



Πανεπιστήμιο Πειραιά

Τμήμα Ψηφιακών Συστημάτων

Π.Μ.Σ. Διδακτική της Τεχνολογίας & Ψηφιακά Συστήματα

Κατεύθυνση: Δικτυοκεντρικά Συστήματα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«ΙΣΤΟΣΕΛΙΔΑ ΑΞΙΟΛΟΓΗΣΗΣ
ΕΡΕΥΝΗΤΩΝ»**

“Msc Thesis Researchers Rating Site”

ΦΟΙΤΗΤΗΣ :

ΓΚΙΝΗΣ ΕΥΑΓΓΕΛΟΣ ΜΕ11056

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ

Δουλκερίδης Χρήστος

ΑΘΗΝΑ, ΝΟΕΒΡΙΟΣ 2014

Πανεπιστήμιο Πειραιώς

*“It Has Become
Appallingly Obvious
That Our Technology
Has Exceeded Our
Humanity.”*

Albert Einstein

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον Δημήτρη Περτέση για την υποστήριξή του όλα αυτά τα χρόνια της μεταπτυχιακής μου φοίτησης, τον Χρήστο Παπαζαχαρίου για τη βοήθειά του στο *data mining* και τον επιβλέπων καθηγητή στην πτυχιακή αυτή και τον κύριο Χρήστο Δουλκερίδη για τον επαγγελματισμό του, τη σοβαρότητα του, την ανεκτικότητα του όποτε έπρεπε και φυσικά την αμέριστη υποστήριξή του στην εκπόνηση αυτής της εργασίας, από την πιο μικρή λεπτομέρεια έως και την πιο σημαντική βοήθεια.

Αφιερώνεται στη Λίζα.

Πανεπιστήμιο Πειραιώς

Table of Contents

Κεφάλαιο 1^ο : Εισαγωγή - Στόχος	8
1.1 Έρευνα.....	8
1.1.1 Είδη ερευνών.....	8
1.2 Ερευνητικά άρθρα.....	9
1.3 Ένας αντικειμενικός τρόπος κατάταξης άρθρων και ερευνητών.....	9
1.4 Στόχος.....	9
1.5 Διάρθρωση εργασίας.....	10
Κεφάλαιο 2^ο : Βασικές αρχές συγγραφής επιστημονικών κειμένων	12
2.1 Είδη δημοσιεύσεων.....	12
2.1.1 Ερευνητικά άρθρα.....	12
2.1.2 Ανασκοπήσεις.....	14
2.1.3 Γράμματα στον εκδότη.....	15
2.2 Βασικές αρχές συγγραφής.....	16
2.2.1 Σαφήνεια.....	16
2.2.2 Λιτότητα.....	16
2.2.3 Απλότητα.....	17
2.2.4 Ακρίβεια.....	17
2.2.5 Συνέπεια.....	17
2.2.6 Ενεργητική ή παθητική φωνή.....	18
2.2.7 Αγγλική γλώσσα.....	19
2.2.8 Οδηγίες συγγραφής από περιοδικά.....	19
Κεφάλαιο 3^ο : Infometrics	20
3.1 Bibliometrics.....	20
3.1.1 Χρησιμότητα.....	20
3.1.2 Ιστορία.....	21
3.2 Citation Index.....	22
3.2.1 Ανάλυση Αναφορών.....	22
3.3 Μέθοδοι υπολογισμού αντίκτυπου άρθρων και παραπομπών (Citation Impact Calculation).....	23
3.3.1 Citation Count.....	23
3.3.2 h-index.....	24
3.3.2.1 Υπολογισμός.....	24

3.3.2.2 Πλεονεκτήματα	25
3.3.2.3 Μειονεκτήματα	26
3.3.3 g-index	27
3.3.4 Google Scholar	28
3.3.4.1 Google Scholar Citations.....	28
3.3.4.2 Google Scholar Metrics.....	29
3.4 Βαθμολόγηση περιοδικών	29
3.4.1 Impact Factor (παράγοντας επιρροής).....	30
3.4.1.1 Υπολογισμός	30
3.4.1.2 Μειονεκτήματα	31
3.4.1.3 Αλγόριθμος PageRank	31
3.4.2 Eigenfactor.....	32
3.4.2.1 Υπολογισμός	32
3.4.2.2 Πλεονεκτήματα	33
3.4.3 Ένδειξη SJR.....	33
3.4.3.1 Υπολογισμός	33
3.5 Altmetrics	34
3.5.1 Κατηγορίες παραγόντων	34
3.5.2 Μειονεκτήματα	35
Κεφάλαιο 4^ο : DBLP (<i>Digital Bibliography & Library Project</i>).....	36
4.1 Ανάγκη για μια συνολική βάση αναφοράς στην πληροφορική.....	36
4.2 Τι είναι το DBLP	36
4.3 Αρχιτεκτονική του DBLP	37
4.3.1 Άρθρα περιοδικών.....	37
4.3.2 Άρθρα συνεδρίων.....	38
4.3.3 Συγγραφείς - Ερευνητές	39
Κεφάλαιο 5^ο : Αρχιτεκτονική Εργασίας.....	41
5.1 Συλλογή των δεδομένων	41
5.2 Κατασκευή της ιστοσελίδας	41
5.2.1 Αναζήτηση ερευνητών	41
5.2.2 Αξιολόγηση ερευνητών	42
5.3 Διάγραμμα ροής.....	43
5.4 Site Diagram	44
Κεφάλαιο 6^ο : Χρησιμοποιούμενες Τεχνολογίες.....	45

6.1 XAMPP	45
6.2 Apache HTTP.....	46
6.3 MySQL.....	47
6.4 PHP	47
6.5 SAX Parser.....	48
Κεφάλαιο 7^ο : Δημιουργία Βάσης Δεδομένων.....	53
7.1 Δομή της βάσης.....	53
7.2 Database Schema	55
7.3 Έτοιμοι πίνακες βαθμολόγησης.....	55
Κεφάλαιο 8^ο : Αναζήτηση Καθηγητών	57
8.1 Απλή αναζήτηση καθηγητών.....	57
8.2 Σύνθετη αναζήτηση καθηγητών.....	58
Κεφάλαιο 9^ο : Αξιολόγηση Καθηγητών	62
9.1 Έτοιμες λίστες βαθμολόγησης	62
9.2 Βαθμολόγηση	63
9.3 Σύγκριση	65
9.4 Προσθήκη νέων λιστών βαθμολόγησης	66
Κεφάλαιο 10^ο : Επεκτασιμότητα - Επίλογος	68
10.1 Επεκτασιμότητα	68
10.2 Επίλογος	68
Βιβλιογραφία.....	69

Κεφάλαιο 1^ο: Εισαγωγή - Στόχος

1.1 Έρευνα

Ο όρος έρευνα στις καθημερινές εκφράσεις φέρεται με πολλά αντικείμενα εφαρμογής. Περισσότερο όμως φέρεται με μια εξειδικευμένη σημασία τη συστηματική, αντικειμενική και εξακριβωμένη αναζήτηση πληροφοριών προς επίλυση κάποιου προβλήματος. Μεταξύ τους οι διάφορες έρευνες διαφέρουν ως προς τις προθέσεις, τα μέσα αλλά και στο βαθμό σεβασμού των επιστημονικών αρχών εκ μέρους των διαφόρων ερευνητών.

Σε γενικές γραμμές η έρευνα αποτελεί πρωταρχικής σημασίας διαδικασία για την ανάδειξη, συγκρότηση και προαγωγή των επιστημών, όπου μέσω αυτών επιχειρείται ευρύτερα η βελτίωση των συνθηκών της ζωής των ανθρώπων. Πολλοί είναι εκείνοι που περιορίζουν την έννοια μόνο στην επιστημονική έρευνα που αναπτύσσεται κατά κλάδο επιστημονικών πεδίων, όπου βασίζεται σε πειραματική μέθοδο, ή την επαλήθευση υποθέσεων αν και σε πολλές των περιπτώσεων η επιστήμη ερευνά "στα τυφλά" όπου οι όποιες επιβεβαιώσεις έπονται στη συνέχεια.

1.1.1 Είδη ερευνών

Βασική διάκριση των διαφόρων ερευνών λαμβάνοντας υπόψη

- α) τον τρόπο της διαδικασίας τους είναι οι "μη πειραματικές" και οι "πειραματικές έρευνες,
- β) ανάλογα του αριθμού των ερευνητών σε ατομικές και ομαδικές έρευνες", και
- γ) ανάλογα του σκοπού σε "γενικές", ή "θεωρητικές" και σε "εφαρμοσμένες έρευνες".

Γενικά οι μη πειραματικές είναι οι επιτόπιες έρευνες που περιορίζονται σε κάποια περιγραφή καταστάσεων παρά σε σχέσεις με αιτιώδη συνάφεια. Αυτές εκτείνονται κυρίως σε πληθυσμιακά δείγματα, απ' όπου οι ερευνητές εξάγουν διάφορα συμπεράσματα "κατανόησης" ή "ερμηνείας" των προς έρευνα "φαινομένων". Σ' αυτού του είδους ερευνών συγκαταλέγονται εκτός του μεγάλου πλήθους δημοσκοπήσεων, των λεγομένων γκάλοπ, η ιστορική έρευνα καθώς και η περιγραφική έρευνα.

