]

5.1 Retry Pattern

Το πρότυπο σχεδιασμού retry pattern επιτρέπει σε μια εφαρμογή να αντιμετωπίζει απροσδόκητες προσωρινές αποτυχίες όταν προσπαθεί να συνδεθεί σε μια υπηρεσία ή έναν πόρο δικτύου, επαναλαμβάνοντας αυτόματα μια λειτουργία που προηγουμένως απέτυχε με την προσδοκία, ο λόγος που απέτυχε να είναι προσωρινός. Αυτό το πρότυπο βελτιώνει τη σταθερότητα της εφαρμογής. [15]

Στη εφαρμογές διαδικτύου οι παροδικές αποτυχίες (π.χ. δικτύου) είναι συχνές και κάθε εφαρμογή θα πρέπει να υλοποιηθεί με τέτοιο τρόπο που να τις διαχειρίζεται με στόχο να ελαχιστοποιεί τις επιπτώσεις της δυσλειτουργίας κατά τη χρήση της. Έτσι εάν η εφαρμογή λάβει ένα μήνυμα σφάλματος θα πρέπει να ακολουθήσει τις παρακάτω στρατηγικές.

- Εάν από το μήνυμα σφάλματος φαίνεται πως η αποτυχία δεν είναι παροδική ή είναι απίθανο να ξεπεραστεί με μια απλή επανάληψη (π.χ. Ένα σφάλμα αυθεντικοποίησης δεν πρόκειται να αλλάξει με την πάροδο του χρόνου) τότε η εφαρμογή θα πρέπει να σταματήσει τη διαδικασία της επανάληψης και να εξάγει ένα κατάλληλο μήνυμα λάθους.
- Εάν το σφάλμα είναι σπάνιο ή ασυνήθιστο μπορεί να οφείλεται σε κάποια δυσλειτουργία του δικτύου οπότε η εφαρμογή θα πρέπει να δοκιμάσει ξανά.
- Εάν το σφάλμα έχει προκληθεί από καθυστερήσεις του δικτύου είτε από κάποιο deadlock κατά τη διάρκεια εισαγωγής εγγραφών σε μια βάση τότε η εφαρμογή θα πρέπει να περιμένει ένα μικρό χρονικό περιθώριο πριν δοκιμάσει ξανά. [15]



Παράδειγμα υλοποίησης του Retry Pattern [15]

Για την ενσωμάτωση του retry pattern εγκαταστάθηκε η βιβλιοθήκη retry της python στο backend.

Η βιβλιοθήκη retry μας επιτρέπει να προσθέτουμε τον decorator `@retry()` πάνω από μια function που θέλουμε να εφαρμοστεί το πρότυπο. Στον decorator έχουμε τη δυνατότητα να περάσουμε όσες από τις παρακάτω παραμέτρους επιθυμούμε:

- `exceptions=Exception`
- `tries=-1`
- `delay=0`
- `max_delay=None`
- `backoff=1`
- `jitter=0`
- `logger=logging_logger`

```

:param exceptions: an exception or a tuple of exceptions to catch. default: Exception.
:param tries: the maximum number of attempts. default: -1 (infinite).
:param delay: initial delay between attempts. default: 0.
:param max_delay: the maximum value of delay. default: None (no limit).
:param backoff: multiplier applied to delay between attempts. default: 1 (no backoff).
:param jitter: extra seconds added to delay between attempts. default: 0.
                fixed if a number, random if a range tuple (min, max)
:param logger: logger.warning(fmt, error, delay) will be called on failed attempts.
                default: retry.logging_logger. if None, logging is disabled.