Οι πειραματικές έρευνες συμπεριφοράς ασκούνται σε περιορισμένα πληθυσμιακά δείγματα όπου προϋποθέτουν μακρά εξέταση φαινομένων προκειμένου να εξαχθεί μια θεμελιωμένη "ερμηνεία".

1.2 Ερευνητικά άρθρα

Η συγγραφή και η δημοσίευση άρθρων αποτελεί πρόκληση, με τα χιλιάδες επιστημονικά περιοδικά να παρέχουν τη δυνατότητα ακόμη και στους πιο αδαείς να δοκιμάσουν τη συγγραφική τους τύχη. Εντούτοις, αρκετοί επιστήμονες αποφεύγουν τη συγγραφή άρθρων, θεωρώντας ότι οι περιορισμένες γνώσεις τους και η απειρία τους αναφορικά με τις δημοσιεύσεις σε επιστημονικά περιοδικά μειώνουν σημαντικά την πιθανότητα επιτυχίας. Η συγγραφή και η δημοσίευση άρθρων αποτελεί επίπονη και χρονοβόρα διαδικασία, απαιτώντας την τήρηση αυστηρά καθορισμένων κριτηρίων. Η αδυναμία συγγραφής ενός άρθρου με τον επιστημονικά αποδεκτό τρόπο μπορεί να καθυστερήσει σημαντικά ή ακόμη χειρότερα να εμποδίσει τη δημοσίευσή του.

Είναι σαφές ότι η ανάγνωση του παρόντος (ή οποιουδήποτε άλλου αντίστοιχου άρθρου ή βιβλίου) δε δημιουργεί αυτόματα εξαιρετικούς συγγραφείς. Εντούτοις, μπορεί να συμβάλλει αποφασιστικά στη βελτίωση της συγγραφικής ικανότητας και στον περιορισμό των λαθών και να αποτελέσει ερέθισμα για μεγαλύτερη συμμετοχή των ερευνητών σε συγγραφικές προσπάθειες. Η συγγραφή είναι δύσκολη, αλλά όχι και αδύνατη και απαιτεί υπομονή και συστηματική προσπάθεια. Η καλοπροαίρετη κριτική είναι απαραίτητη για τον εντοπισμό και τη διόρθωση λαθών και παραλείψεων.

1.3 Ένας αντικειμενικός τρόπος κατάταξης άρθρων και ερευνητών

Στόχος αυτής της πτυχιακής είναι η δημιουργία μιας ιστοσελίδας που σκοπό έχει την αξιολόγηση των καθηγητών που ασχολούνται με την έρευνα ανά τον κόσμο με βάση τις υπάρχουσες δημοσιεύσεις τους σε γνωστά και μη συνέδρια και περιοδικά. Η αξιολόγηση δεν θα είχε νόημα αν δεν γινόταν όσο το δυνατόν πιο αντικειμενικά και για αυτό το σκοπό πραγματοποιείται με διάφορα filters που θα μπορούν να παραμετροποιούνται κάθε φορά όπως επίσης με έτοιμες βαθμολογίες συνεδρίων/περιοδικών που θα είναι στη διάθεση του χρήστη προς επιλογή.

1.4 Στόχος

Στόχος της εργασίας όπως προαναφέρθηκε είναι η βαθμολόγηση των καθηγητών με βάση τις δημοσιεύσεις τους.

Για να επιτευχθεί αυτό πρέπει να αντληθούν δεδομένα από μία πηγή πληροφοριών που θα περιέχει συνέδρια, περιοδικά, άρθρα δημοσιευμένα σε αυτά, στοιχεία καθηγητών κλπ. Αυτή η πηγή για εμάς είναι το [dblp](#) και συγκεκριμένα όλα τα δεδομένα είναι συγκεντρωμένα σ' ένα αρχείο το dblp.xml. Επειδή όμως δεν χρειάζονται όλα τα δεδομένα από αυτό το αρχείο, πρέπει να γίνει επιλογή αυτών (parsing), μετατροπή αυτών σε κατάλληλη μορφή για να μπορούν να εισαχθούν σε μια βάση δεδομένων (editing) και τέλος εισαγωγή αυτών στη βάση.

Εν συνεχεία πάνω σε αυτή τη database θα "χτιστεί" η ιστοσελίδα η οποία πέρα από βαθμολόγηση θα παρέχει και μια μηχανή αναζήτησης καθηγητών, πάλι με διάφορα filters, για να μπορεί ο χρήστης να διαλέξει ποιον καθηγητή επιθυμεί να βαθμολογήσει.

Τέλος, για να μπορεί ο χρήστης να αξιολογήσει έναν καθηγητή θα έχει και τις κατάλληλες πληροφορίες δηλαδή το ιστορικό του καθηγητή σε δημοσιεύσεις άρθρων και λίστες με έτοιμες βαθμολογίες συνεδρίων/περιοδικών για να διαλέξει ποιές θεωρεί πιο αντικειμενικές για να συμπεριλάβει στη βαθμολόγηση του.

1.5 Διάρθρωση εργασίας

Στο πρώτο κεφάλαιο γίνεται μια εισαγωγή στην έρευνα και τα ερευνητικά άρθρα και διατυπώνεται ο στόχος της εργασίας.

Στο δεύτερο κεφάλαιο δίνεται ένας ορισμός για το τι είναι ένα ερευνητικό άρθρο, περιγράφεται η δομή του, ο σκοπός του, είδη άρθρων κλπ.

Στο τρίτο κεφάλαιο γίνεται μια εισαγωγή στα Infometrics (πληροφοριομετρία) και περιγράφεται πως γίνεται σήμερα η καταγραφή και βαθμολόγηση των ερευνητικών δημοσιεύσεων.

Στο τέταρτο κεφάλαιο γίνεται ανάλυση του dblp αρχείου ως προς τα δεδομένα που περιέχει και τη δομή του.

Στο πέμπτο κεφάλαιο περιγράφεται η αρχιτεκτονική της εργασίας μαζί με το διάγραμμα ροής.

Στο έκτο κεφάλαιο γίνεται μια αναφορά στις χρησιμοποιούμενες τεχνολογίες για την υλοποίηση αυτής της εργασίας.

Στο έβδομο κεφάλαιο αναλύεται η διαδικασία της δημιουργίας της βάσης δεδομένων καθώς και η δομή της.

Στο όγδοο κεφάλαιο περιγράφεται με ποιους τρόπους μπορεί να γίνει αναζήτηση καθηγητών.

Στο ένατο κεφάλαιο περιγράφεται με ποιους τρόπους μπορεί να γίνει η αξιολόγηση ενός καθηγητή.

Στο δέκατο κεφάλαιο περιγράφεται πως μπορεί να επεκταθεί η εργασία μαζί με τον επίλογο.

Στο ενδέκατο κεφάλαιο υπάρχει η βιβλιογραφία της εργασίας.

Κεφάλαιο 2^ο : Βασικές αρχές συγγραφής επιστημονικών κειμένων

2.1 Είδη δημοσιεύσεων

Πριν από την ανάλυση των βασικών αρχών συγγραφής κρίνεται σκόπιμο να γίνει μια σύντομη αναφορά στα είδη των δημοσιεύσεων. Το ενδιαφέρον επικεντρώνεται συνήθως στη συγγραφή και στη δημοσίευση ερευνητικών άρθρων, καθώς η πλειονότητα των δημοσιεύσεων στα επιστημονικά περιοδικά (scientific journals) αφορά σε ερευνητικά άρθρα, ενώ λιγότερο συχνά εμφανίζονται οι ανασκοπήσεις, οι αναφορές περιπτώσεων, τα γράμματα στον εκδότη, τα σχόλια, οι παρουσιάσεις βιβλίων κ.ά. Επισημαίνεται, ότι ο μεγάλος αριθμός των περιοδικών που αφορούν στις επιστήμες υγείας για παράδειγμα έχει ως αποτέλεσμα, σε ορισμένες περιπτώσεις, να χρησιμοποιούνται διαφορετικοί όροι για τα διάφορα είδη δημοσιεύσεων, με αποτέλεσμα να δημιουργείται σύγχυση. Για παράδειγμα, το “The Journal of the American Medical Association” χρησιμοποιεί τον όρο “original contribution”, ενώ το “Pediatrics” χρησιμοποιεί τον όρο “regular article” αντί των όρων “research article” ή “research paper” που χρησιμοποιούνται στην πλειονότητα των περιοδικών για να δηλώσουν τα ερευνητικά άρθρα που αφορούν σε μελέτες. Για την αποφυγή παρερμηνειών και για τη διευκόλυνση των συγγραφέων, των εκδοτών, των κριτών και των αναγνωστών είναι απαραίτητη η προσεκτική ανάγνωση και η αυστηρή τήρηση των οδηγιών συγγραφής των περιοδικών.