```

Επεξήγηση παραμέτρων που δέχεται η Retry Function

Το Retry εφαρμόστηκε σε όλες τις συναρτήσεις που πραγματοποιούν εγγραφές στη βάση δεδομένων. Με αυτό τον τρόπο εάν η βάση για κάποιο λόγο κλειδώσει προσωρινά από μια άλλη διαδικασία, το retry θα ξανά προσπαθήσει

```

@retry(exceptions=Exception, tries=5, delay=1, backoff=2, jitter=1)
def article_meta_data(article: Article):
    """Αποθηκεύει τα "Meta" data του άρθρου
    Ονόματα, Περιοχές, Οργανισμούς σε μορφή Json"""
    with Session(engine) as session:

```

Retry decorator

Στην παραπάνω συνάρτηση ζητήθηκε από τον decorator, σε περίπτωση που βρεθεί σφάλμα, να προσπαθήσει έως 5 φορές με καθυστέρηση 1 δευτερολέπτου και backoff = 2. Το backoff διπλασιάζει τον χρόνο καθυστέρησης μετά από κάθε προσπάθεια. Επιπλέον το jitter = 1 προσθέτει ένα επιπλέον δευτερόλεπτο στην τιμή του χρόνου καθυστέρησης μετά από κάθε προσπάθεια.

Το ίδιο πρότυπο εφαρμόστηκε και στο Front-End σε περίπτωση που το API για κάποιο λόγο δεν είναι διαθέσιμο. Με χρήση της βιβλιοθήκης rxjs δημιουργήθηκε η παρακάτω συνάρτηση που παίρνει ως ορίσματα τον χρόνο καθυστέρησης σε milliseconds καθώς και το max Retry που δηλώνει τον μέγιστο αριθμό προσπαθειών. Η συγκεκριμένη συνάρτηση παίρνει το observable που δημιουργεί το http request και το περνάει από ένα pipe. Εκεί εάν η τιμή retries είναι μεγαλύτερη του 0 και εξακολουθεί να επιστρέφεται σφάλμα το πρόγραμμα παγώνει για x milliseconds. Ανάλογα με το πόσα έχουμε επιλέξει.

```

export function delayedRetry(delayMS: number, maxRetry = 5, ) {
    let retries = maxRetry;

    return (src: Observable<any>)=>
    src.pipe(
        retryWhen((errors:Observable<any>)=> errors.pipe(
            delay(delayMS),
            mergeMap(error=>retries-- > 0 ? of(error): throwError(1))
        ))
    )
}

```

Retry pattern στο Front End

Με αυτό τον τρόπο σε κάθε call του service προστέθηκαν σε pipe οι παρακάτω γραμμές κώδικα που σε περίπτωση σφάλματος από το API ξανά προσπαθούν μετά από 1 δευτερόλεπτο για πέντε φορές.:

```

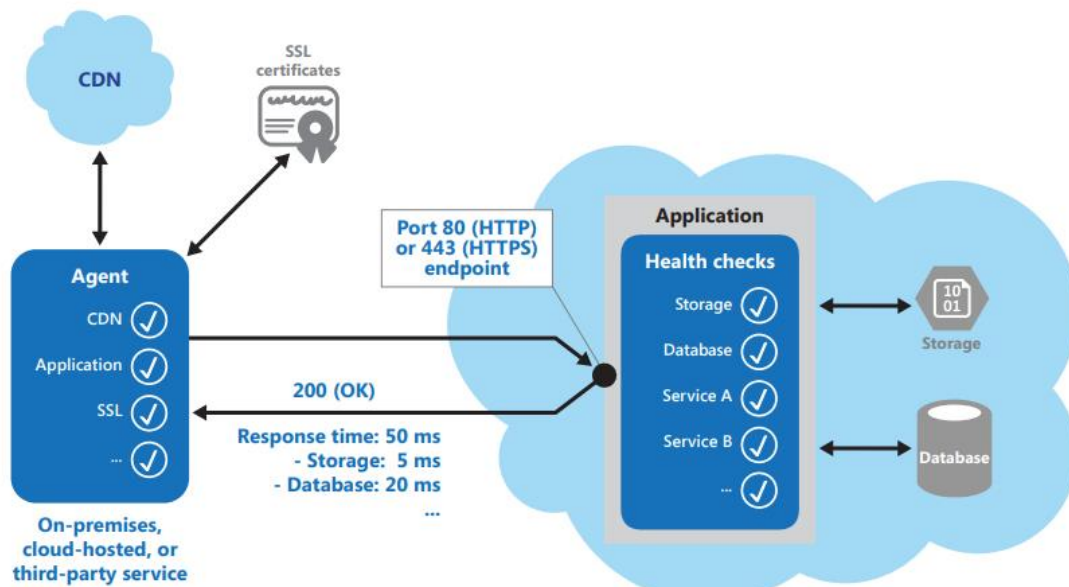
getUserTransactions(){
  return this.http.get<any>(`${this.API_URL}/user-transactions`, {}).pipe(
    delayedRetry(1000),
    catchError(()=>{
      return EMPTY;
    }),
    shareReplay()
  );
}