2.1.1 Ερευνητικά άρθρα

Η συγγραφή ερευνητικών άρθρων αποτελεί την πιο σύνθετη και επίπονη διαδικασία και αφορά ουσιαστικά στη δημοσίευση μελετών. Το αποτέλεσμα (ή, αλλιώς, η πληροφορία) που προέρχεται από μια μελέτη (study) γίνεται ευρέως γνωστό στην επιστημονική κοινότητα μέσω της δημοσίευσης σε έγκριτα επιστημονικά περιοδικά. Η ολοκλήρωση μιας μελέτης επιτυγχάνεται με τη δημοσίευση των αποτελεσμάτων της.

Στην επιστήμη της πληροφορικής, οι ερευνητές αρχικά καθορίζουν το ερευνητικό τους ερώτημα και πραγματοποιούν την αντίστοιχη συστηματική βιβλιογραφική ανασκόπηση, έπειτα σχεδιάζουν και πραγματοποιούν την κατάλληλη μελέτη, ακολούθως συλλέγουν και αναλύουν τα απαιτούμενα δεδομένα και τελικά γράφουν και υποβάλλουν για δημοσίευση το αντίστοιχο ερευνητικό άρθρο (εικόνα 2.1). Σημειώνεται ότι με τον όρο «ερευνητικό

άρθρο» αποδίδονται οι αγγλικοί όροι “research article” και “research paper”, με τον πρώτο όρο να θεωρείται ορθότερος και να προτιμάται έναντι του δεύτερου.



Εικόνα 2.1 Διαδικασία πραγματοποίησης μελετών

Σήμερα, η πλειονότητα των επιστημονικών περιοδικών έχει υιοθετήσει και εφαρμόζει μια κοινή δομή για τα ερευνητικά άρθρα, γεγονός που διευκολύνει σημαντικά τους ερευνητές, τους εκδότες, τους κριτές και τους αναγνώστες. Σύμφωνα με τη δομή αυτή, τα ερευνητικά άρθρα αποτελούνται από τα εξής μέρη:

- Εισαγωγή (introduction)
- Μέθοδοι (methods)
- Αποτελέσματα (results)
- Συζήτηση (discussion).

Επιπλέον, σε ένα ερευνητικό άρθρο υπάρχουν οπωσδήποτε τίτλος, συγγραφείς, περίληψη, λέξεις-κλειδιά, βιβλιογραφικές παραπομπές και πίνακες και ενδεχομένως γραφήματα, φωτογραφίες, ευχαριστίες, αντικρουόμενα συμφέροντα και οικονομική υποστήριξη για τη διεξαγωγή της μελέτης.

2.1.2 Ανασκοπήσεις

Η ανασκόπηση (review) αφορά στην ανασκόπηση της βιβλιογραφίας (literature review) σχετικά με ένα συγκεκριμένο επιστημονικό θέμα και αποβλέπει στην αναγνώριση, την εκτίμηση και την ερμηνεία όσο το δυνατόν περισσότερων μελετών σχετικά με το συγκεκριμένο θέμα. Μια ανασκόπηση δεν αποτελεί πρωτογενή έρευνα, δεν αποτελεί δηλαδή νέα μελέτη των ερευνητών, αλλά αποτελεί ουσιαστικά περιγραφή και αξιολόγηση των μελετών που ήδη έχουν πραγματοποιηθεί. Οι ανασκοπήσεις δεν περιλαμβάνουν ανάλυση δεδομένων, αλλά μια εκτεταμένη και σε βάθος προσέγγιση ενός επιστημονικού θέματος, αποβλέποντας στην εξαγωγή ασφαλών συμπερασμάτων με βάση την υπάρχουσα βιβλιογραφία. Οι ανασκοπήσεις της σχετικής κατηγορίας είναι γνωστές και ως αφηγηματικές ανασκοπήσεις (narrative reviews).

Οι αφηγηματικές ανασκοπήσεις δε χαρακτηρίζονται από μια συγκεκριμένη δομή σε αντίθεση με τα ερευνητικά άρθρα. Σίγουρα, πάντως, στις ανασκοπήσεις αυτές, υπάρχει εισαγωγή, συζήτηση, συμπεράσματα, βιβλιογραφικές παραπομπές και περίληψη. Πρόκειται για εκτεταμένες δημοσιεύσεις, φθάνοντας μέχρι και τις 5.000 λέξεις, με μεγάλο αριθμό βιβλιογραφικών παραπομπών (μέχρι και 100). Ο τίτλος και η εισαγωγή μιας αφηγηματικής ανασκόπησης είναι καθοριστικοί παράγοντες στο να προκαλέσει η ανασκόπηση το ενδιαφέρον των αναγνωστών.

Σημειώνεται η σημαντική διαφορά μεταξύ αφηγηματικής ανασκόπησης και συστηματικής ανασκόπησης, με τη δεύτερη να περιλαμβάνει και τη διενέργεια στατιστικής ανάλυσης, τη γνωστή μετα-ανάλυση. Η συστηματική ανασκόπηση (systematic review) αποτελεί τη θεωρητική διαδικασία καθορισμού με ορισμένα κριτήρια των καλύτερα μεθοδολογικά σχεδιασμένων μελετών σχετικά με μια συγκεκριμένη επιστημονική υπόθεση, ενώ η μετα-ανάλυση (meta-analysis) αποτελεί μια μαθηματική διαδικασία που συνδυάζει στατιστικά τα αποτελέσματα των μελετών που επιλέχθηκαν έπειτα από τη συστηματική ανασκόπηση και εξάγει ένα συγκεντρωτικό αποτέλεσμα. Είναι σαφές ότι η εξαγωγή ασφαλών συμπερασμάτων με την εφαρμογή της μετα-ανάλυσης προϋποθέτει μια καλά σχεδιασμένη συστηματική ανασκόπηση της βιβλιογραφίας, έτσι ώστε να συμπεριληφθούν στη μετα-ανάλυση οι πλέον κατάλληλες μελέτες. Συνεπώς, τα συμπεράσματα της μετα-ανάλυσης είναι ασφαλή μόνον εφόσον έχει προηγηθεί ενδελεχής συστηματική ανασκόπηση όλων των μελετών που ενδεχομένως θα μπορούσαν να συμπεριληφθούν στη μετα-ανάλυση. Επιπλέον, η διεξαγωγή μόνο της συστηματικής ανασκόπησης χωρίς την εφαρμογή της μετα-ανάλυσης για τον υπολογισμό ενός συνδυαστικού αποτελέσματος με βάση τα ξεχωριστά

αποτελέσματα των επιμέρους μελετών αποτελεί μια μη ολοκληρωμένη διαδικασία, καθώς δεν εξάγεται ένα συγκεντρωτικό αποτέλεσμα.

Τα άρθρα που αφορούν σε μετα-αναλύσεις συνεχώς αυξάνονται, προσφέροντας πολύτιμη πληροφορία, καθώς συγκεντρώνουν την πληροφορία που προέρχεται από τις επιμέρους μελέτες μειώνοντας έτσι σημαντικά το τυχαίο σφάλμα. Στις μετα-αναλύσεις δεν πραγματοποιείται ανάλυση πρωτογενών δεδομένων, αλλά ανάλυση των αποτελεσμάτων (ή σπανιότερα των δεδομένων) που προέρχονται από μελέτες που έχουν ήδη διεξαχθεί. Σημειώνεται, ότι η δομή των άρθρων που αφορούν στις μετα-αναλύσεις είναι παρόμοια με τη δομή των ερευνητικών άρθρων.

2.1.3 Γράμματα στον εκδότη

Τα γράμματα στον εκδότη (letters to the editor) κατέχουν ιδιαίτερη θέση στα επιστημονικά περιοδικά, αποτελώντας σύντομα και περιεκτικά σχόλια σε άρθρα που έχουν ήδη δημοσιευτεί. Τα γράμματα στον εκδότη είναι ιδιαίτερης σημασίας, καθώς παρέχουν τη δυνατότητα επικοινωνίας και ανταλλαγής απόψεων μεταξύ των ερευνητών. Με τον τρόπο αυτόν, προάγεται ο επιστημονικός διάλογος και στην πράξη παρέχεται η δυνατότητα σε όλους τους αναγνώστες να κρίνουν και να σχολιάσουν τις ερευνητικές προσπάθειες που πραγματοποιούνται διεθνώς.

Ενδεικτικό της αξίας που έχουν τα γράμματα στον εκδότη είναι το γεγονός ότι το Δεκέμβριο του 1961, το *Lancet* δημοσίευσε την παρατήρηση του Αυστραλού γυναικολόγου και μαιευτήρα William McBride ότι τρία νεογνά (σε χρονικό διάστημα ενός μήνα), οι μητέρες των οποίων έλαβαν θαλιδομίδη στους πρώτους μήνες της κύησης, γεννήθηκαν με σοβαρές γενετικές ανωμαλίες (μη αρτιμελή μωρά). Η δημοσίευση αυτή είχε ως αποτέλεσμα να διακοπεί άμεσα η χορήγηση της θαλιδομίδης και να αποφευχθούν περαιτέρω τραγικές συνέπειες.

Τα γράμματα στον εκδότη πρέπει να είναι σύντομα και περιεκτικά και να υποβάλλονται για δημοσίευση μέσα σε μικρό χρονικό διάστημα (συνήθως εντός δύο μηνών) από την ημέρα δημοσίευσης του άρθρου, στο οποίο αναφέρονται. Οι οδηγίες συγγραφής των περιοδικών διαφέρουν σημαντικά μεταξύ τους και για το λόγο αυτόν ακολουθήστε πιστά τις οδηγίες του περιοδικού, στο οποίο πρόκειται να υποβάλλετε το γράμμα σας. Για παράδειγμα, στο *Lancet*, η υποβολή του γράμματος πρέπει να γίνεται εντός 15 ημερών από την ημέρα δημοσίευσης του άρθρου στο οποίο αναφέρεται, ο αριθμός των λέξεων δεν πρέπει να υπερβαίνει τις 250 και επιτρέπονται ένας πίνακας ή ένα γράφημα, 1–5 βιβλιογραφικές παραπομπές και 1–5 συγγραφείς, ενώ στο “*The Journal of the American Medical Association*”, η υποβολή του γράμματος πρέπει να γίνεται εντός 30 ημερών από την ημέρα δημοσίευσης του άρθρου στο οποίο αναφέρεται, ο αριθμός των λέξεων δεν πρέπει να υπερβαίνει τις 400 και επιτρέπονται 1–5 βιβλιογραφικές παραπομπές και 1–3 συγγραφείς.

2.2 Βασικές αρχές συγγραφής

2.2.1 Σαφήνεια

Χαρακτηριστική είναι η άποψη του Γάλλου συγγραφέα Albert Camus ότι «αυτοί που γράφουν με σαφήνεια έχουν αναγνώστες, ενώ αυτοί που γράφουν δυσνόητα έχουν σχολιαστές».

Το βασικό χαρακτηριστικό της επιτυχούς επιστημονικής συγγραφής είναι η σαφήνεια. Η ξεκάθαρη και αποκρυσταλλωμένη σκέψη είναι απαραίτητη προϋπόθεση για τη συγγραφή ενός κειμένου με σαφή και κατανοητό τρόπο. Η αδυναμία του συγγραφέα να εστιαστεί σε ένα ξεκάθαρο μήνυμα αποτελεί βασική αιτία αποτυχίας. Το κύριο μήνυμα του συγγραφέα μεταφέρεται ευκολότερα στους αναγνώστες με την παράθεση μιας απλής, σύντομης και εύληπτης πρότασης. Συνήθως, ο τίτλος της δημοσίευσης εξυπηρετεί το σκοπό αυτόν.

Εάν ο συγγραφέας δεν έχει κατανοήσει σε βάθος το θέμα το οποίο πραγματεύεται, τότε είναι σχεδόν απίθανο να συγγράψει ένα κατανοητό κείμενο. Η σύγχυση στη σκέψη του μεταφέρεται στο κείμενό του και βέβαια και στους αναγνώστες. Απαιτείται δομημένη επιχειρηματολογία, έτσι ώστε το μήνυμα του συγγραφέα να μεταφέρεται βήμα προς βήμα με λογικό τρόπο. Οι προτάσεις πρέπει να είναι απλές, σαφείς και εύκολα κατανοητές, έτσι ώστε οι αναγνώστες να μην κοπιάζουν και να μπορούν να αντιλαμβάνονται τα συγγραφικά μηνύματα. Συνήθως, οι συντομότερες προτάσεις είναι απλούστερες και ευκολότερα κατανοητές. Είναι απαραίτητο εξάλλου οι λέξεις, οι προτάσεις και οι παράγραφοι να χαρακτηρίζονται από λογική συνέπεια και συνάφεια μεταξύ τους, έτσι ώστε το κείμενο να εμφανίζει μια λογική ροή.

2.2.2 Λιτότητα

Η λιτότητα είναι απαραίτητη στη συγγραφή επιστημονικών κειμένων, καθώς η χρήση περιττών λέξεων, προτάσεων και παραγράφων συχνά δημιουργεί σύγχυση και απογοήτευση στους αναγνώστες. Οι προτάσεις πρέπει να είναι όσο το δυνατό πιο απλές, ενώ είναι προτιμότερο οι μεγάλες προτάσεις να διασπώνται σε επιμέρους μικρότερες και απλούστερες. Επιπλέον, πρέπει να αποφεύγονται οι επαναλήψεις και να αφαιρούνται οι προτάσεις που δεν προσφέρουν χρήσιμη πληροφορία. Δεν πρέπει να παραγνωρίζεται εξάλλου το γεγονός ότι τα περιοδικά διαθέτουν περιορισμένο διαθέσιμο χώρο και γι' αυτό τα μικρότερα σε έκταση άρθρα πλεονεκτούν έναντι των μεγαλύτερων και έχουν μεγαλύτερη πιθανότητα δημοσίευσης.

Σύμφωνα με την αρχή της οικονομίας ή την αρχή της λιτότητας ή απλούστερα «ξυράφι του Όκαμ» (Ockham's razor), δεν πρέπει να υπάρχουν περισσότερες εικασίες από όσες είναι

απαραίτητες ή, με άλλη διατύπωση, μεταξύ δύο θεωριών ή εξηγήσεων, όταν όλοι οι υπόλοιποι παράγοντες είναι ταυτόσημοι, προτιμάται η απλούστερη.

2.2.3 Απλότητα

Αρκετοί συγγραφείς, λανθασμένα, προσπαθούν να κεντρίσουν το ενδιαφέρον του αναγνωστικού κοινού χρησιμοποιώντας εντυπωσιακές ή πομπώδεις λέξεις και εκφράσεις. Με τον τρόπο αυτόν δημιουργούνται παρερμηνείες και οι αναγνώστες όχι μόνο δεν εντυπωσιάζονται αλλά αντίθετα δυσκολεύονται να κατανοήσουν το κείμενο και απογοητεύονται. Είναι σαφές ότι σκοπός των συγγραφέων πρέπει να είναι η ενημέρωση των αναγνωστών με απλό, ξεκάθαρο και επιστημονικό τρόπο και όχι ο εντυπωσιασμός τους. Τα επιστημονικά κείμενα δεν είναι λογοτεχνικά και για το λόγο αυτόν πρέπει να αποφεύγονται εκφράσεις που ενδέχεται να δημιουργήσουν σύγχυση, όπως μεταφορικές έννοιες, παρομοιώσεις, ιδιωματισμοί κ.ά.

2.2.4 Ακρίβεια

Τα επιστημονικά κείμενα πρέπει να χαρακτηρίζονται από ακρίβεια και όχι από αοριστίες ή ασάφειες. Για παράδειγμα, η πρόταση «η χειρουργική παρέμβαση πραγματοποιήθηκε σε 50 νεαρές γυναίκες» δημιουργεί σύγχυση. Είναι σαφές ότι πρόκειται για νεαρές γυναίκες, αλλά δεν είναι σαφής η ηλικιακή ομάδα στην οποία ανήκουν. Ορθότερη και ακριβέστερη θα ήταν η πρόταση "η χειρουργική παρέμβαση πραγματοποιήθηκε σε 50 γυναίκες, ηλικίας 20–30 ετών» ή ακόμη καλύτερα «η χειρουργική παρέμβαση πραγματοποιήθηκε σε 50 γυναίκες με μέση ηλικία τα 26,2 έτη (τυπική απόκλιση: 5,5)".

Ιδιαίτερη προσοχή χρειάζεται στην περίπτωση που χρησιμοποιούνται στατιστικοί όροι. Για παράδειγμα, δεν πρέπει να χρησιμοποιούνται οι όροι προοπτικές και αναδρομικές μελέτες, αλλά να ορίζεται επακριβώς το είδος της μελέτης. Ο όρος "δείγμα" εξάλλου προκαλεί ιδιαίτερη σύγχυση και παρερμηνείες και ως ορθότερος προτείνεται ο όρος "μελετώμενος πληθυσμός".

2.2.5 Συνέπεια

Το κείμενο γίνεται απλούστερο και ευκολότερα κατανοητό, όταν χαρακτηρίζεται από συνέπεια. Όταν δηλαδή, οι ίδιοι όροι χρησιμοποιούνται με τον ίδιον τρόπο σε όλη την

έκταση του κειμένου. Για παράδειγμα, εάν χρησιμοποιηθεί για πρώτη φορά στο κείμενο ο όρος «πάσχοντες» είναι προτιμότερο να διατηρηθεί ο όρος αυτός σε όλο το κείμενο παρά να χρησιμοποιούνται εναλλακτικά οι όροι «άρρωστοι», «ασθενείς», «νοσηλεύόμενοι», «συμμετέχοντες» κ.ά.