```

Χρήση της συνάρτησης

5.2 Health End Point Monitoring Pattern

Το Health End Point Monitoring pattern χρησιμοποιείται προκειμένου να ελεγχθεί η ορθή λειτουργία της εφαρμογής. [15]



Παράδειγμα υλοποίησης του Health End Point Monitoring Pattern [15]

Για την υλοποίησή του δημιουργήθηκε ένα end point όπου όταν καλείται εκτελεί ελέγχους ως προς τη χρήση του επεξεργαστή του server, τη χρήση της Ram του υπολογιστή που τρέχει η εφαρμογή και τέλος πραγματοποιεί έλεγχο ορθής λειτουργίας της βάσης δεδομένων της εφαρμογής.

```

@app.get('/api-health')
async def transactions():
    """Τσεκάρει την υγεία του server και της Database"""
    result = health_check()
    return result

```

Health end point

Αρχικά πραγματοποιείται έλεγχος στη βάση δεδομένων με select προκειμένου να δούμε πως η βάση ανταποκρίνεται και σε περίπτωση που η διαδικασία ολοκληρωθεί επιστρέφει True.

```
def check_database():
    try:
        with Session(engine) as session:
            statement = select(Article)
            statement = session.exec(statement).all()
            if statement is not None:
                return True
            else:
                return False
    except:
        return False
```

Έλεγχος Database

Στη συνέχεια πραγματοποιείται έλεγχος στη χρήση της CPU του server που φιλοξενεί το API και σε περίπτωση που βρεθεί με πάνω από 90% επιστρέφει False.

```
def check_cpu():
    load1, load5, load15 = psutil.getloadavg()
    cpu_usage = (load15 / os.cpu_count()) * 100
    if cpu_usage > 90:
        return False
    else:
        return True
```

Έλεγχος CPU

Τέλος πραγματοποιείται έλεγχος στη χρήση της RAM του server που φιλοξενεί το API και σε περίπτωση που βρεθεί με πάνω από 90% επιστρέφει False.

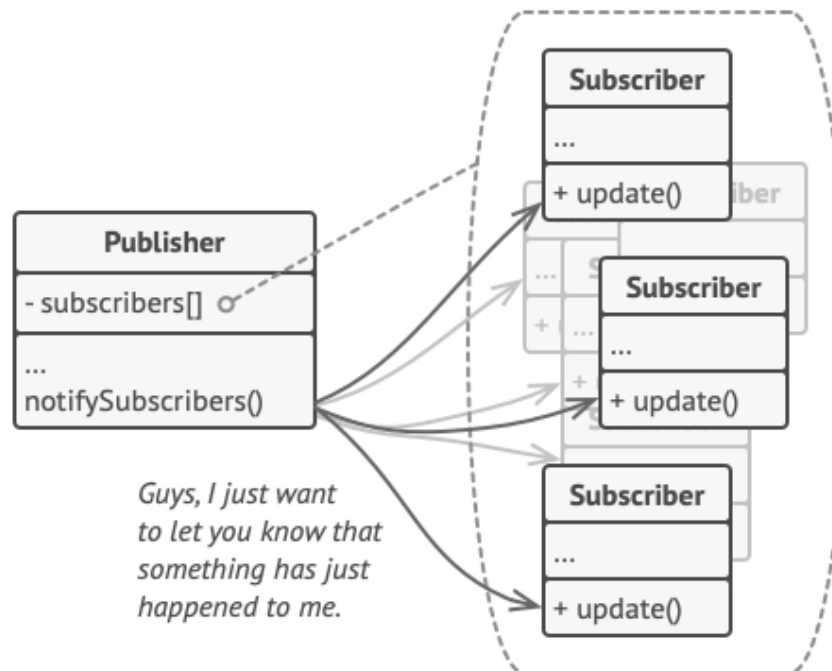
```
def check_ram():
    if psutil.virtual_memory()[2] > 90:
        return False
    else:
        return True
```