Επισημαίνεται ότι η σωστή χρήση των χρόνων των ρημάτων είναι καθοριστικής σημασίας για την ορθή κατανόηση ενός ερευνητικού άρθρου. Δυστυχώς όμως, σε αρκετές περιπτώσεις, χρησιμοποιούνται λανθασμένα οι χρόνοι των ρημάτων, με αποτέλεσμα να δημιουργούνται παρερμηνείες. Ουσιαστικά, δύο χρόνοι χρησιμοποιούνται στα ερευνητικά άρθρα, ο ενεστώτας και ο αόριστος. Ο ενεστώτας χρησιμοποιείται για να παρουσιαστεί η ήδη υπάρχουσα γνώση, ενώ ο αόριστος για να παρουσιαστούν οι μέθοδοι που χρησιμοποιήθηκαν στη μελέτη, καθώς και τα αποτελέσματά της.

2.2.6 Ενεργητική ή παθητική φωνή

Η ενεργητική φωνή δηλώνει ότι το υποκείμενο του ρήματος (πρόσωπο ή πράγμα) δρα, ενώ η παθητική φωνή δηλώνει ότι το υποκείμενο δέχεται την ενέργεια του ρήματος ή επηρεάζεται από αυτή. Το άμεσο αντικείμενο της ενεργητικής φωνής αντιστοιχεί στο υποκείμενο της παθητικής, ενώ το υποκείμενο της ενεργητικής αντιστοιχεί στο ποιητικό αίτιο της παθητικής. Με την ενεργητική σύνταξη τονίζεται η δράση του υποκειμένου, δηλαδή δίνεται έμφαση σε αυτόν που πραγματοποιεί την ενέργεια, ενώ με την παθητική σύνταξη τονίζεται το αποτέλεσμα της ενέργειας του υποκειμένου, δηλαδή δίνεται έμφαση στην ενέργεια.

Η επιλογή της χρήσης της ενεργητικής ή της παθητικής φωνής έχει ιδιαίτερη σημασία στα ερευνητικά άρθρα. Η ενεργητική φωνή είναι απλούστερη, ευκολότερη και συνήθως χρησιμοποιούνται λιγότερες λέξεις. Η παθητική φωνή χαρακτηρίζεται από μεγαλύτερη αντικειμενικότητα, καθώς δίνεται έμφαση στην ενέργεια και όχι σε αυτόν που πραγματοποιεί την ενέργεια, αλλά είναι λιγότερο ευχάριστη και περισσότερο απρόσωπη και βαρετή για τους αναγνώστες.

Παλαιότερα, η παθητική φωνή χρησιμοποιείτο σε μεγαλύτερο βαθμό συγκριτικά με σήμερα. Τα επιστημονικά περιοδικά που υιοθετούν, μέσα από τις κατευθυντήριες οδηγίες τους για τους συγγραφείς, τη χρήση της ενεργητικής φωνής έναντι της παθητικής αυξάνονται συνεχώς. Η παθητική φωνή προτιμάται έναντι της ενεργητικής μόνο στην περίπτωση επιστημονικών περιοδικών που δηλώνουν ρητά ότι δεν αποδέχονται τη χρήση ενεργητικής φωνής.

Σε γενικές γραμμές, συστήνεται η χρήση της ενεργητικής φωνής έναντι της παθητικής. Σημειώνεται πάντως ότι ιδιαίτερη προσοχή και περίσκεψη απαιτείται στη συγγραφή των μεθόδων (κυρίως) και των αποτελεσμάτων ενός ερευνητικού άρθρου αναφορικά με τη χρήση της ενεργητικής ή της παθητικής φωνής. Αυτό οφείλεται στο γεγονός ότι στις μεθόδους και στα αποτελέσματα ενός ερευνητικού άρθρου δίνεται μεγαλύτερη έμφαση

στην ενέργεια παρά σε αυτόν (ή σε αυτούς) που πραγματοποιεί την ενέργεια. Έτσι, στη συγκεκριμένη περίπτωση είναι δυνατή η χρήση τόσο της ενεργητικής όσο και της παθητικής φωνής, χωρίς να επηρεάζεται η ποιότητα του κειμένου.

2.2.7 Αγγλική γλώσσα

Ο αριθμός των δημοσιεύσεων στην αγγλική γλώσσα συνεχώς αυξάνεται. Ενδεικτικά αναφέρεται ότι στη μηχανή αναζήτησης βιβλιογραφικών δεδομένων PubMed, το 1970, το 56% των δημοσιεύσεων ήταν στην αγγλική γλώσσα, ενώ το 2011, το ποσοστό αυτό αυξήθηκε σημαντικά, φθάνοντας το 80%.

Η σωστή συγγραφή ενός άρθρου στην αγγλική γλώσσα είναι εξαιρετικά δύσκολη για αυτούς, των οποίων η αγγλική γλώσσα δεν είναι η μητρική. Και αυτοί πάντως των οποίων η αγγλική γλώσσα είναι η μητρική τους, δε σημαίνει απαραίτητα ότι μπορούν και να γράψουν καλά στην αγγλική γλώσσα, γεγονός βέβαια που ισχύει και για οποιαδήποτε άλλη γλώσσα. Σίγουρα όμως για αυτούς, που η μητρική τους γλώσσα είναι η ελληνική, το να γράψουν καλά στην αγγλική γλώσσα είναι πιο δύσκολο σε σχέση με εκείνους των οποίων η μητρική τους γλώσσα είναι η αγγλική.

2.2.8 Οδηγίες συγγραφής από περιοδικά

Τα επιστημονικά περιοδικά διαθέτουν σαφείς οδηγίες συγγραφής (writing guidelines) των διαφόρων ειδών δημοσιεύσεων, έτσι ώστε να εξασφαλίζεται η ομοιομορφία και η σαφήνεια και να διευκολύνονται τόσο οι εκδότες και οι κριτές όσο και οι αναγνώστες. Οι εκδότες των περιοδικών δέχονται μεγάλο αριθμό άρθρων για κρίση και δεν σπαταλούν πολύτιμο χρόνο για άρθρα που δεν τηρούν τις σαφείς και ξεκάθαρες οδηγίες συγγραφής. Για παράδειγμα, εάν σε ένα περιοδικό, για τις βιβλιογραφικές παραπομπές εφαρμόζεται το σύστημα Vancouver και υποβληθεί ένα άρθρο στο οποίο έχει εφαρμοστεί το σύστημα Harvard, τότε είναι σαφές ότι οι εκδότες του περιοδικού δεν πρόκειται καν να στείλουν το άρθρο για κρίση. Στην πλειονότητα των περιπτώσεων, η μη τήρηση των οδηγιών συγγραφής των περιοδικών συνοδεύεται και από άρθρα χαμηλής ποιότητας

Κεφάλαιο 3^ο : Infometrics

Infometrics (Πληροφοριομετρία) είναι η μελέτη των ποσοτικών πτυχών των πληροφοριών. Αυτό περιλαμβάνει την παραγωγή, διάδοση και χρήση όλων των μορφών των πληροφοριών, ανεξάρτητα από τη μορφή ή την προέλευσή τους. Η πληροφοριομετρία περιλαμβάνει τους ακόλουθους τομείς:

- Scientometrics, η οποία μελετά ποσοτικές πτυχές της επιστήμης
- Webometrics, η οποία μελετά ποσοτικές πτυχές του World Wide Web
- Cybermetrics, η οποία είναι παρόμοια με webometrics, αλλά διευρύνει τον ορισμό του ώστε να συμπεριλάβει τις ηλεκτρονικές πηγές
- Bibliometrics, η οποία μελετά ποσοτικές πτυχές των καταγεγραμμένων πληροφοριών

Στην ποσοτική ανάλυση των βιβλιογραφικών δεδομένων ήταν πρωτοπόρος ο Robert K. Merton σε ένα άρθρο του που ονομάζεται Επιστήμη, Τεχνολογία και Κοινωνία τον δέκατο έβδομο αιώνα, στην Αγγλία και αρχικά δημοσιεύθηκε από τον Merton το 1938.

Εμείς θα ασχοληθούμε με την βιβλιομετρία όσο αναφορά τα δημοσιευμένα άρθρα.

3.1 Bibliometrics

Βιβλιομετρία είναι ένα σύνολο μεθόδων για την ποσοτική ανάλυση της ακαδημαϊκής βιβλιογραφίας. Η ανάλυση βιβλιογραφικών παραπομπών και η ανάλυση περιεχομένου είναι σύνηθες βιβλιομετρικές μέθοδοι. Παρά το γεγονός ότι οι βιβλιομετρικές μέθοδοι χρησιμοποιούνται πιο συχνά στις βιβλιοθήκες και στον τομέα της πληροφορικής, η βιβλιομετρία έχει ευρείες εφαρμογές και σε άλλες περιοχές. Πολλά ερευνητικά πεδία χρησιμοποιούν βιβλιομετρικές μεθόδους για να διερευνήσουν την επίδραση του πεδίου τους, τον αντίκτυπο ενός συνόλου των ερευνητών, ή την επίδραση ενός συγκεκριμένου εγγράφου.

3.1.1 Χρησιμότητα

Ιστορικά βιβλιομετρικές μέθοδοι έχουν χρησιμοποιηθεί για την ανίχνευση των σχέσεων μεταξύ αναφορών σε ακαδημαϊκά περιοδικά. Η ανάλυση της αιτιολογικής αναφοράς, η

οποία συνεπάγεται την εξέταση των αναφερόμενων εγγράφων ενός στοιχείου, χρησιμοποιείται στην αναζήτηση πληροφοριών και την ανάλυση της αξίας τους. Δείκτες παραπομπής, όπως το Institute for Scientific Information's Web of Science, επιτρέπει στους χρήστες να αναζητούν μπροστά στο χρόνο από ένα γνωστό άρθρο στις πιο πρόσφατες δημοσιεύσεις που το αναφέρουν.

Δεδομένα από δείκτες παραπομπών μπορούν να αναλυθούν για να καθορίσουν τη δημοτικότητα και τις επιπτώσεις συγκεκριμένων άρθρων, των συγγραφέων, και δημοσιεύσεων. Επίσης επιστήμονες χρησιμοποιούν την ανάλυση παραπομπών για να εκτιμήσουν ποσοτικά τους βασικούς τίτλους περιοδικών και την πληθώρα δημοσιεύσεων σε συγκεκριμένες ειδικότητες, τις αλληλεξαρτήσεις μεταξύ των δημιουργών από διάφορα ιδρύματα και σχολές σκέψης και των συναφών δεδομένων σχετικά με την ακαδημαϊκή κοινότητα. Μια πιο ρεαλιστική εφαρμογή αυτών των πληροφοριών περιλαμβάνει το σχεδιασμό της αναδρομικής βιβλιογραφίας, δίνοντας έτσι κάποια ένδειξη όσον αφορά την ηλικία του υλικού που χρησιμοποιείται και του βαθμού στον οποίο πιο πρόσφατες δημοσιεύσεις αντικαθιστούν τις παλαιότερες υποδεικνύοντας έτσι μέσω μιας υψηλής συχνότητας παραπομπών ποια έγγραφα θα πρέπει να αρχειοθετούνται, συγκρίνοντας επίσης την κάλυψη των δευτερευουσών υπηρεσιών που μπορεί να βοηθήσουν τους εκδότες να μετρήσουν τα επιτεύγματα και τον ανταγωνισμό τους, και μπορεί να βοηθήσει και τους βιβλιοθηκάρχους στην αξιολόγηση του έργου τους.

Υπάρχουν επίσης και ορισμένοι περιορισμοί σχετικά με την αξία των δεδομένων των παραπομπών. Αυτά είναι συχνά ελλιπή ή προκατειλημμένα, πολλά δεδομένα έχουν συλλεχθεί σε μεγάλο βαθμό με το χέρι (το οποίο είναι χρονοβόρο) και τυχών λανθασμένη αναφορά των πηγών επεκτείνεται κατά συνέπεια να απαιτείται περαιτέρω έρευνα για να καταλάβει κάποιος πραγματικά την λογική πίσω επικαλούμενες αναφορές ώστε να τις χρησιμοποιήσει με σιγουριά.

3.1.2 Ιστορία

Ο όρος βιβλιομετρία επινοήθηκε από τον Alan Pritchard σε ένα έγγραφο του που δημοσιεύθηκε το 1969, με τίτλο Στατιστική Βιβλιογραφία ή Βιβλιομετρία. Όρισε τον όρο ως «η εφαρμογή των μαθηματικών και στατιστικών μεθόδων για βιβλία και άλλα μέσα επικοινωνίας».

Αν και η ανάλυση παραπομπών δεν είναι νέα έννοια (η Science Citation Index άρχισε από το 1961), πριν να μπορέσει να υπολογιστεί από υπολογιστές και κατά συνέπεια γινόταν με το χέρι και έτσι ήταν χρονοβόρα. Αυτοματοποιημένοι αλγόριθμοι την καθιστούν πολύ πιο χρήσιμη, ευέλικτη και διαδεδομένη. Αυτό οδήγησε στη δημιουργία ενός νέου πεδίου: της υπολογιστικής βιβλιομετρίας. Ο πρώτος αλγόριθμος για την αυτοματοποιημένη εξαγωγή παραπομπών και ευρετηρίων ήταν από το CiteSeer. Το PageRank της Google βασίζεται στην αρχή της ανάλυσης παραπομπών. Χάρτες παραπομπών ευρεσιτεχνίας επίσης βασίζονται σε

ανάλυση παραπομπών (σε αυτή την περίπτωση, η παραπομπή από τη μία ευρεσιτεχνία στην άλλη).

3.2 Citation Index

Ένας δείκτης παραπομπών (citation index) είναι ένα είδος βιβλιογραφικής βάσης δεδομένων, ένας δείκτης των παραπομπών μεταξύ διαφορετικών εκδόσεων επιτρέποντας στο χρήστη να βρει εύκολα ποιες τωρινές δημοσιεύσεις αναφέρονται σε συγκεκριμένες παλαιότερες δημοσιεύσεις.

Το 1960, το Eugene Garfield's Institute for Scientific Information (ISI), εισήγαγε το πρώτο δείκτη παραπομπών για δημοσιεύσεις σε ακαδημαϊκά περιοδικά, πρώτα το Science Citation Index (SCI), και αργότερα Social Sciences Citation Index (SSCI) και το Arts and Humanities Citation Index (AHCI). Η πρώτη αυτοματοποιημένη παραπομπή ευρετηρίων έγινε από CiteSeer το 1997.

3.2.1 Ανάλυση Αναφορών

Παρόλο που οι δείκτες παραπομπών είχαν αρχικά σχεδιαστεί για την ανάκτηση πληροφοριών, χρησιμοποιούνται όλο και περισσότερο για βιβλιομετρία και άλλες μελέτες που αφορούν την αξιολόγηση της έρευνας. Τα βιβλιογραφικά δεδομένα είναι επίσης η βάση του παράγοντα της επιρροής ενός περιοδικού.

Το πεδίο άνθισε με την έλευση του Science Citation Index, το οποίο καλύπτει πλέον τις πηγές βιβλιογραφίας από το 1900. Τα κορυφαία περιοδικά του πεδίου είναι Scientometrics, Infometrics, και το Journal of the American Society of Information Science and Technology (ASIST). ASIST φιλοξενεί επίσης μια ηλεκτρονική λίστα αλληλογραφίας που ονομάζεται SIGMETRICS. Η μέθοδος αυτή υφίσταται μια αναβίωση η οποία βασίζεται στην ευρεία διάδοση της επιστήμης στο διαδίκτυο και σε βάσεις δεδομένων με συνδρομή σε πολλά πανεπιστήμια καθώς και στα διαθέσιμα δωρεάν εργαλεία παραπομπής όπως το Citebase, CiteSeerX, το Google Scholar, και το πρώην Windows Live Academic (τώρα διαθέσιμο με επιπλέον χαρακτηριστικά, όπως το Microsoft Academic Search).

Ανάλυση παραπομπών νομικού περιεχομένου είναι μια τεχνική για την ανάλυση νομικών εγγράφων από την εξερεύνηση στις αναφορές που συνδέουν διατάξεις σε άλλες διατάξεις μέσα στο ίδιο έγγραφο ή μεταξύ διαφορετικών εγγράφων. Αυτή η ανάλυση χρησιμοποιεί ένα γράφημα παραπομπής που εξάγεται από ένα ρυθμιστικό έγγραφο, το οποίο θα μπορούσε να συμπληρώσει το E- discovery - μια διαδικασία που δραστηριοποιείται σε τεχνολογικές καινοτομίες σε big data analytics.

3.3 Μέθοδοι υπολογισμού αντίκτυπου άρθρων και παραπομπών (Citation Impact Calculation)

Ο αντίκτυπος των παραπομπών μπορεί να μετρηθεί με διάφορους τρόπους.

3.3.1 Citation Count

Ένα προφανές μέτρο είναι η καταμέτρηση των παραπομπών, η οποία ποσοτικοποιεί τόσο τη χρήση και τον αντίκτυπο του αναφερόμενου έργου. Αυτό ονομάζεται ανάλυση παραπομπών ή βιβλιομετρία (βλ. 3.1).