Έλεγχος RAM

5.3 Observer Design Pattern

Το πρότυπο σχεδίασης Observer είναι ένα behavioral design pattern που επιτρέπει να οριστεί ένας μηχανισμός που ειδοποιεί πολλά αντικείμενα για διάφορα γεγονότα που συμβαίνουν στο αντικείμενο που παρατηρούν και στο οποίο έχουν πρώτα κάνει εγγραφή. [17]

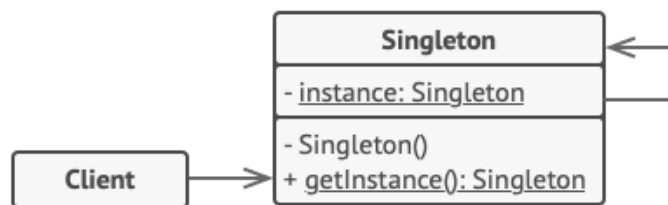
Το συγκεκριμένο πρότυπο σχεδίασης χρησιμοποιείται σε όλο το framework της Angular για τη μεταφορά μηνυμάτων μεταξύ τμημάτων της εφαρμογής, για τη διαχείριση ενεργειών είτε του χρήστη είτε της εφαρμογής καθώς και για ασύγχρονο προγραμματισμό με τη διαχείριση πολλαπλών τιμών. [18]



Observer [17]

5.4 Singleton Design Patter

Το πρότυπο σχεδίασης Singleton είναι ένα creational design pattern και διασφαλίζει πως μια κλάση μπορεί να έχει μόνο ένα instance το οποίο είναι διαθέσιμο σε όλη την εφαρμογή. [17]



1 The **Singleton** class declares the static method `getInstance` that returns the same instance of its own class.

The Singleton's constructor should be hidden from the client code. Calling the `getInstance` method should be the only way of getting the Singleton object.

```

if (instance == null) {
    // Note: if you're creating an app with
    // multithreading support, you should
    // place a thread lock here.
    instance = new Singleton()
}
return instance
    
```

Singleton [17]

Το framework Angular χρησιμοποιεί το patterns Singleton σε κάθε component που θέλει να συνδεθεί με οποιοδήποτε service προκειμένου να λάβει, να στείλει ή να επεξεργαστεί δεδομένα. Η χρήση του Singleton πραγματοποιείται μέσω του Dependency Injection.

Το Dependency Injection (DI) είναι ένα δομικό στοιχείο της Angular. Το DI είναι συνδεδεμένο μέσα στο Angular framework και επιτρέπει στις κλάσεις που έχουν decorators όπως Components, Directives, Pipes και Injectables να ορίζουν τις εξαρτήσεις που χρειάζονται. [18]

6 Συμπεράσματα

Η εφαρμογή αυτή τη στιγμή αντλεί ειδησεογραφικά άρθρα από τρεις πηγές. Για την επέκτασή της θα μπορούσαν να προστεθούν στην αράχνη περισσότεροι ιστότοποι.

Κατά τη διάρκεια της υλοποίησης της εφαρμογής και με τη συνεχή αποθήκευση νέων άρθρων παρατηρήθηκε πως σε μια μεγάλη είδηση πολλά ειδησεογραφικά πρακτορεία αναπαράγουν την είδηση με ελάχιστες διαφοροποιήσεις σε τίτλους και κείμενα.

id	title	date	body
41	Ήρθε στην Ελλάδα η «κίρακεν» - Πόσα κρούσματα εντοπίστηκαν	12/01/2023 17:43	Συνολικά έχει ανιχνευτεί σε 6 δείγματα στη χώρα μας...
42	Ήρθε στην Ελλάδα η μετάλλαξη «κίρακεν» – Πόσα κρούσματα εντοπίστηκαν	12.01.2023 - 17:58	Συνολικά έχει ανιχνευτεί σε 6 δείγματα στη χώρα μας. Στην Ελλάδα έφθασε η μετάλλαξη «κίρακεν» όπως αναφέρεται στην εβδομαδιαία ...
43	«Ήρθε» στην Ελλάδα η άκρως μεταδοτική μετάλλαξη Κράκεν – Η ανακοίνωση ...	12/01/2023 17:59	Σύμφωνα με την τελευταία επιδημιολογική έκθεση του Εθνικού Οργανισμού Δημόσιας Υγείας (ΕΟΔΥ), η μετάλλαξη Κράκεν ήρθε και επίσημ...

Παράδειγμα παρόμοιων κειμένων από διαφορετικούς ειδησεογραφικούς ιστότοπους

Επιπλέον σε επέκταση της εφαρμογής θα μπορούσε να υλοποιηθεί ανάλυση για την εύρεση της πρωταρχικής πηγής μιας είδησης.

Η προσπάθεια αξιολόγησης των αξιολογήσεων των χρηστών είναι μια αρκετά δύσκολη και περίπλοκη διαδικασία. Τα κριτήρια που υλοποιήθηκαν ήταν τρία και αφορούσαν την επαφή του χρήστη με την εφαρμογή:

- Η τοποθεσία του χρήστη κατά την χρήση της εφαρμογής
- Ο χρόνος που χρειάστηκε για να διαβάσει ένα άρθρο
- Καθώς και εάν πλοηγήθηκε μέχρι το τέλος του άρθρου

Όλα τα παραπάνω κριτήρια αποθηκεύονται χωρίς ο χρήστης να το αντιλαμβάνεται. Για την περαιτέρω ανάπτυξη της εφαρμογής θα μπορούσε να υλοποιηθεί επιπλέον έλεγχος για το πόση ώρα ο χρήστης είναι συγκεντρωμένος πάνω στο κείμενο με χρήση της κάμερας της συσκευής του.

Επιπλέον θα μπορούσε η εφαρμογή να αποθηκεύει περισσότερα στοιχεία του χρήστη κατά την εγγραφή του όπως η ημερομηνία γέννησης, το επίπεδο εκπαίδευσης, το εάν εργάζεται ή όχι κ.α. και με αυτό τον τρόπο θα μπορούσε να δημιουργεί clusters χρηστών που θα βοηθούσαν στην καλύτερη αξιολόγηση των αξιολογήσεων.

Τέλος εάν η εφαρμογή λειτουργούσε με ένα μεγάλο αριθμό χρηστών θα μπορούσε μέσα σε λίγους μήνες να εξάγει Ratings για τα μέσα και τους συντάκτες των κειμένων από τις αξιολογήσεις των χρηστών. Αυτό θα οδηγούσε σε μια άτυπη αξιολόγηση των συντακτών και του ηλεκτρονικού τύπου συνολικά, αναγκάζοντας τις διαχειρίστριες εταιρείες να ελέγχουν περισσότερο τις πηγές τους και να δίνουν μεγαλύτερη προσοχή στα κείμενα που αναρτούν.