Μεταξύ των μέτρων που έχουν προκύψει από την ανάλυση παραπομπών είναι αυτά που μετράνε για:

- α) ένα μεμονωμένο άρθρο (πόσο συχνά αναφέρθηκε)
- β) έναν συγγραφέα (σύνολο αναφορών, ή ο μέσος αριθμός των αναφορών ανά άρθρο)
- γ) ένα περιοδικό (μέσος αριθμός παραπομπών για τα άρθρα στο περιοδικό)

Πολλά μέτρα έχουν προταθεί, πέρα από την απλή καταμέτρηση παραπομπών, για να ποσοτικοποιηθεί καλύτερα ο αντίκτυπος μιας παραπομπής. Τα πιο γνωστά μέτρα περιλαμβάνουν το h-index και το G-INDEX. Κάθε μέτρο έχει πλεονεκτήματα και μειονεκτήματα, που εκτείνονται από την προκατάληψη και την εξάρτηση πειθαρχίας ως τους περιορισμούς της πηγής δεδομένων των αναφορών.

Μια εναλλακτική προσέγγιση για τη μέτρηση της επίδρασης του συγγραφέα βασίζεται στη χρησιμότητα των δεδομένων, όπως ο αριθμός των downloads από τους εκδότες και την ανάλυση των παραπομπών σε επίπεδο άρθρου. Μια σημαντική πρόσφατη εξέλιξη στον τομέα της έρευνας σχετικά με τις επιπτώσεις παραπομπών είναι η ανακάλυψη της καθολικότητας ή πρότυπα που είναι μεταξύ διαφόρων κλάδων των επιστημών όπως των κοινωνικών επιστημών και των ανθρωπιστικών επιστημών. Για παράδειγμα, έχει αποδειχθεί ότι ο αριθμός των αναφορών που λαμβάνονται από μια δημοσίευση, αν υπολογιστούν από το μέσο όρο τους σε διάφορα άρθρα που δημοσιεύονται στον ίδιο τομέα και την ίδια χρονιά, ακολουθούν μια καθολική λογαριθμική-κανονική κατανομή η οποία είναι ίδια τομέα. Κατά συνέπεια αυτό το εύρημα προτείνει ένα παγκόσμιο μέτρο όσο αναφορά το αντίκτυπο των αναφορών που επεκτείνει το h-index αν γίνει με επαναυπολογισμό των αριθμών των παραπομπών και στα αναφερόμενα άρθρα τους, ωστόσο ο υπολογισμός ενός τέτοιου παγκόσμιου μέτρου απαιτεί τη συλλογή των εκτεταμένων δεδομένων αναφορών και στατιστικών στοιχείων για κάθε τομέα και έτος. Κοινωνικά crowdsourcing εργαλεία, όπως Scholarometer έχουν προταθεί για την αντιμετώπιση αυτής της ανάγκης.

Η αυτοματοποιημένη ευρετηρίαση παραπομπών έχει αλλάξει τη φύση της έρευνας της ανάλυσης παραπομπών, επιτρέποντας σε εκατομμύρια αναφορές να αναλυθούν σε μεγάλου μεγέθους κλίμακα για την ανακάλυψη γνώσης. Το Google Scholar πρόσφατα, πρότεινε προηγμένα μοντέλα για τη δυναμική ανάλυση παραπομπών ως προς την παλαιότητά τους. Το τελευταίο μοντέλο ακόμα χρησιμοποιείται ως εργαλείο πρόβλεψης για τον καθορισμό βιβλιογραφικών αναφορών που μπορούν να ανακτηθούν ανά πάσα στιγμή στο κύκλο ζωής μιας έκδοσης.

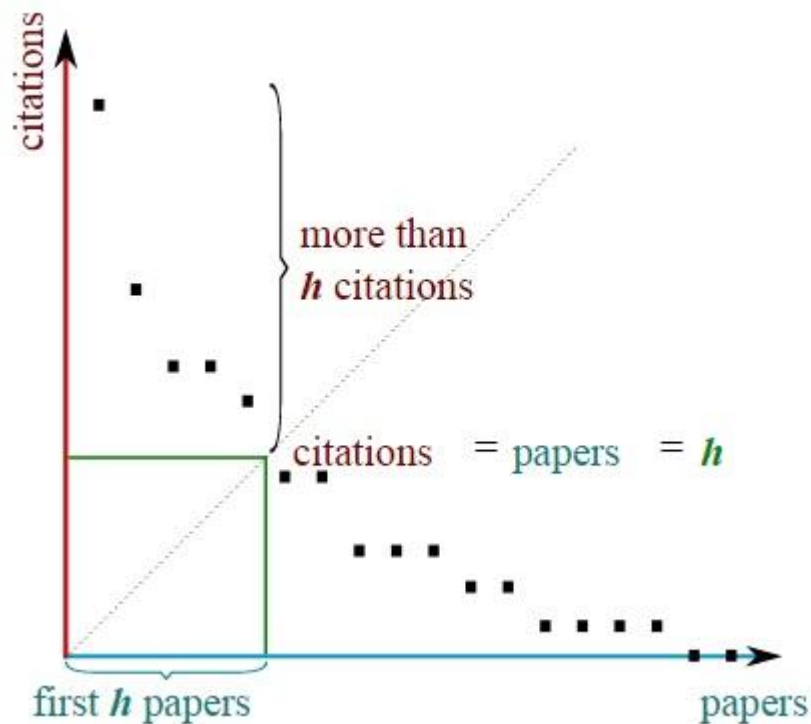
3.3.2 h-index

Το h-index είναι ένας δείκτης που προσπαθεί να μετρήσει τόσο την παραγωγικότητα όσο και την παραπομπή των επιπτώσεων των αναφορών από το δημοσιευμένο έργο ενός ερευνητή. Ο δείκτης βασίζεται στο σύνολο των πιο αναφερόμενων δημοσιεύσεων ενός επιστήμονα και στον αριθμό των αναφορών που έχει λάβει από άλλες δημοσιεύσεις. Ο δείκτης μπορεί επίσης να εφαρμοστεί και για την παραγωγικότητα και το αντίκτυπο ενός επιστημονικού περιοδικού, καθώς και για μια ομάδα επιστημόνων, όπως ένα τμήμα ή ένα πανεπιστήμιο ή χώρα. Ο δείκτης προτάθηκε το 2005 από τον Jorge E. Hirsch, φυσικός στο UCSD, ως εργαλείο για τον προσδιορισμό της "σχετικής ποιότητας" των θεωρητικών φυσικών και καλείται μερικές φορές ως δείκτης Hirsch ή αριθμός Hirsch.

3.3.2.1 Υπολογισμός

Ο δείκτης βασίζεται στην κατανομή των αναφορών που έλαβε από τις δημοσιεύσεις ενός δεδομένου ερευνητή. Ο Hirsch γράφει: *Ένας ερευνητής έχει index h , αν h των εγγράφων του έχουν τουλάχιστον h αναφορές το καθένα και τα άλλα δεν έχουν πάνω από h αναφορές το καθένα.*

Με άλλα λόγια, ένας ερευνητής με δείκτη h έχει δημοσιεύσει h εργασίες όπου η κάθε μια έχει αναφερθεί και σε άλλες μελέτες, τουλάχιστον h φορές. Έτσι, ο δείκτης h αντανακλά τόσο τον αριθμό των δημοσιεύσεων όσο και τον αριθμό των αναφορών ανά δημοσίευση. Ο δείκτης σχεδιάστηκε για να βελτιώσει απλούστερους τρόπους μέτρησης όπως ο συνολικός αριθμός των αναφορών ή δημοσιεύσεων και λειτουργεί σωστά μόνο για τη σύγκριση επιστημόνων που εργάζονται στο ίδιο πεδίο γιατί τα πρότυπα που ακολουθούνται στις παραπομπές διαφέρουν σε μεγάλο βαθμό μεταξύ των διαφόρων πεδίων. Χρησιμοποιείται επίσης και ως εναλλακτική λύση για μέτρηση του αντίκτυπου μιας δημοσίευσης σε ένα περιοδικό. Επειδή μόνο τα πιο αναφερόμενα άρθρα σε αριθμό συμβάλλουν στο προσδιορισμό του h-index, η εύρεσή του είναι μια απλούστερη διαδικασία.



Εικόνα 3.1 h-index μειωμένων αναφορών από δημοσιεύσεις

Ο δείκτης μπορεί να καθορισθεί χειροκίνητα χρησιμοποιώντας βάσεις δεδομένων παραπομπών ή με αυτόματα εργαλεία. Subscription-based βάσεις δεδομένων, όπως το Scopus και το Web of Knowledge παρέχουν αυτοματοποιημένους υπολογισμούς. Το [Publish or Perish](#) πρόγραμμα του Harzing υπολογίζει το δείκτη με βάση τις εγγραφές ενός ερευνητή στο Google Scholar. Τον Ιούλιο του 2011 η Google δοκίμασε ένα εργαλείο που επιτρέπει στους μελετητές να παρακολουθούν τις δικές τους αναφορές και επίσης να παράγει ένα h-index και ένα i10-index (δείκτης που εμφανίζει τον αριθμό των δημοσιεύσεων που έχουν τουλάχιστον 10 αναφορές από άλλους). Επιπλέον, ειδικές βάσεις δεδομένων, όπως η INSPIRE-HEP μπορεί να υπολογίσει αυτόματα το h-index για τους ερευνητές που εργάζονται σε φυσική υψηλών ενεργειών.