7 Disclaimer

Σύμφωνα με την παράγραφο 5 του άρθρο 2 του νόμου 2121/1993 που παρατίθεται στο κεφάλαιο 8, οι ειδήσεις και τα απλά γεγονότα εξαιρούνται από το νόμο περί προστασίας πνευματικής ιδιοκτησίας [18]

Η αποθήκευση και η εμφάνιση του περιεχομένου που ανακτήθηκε από τις ιστοσελίδες έγινε στα πλαίσια της παρούσας διπλωματικής εργασίας σε ιδιωτικό περιβάλλον. Ολόκληρο το υλικό διαγράφηκε με την ολοκλήρωση της παρούσας διπλωματικής εργασίας.

8 ΠΑΡΑΘΕΣΗ

Άρθρο 2 - Νόμος 2121/1993 - Αντικείμενο του δικαιώματος

Ως έργο νοείται κάθε πρωτότυπο πνευματικό δημιούργημα λόγου, τέχνης ή επιστήμης, που εκφράζεται με οποιαδήποτε μορφή, ιδίως τα γραπτά ή προφορικά κείμενα, οι μουσικές συνθέσεις, με κείμενο ή χωρίς, τα θεατρικά έργα, με μουσική ή χωρίς, οι χορογραφίες και οι παντομίμες, τα οπτικοακουστικά έργα, τα έργα των εικαστικών τεχνών, στα οποία περιλαμβάνονται τα σχέδια, τα έργα ζωγραφικής και γλυπτικής, τα χαρακτηριστικά έργα και οι λιθογραφίες, τα αρχιτεκτονικά έργα, οι φωτογραφίες, τα έργα των εφαρμοσμένων τεχνών, οι εικονογραφήσεις, οι χάρτες, τα τρισδιάστατα έργα που αναφέρονται στη γεωγραφία, την τοπογραφία, την αρχιτεκτονική ή την επιστήμη.

2. Νοούνται επίσης ως έργα οι μεταφράσεις, οι διασκευές, οι προσαρμογές και οι άλλες μετατροπές έργων ή εκφράσεων της λαϊκής παράδοσης, καθώς και οι συλλογές έργων ή συλλογές εκφράσεων της λαϊκής παράδοσης ή απλών γεγονότων και στοιχείων, όπως οι εγκυκλοπαίδειες και οι ανθολογίες, εφόσον η επιλογή ή η διευθέτηση του περιεχομένου τους είναι πρωτότυπη. Η προστασία των έργων της παρούσας παραγράφου γίνεται με την επιφύλαξη των δικαιωμάτων στα προϋπάρχοντα έργα, που χρησιμοποιήθηκαν ως αντικείμενο των μετατροπών ή των συλλογών.

2.α. Αντικείμενο προστασίας είναι και οι βάσεις δεδομένων οι οποίες λόγω της επιλογής ή διευθέτησης του περιεχομένου τους αποτελούν πνευματικά δημιουργήματα. Η προστασία αυτή δεν εκτείνεται στο περιεχόμενο των βάσεων δεδομένων και δεν θίγει κανένα από τα δικαιώματα που υφίστανται στο περιεχόμενο αυτό. Ως βάση δεδομένων νοείται η συλλογή έργων, δεδομένων ή άλλων ανεξάρτητων στοιχείων, διευθετημένων κατά συστηματικό ή μεθοδικό τρόπο και ατομικώς προσιτών με ηλεκτρονικά μέσα ή με άλλο τρόπο.

3. Με την επιφύλαξη των διατάξεων του κεφαλαίου 7 του παρόντος νόμου, θεωρούνται ως έργα λόγου προστατευόμενα κατά τις διατάξεις περί πνευματικής ιδιοκτησίας τα προγράμματα ηλεκτρονικών υπολογιστών και το προπαρασκευαστικό υλικό του σχεδιασμού τους. Η προστασία παρέχεται σε κάθε μορφή έκφρασης ενός προγράμματος ηλεκτρονικού υπολογιστή. Οι ιδέες και οι αρχές στις οποίες βασίζεται οποιοδήποτε στοιχείο προγράμματος ηλεκτρονικού υπολογιστή, συμπεριλαμβανομένων και εκείνων στις οποίες βασίζονται τα συστήματα διασύνδεσής του, δεν προστατεύονται κατά τον παρόντα νόμο. Ένα πρόγραμμα ηλεκτρονικού υπολογιστή θεωρείται πρωτότυπο εφόσον είναι προσωπικό πνευματικό δημιούργημα του δημιουργού του.

4. Η προστασία του παρόντος νόμου είναι ανεξάρτητη από την αξία και τον προορισμό του έργου, καθώς και από το γεγονός ότι το έργο προστατεύεται ενδεχομένως και από άλλες διατάξεις.

5. Η προστασία του παρόντος νόμου δεν εκτείνεται σε επίσημα κείμενα με τα οποία εκφράζεται η άσκηση πολιτειακής αρμοδιότητας και ιδίως σε νομοθετικά, διοικητικά ή δικαστικά κείμενα, καθώς και στις εκφράσεις της λαϊκής παράδοσης, στις ειδήσεις και στα απλά γεγονότα ή στοιχεία.[18]

9 Βιβλιογραφία

- [1] BeautifulSoup Python Library [online]. Available: <https://pypi.org/project/beautifulsoup4/> [Accessed 17/11/2022]
- [2] Jan Goyvaerts, Steven Levithan “Regular Expressions Cookbook Detailed Solutions in Eight Programming Languages 2nd Edition” O'Reilly Media (Οκτώβριος 2012)
- [3] Spacy Named-entity Recognition [online]. Available: <https://spacy.io/usage/linguistic-features#named-entities> [Accessed 10/08/2023]
- [4] Nitin Indurkha, Fref J. Damerau “HANDBOOK OF NATURAL LANGUAGE PROCESSING 2nd Edition” CRC Press Taylor & Francis Group (2010)
- [5] Spacy Trained Pipelines Greek el_core_news_lg [online]. Available: https://spacy.io/models/el#el_core_news_lg [Accessed 10/08/2023]
- [6] Scikit Learn 1.3.1 TfidfVectorizer [online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html [Accessed 15/01/2023]
- [7] Cosine Similarity [online]. Available: https://machinelearninggeek.com/text-similarity-measures/#Cosine_Similarity [Accessed 17/05/2023]
- [8] Natural Language Process - Text Summarization [online]. Available: https://github.com/DivakarPM/NLP/blob/master/Text_Summarization/Text_Summarization.ipynb [Accessed 10/02/2023]
- [9] Google Creating API requests and handling responses [online]. Available: <https://cloud.google.com/compute/docs/api/how-tos/api-requests-responses> [Accessed 10-03-2023]
- [10] Adam Freeman, Steven Sanderson «Pro ASP.NET MVC 4 4th Edition» Apress (Ιανουάριος 2013)
- [11] FastApi Repeat Feature [online]. Available: https://github.com/dmontagu/fastapi-utils/blob/master/fastapi_utils/tasks.py [Accessed 01/05/2023]
- [13] SQLAlchemy [online]. Available: <https://sqlmodel.tiangolo.com/> [Accessed 01/02/2023]
- [14] SQLAlchemy 1.4 [online]. Available: <https://docs.sqlalchemy.org/en/14/tutorial/> [Accessed 01/02/2023]
- [15] MVC Overview by Joseph Spinelli [online]. Available: https://medium.com/@joespinelli_6190/mvc-model-view-controller-ef878e2fd6f5 [Accessed 01/10/2023]
- [16] Alex Homer, John Sharp, Larry Brader, Masashi Narumoto, Trent Swanson «Cloud Design Patterns» Microsoft (2014)
- [17] Refactoring Guru [online]. Available: <https://refactoring.guru/design-patterns> [Accessed 01/06/2023]
- [18] Angular Documentation [online]. Available: <https://angular.io/> [Accessed 01/09/2023]
- [19] Εφημερίδα της Κυβερνήσεως – Νόμος 2121/1993 [online]. Available: <http://elib.aade.gr/elib/view?d=/gr/act/1993/2121> [Accessed 10/10/2023]