Μια ιστοσελίδα υπολογισμού του h-index είναι από το πανεπιστήμιο του [waterloo](#).

3.3.2.2 Πλεονεκτήματα

Ο Hirsch προόριζε το h-index για να αντιμετωπιστούν τα κύρια μειονεκτήματα των άλλων βιβλιομετρικών δεικτών, όπως ο συνολικός αριθμός των εγγράφων ή του συνολικού

αριθμού των αναφορών. Ο συνολικός αριθμός των δημοσιεύσεων δεν αντιπροσωπεύει την ποιότητά τους, ενώ ο συνολικός αριθμός των παραπομπών μπορεί να επηρεαστεί δυσανάλογα από μια δημοσίευση σημαντικής επιρροής (για παράδειγμα, δημοσίευση προτείνοντας επιτυχημένες νέες τεχνικές, μέθοδοι ή προσεγγίσεις, που μπορεί να δημιουργήσουν μεγάλο αριθμό παραπομπών), ή έχει πολλές δημοσιεύσεις με λίγες αναφορές στο καθένα. Ο δείκτης h προορίζεται για την ταυτόχρονη μέτρηση της ποιότητας και της ποσότητας της επιστημονικού έργου.

Επίσης ο Hirsch έχει αποδείξει ότι ο δείκτης h έχει υψηλή προγνωστική αξία για το αν ένας επιστήμονας έχει κερδίσει τιμητικές διακρίσεις όπως η ένταξη στην Εθνική Ακαδημία ή βραβείο Νόμπελ. Ο δείκτης h μεγαλώνει όσο οι αναφορές συσσωρεύονται και έτσι εξαρτάται και από την "ακαδημαϊκή ηλικία" του ερευνητή.

3.3.2.3 Μειονεκτήματα

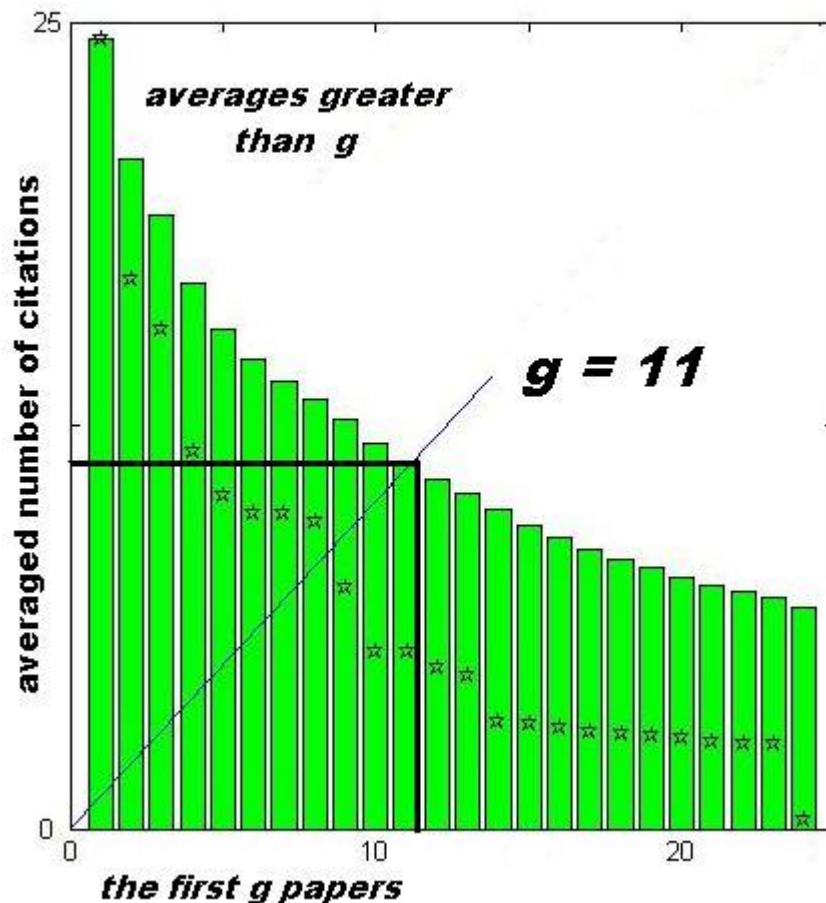
Μερικά από αυτά είναι:

- Ο δείκτης h δεν αντιπροσωπεύει τον αριθμό των συγγραφέων μιας έρευνας. Στο αρχική έκδοση του, ο Hirsch προτείνει να μοιραστούν οι αναφορές μεταξύ των συν-συγγραφέων. Παρόλαυτά, το h -index και παρόμοιοι δείκτες τείνουν να ευνοούν τους τομείς με μεγαλύτερες ομάδες, π.χ. πειραματικές έρευνες πάνω από θεωρητικές.
- Δεν αντιπροσωπεύει τον τυπικό αριθμό των παραπομπών σε διάφορους τομείς. Έχει αναφερθεί ότι τρόπος παραπομπών σε γενικές γραμμές επηρεάζεται από παράγοντες του κάθε κλάδου και πεδίου. Έτσι ενδέχεται να καταστήσει άκυρες συγκρίσεις όχι μόνο από διάφορους κλάδους, αλλά ακόμα και σε διαφορετικούς τομείς του ίδιου κλάδου.
- Περιορίζεται από το συνολικό αριθμό των δημοσιεύσεων. Αυτό σημαίνει ότι οι επιστήμονες με μια σύντομη καριέρα είναι σε εμφανές μειονέκτημα, ανεξάρτητα από τη σημασία των ανακαλύψεών τους. Για παράδειγμα αν ο Άλμπερτ Αϊνστάιν πέθαινε μετά τη δημοσίευση τέσσερις πρωτοποριακές Annus Mirabilis εργασίες του το 1905, το h -index του θα είχε κολλήσει στο 4 ή 5. Ωστόσο, όπως Hirsch αναφέρει, ο δείκτης προορίζεται ως ένα εργαλείο για την αξιολόγηση των ερευνητών στο ίδιο στάδιο της σταδιοδρομίας τους. Δεν προορίζεται ως ένα εργαλείο για ιστορικές συγκρίσεις.
- Δεν λαμβάνει υπόψη του το περιεχόμενο των παραπομπών. Για παράδειγμα κάποιες παραπομπές μπορεί να γίνονται και ως αρνητικά σχόλια για μια άλλη έρευνα, όχι μόνο ως θετικά.

3.3.3 g-index

Ο g-δείκτης είναι ένας δείκτης για την ποσοτικοποίηση της επιστημονικής παραγωγικότητας με βάση το αρχείο δημοσίευσης. Προτάθηκε το 2006 από τον Leo Egghe. Υπολογίζεται με βάση την κατανομή των αναφορών που λήφθηκαν από τις εκδόσεις ενός δεδομένου ερευνητή: Δεδομένου ενός σετ άρθρων που βαθμολογήθηκαν κατά φθίνουσα σειρά με βάση τις παραπομπές που έλαβαν, ο g-δείκτης είναι ο μοναδικός μεγαλύτερος αριθμός τέτοιος ώστε τα κορυφαία άρθρα g έχουν λάβει (από κοινού) τουλάχιστον G^2 αναφορές.

Ακριβώς όπως με τον h-index, ο G-δείκτης είναι ένας αριθμός που είναι το ίδιο για δύο διαφορετικές ποσότητες: g είναι ο αριθμός των εξαιρετικά αναφερόμενων άρθρων, έτσι ώστε το καθένα από αυτά έλαβε κατά μέσο όρο g αναφορές.



Εικόνα 3.2 Γραφική g-index. Τα δεδομένα αναφορών σχεδιασμένα με αστέρια επιτρέπουν την εξαγωγή του δείκτη h-index επίσης για σύγκριση.

Με άλλα λόγια, αυτό σημαίνει ότι προκειμένου να έχουμε g-index ενός συγγραφέα που παράγει n άρθρα θα πρέπει να έχει, κατά μέσο όρο n αναφορές για το καθένα από αυτά. Σε

αντίθεση με το h-index, ο G-INDEX εξαρτάται από την πλήρη καταμέτρηση παραπομπών σε έρευνες με πάρα πολλές αναφορές. Γενικά, h είναι ο αριθμός των δημοσιεύσεων ενός ελάχιστου ορίου ποιότητας που ανεβαίνει όσο ανεβαίνει και το h, ενώ το g επιτρέπει αναφορές από την υψηλότερα αναφερόμενες έρευνες να χρησιμοποιούνται για την ενίσχυση χαμηλότερων αναφερόμενων ερευνών για την επίτευξη αυτού του ελάχιστου ορίου. Ως εκ τούτου $g \geq h$.

Ωστόσο, σε αντίθεση με το h-index, ο G-index φθίνει όποτε ο μέσος αριθμός των παραπομπών για όλες τις δημοσιευμένες εργασίες υπερβαίνει τον συνολικό αριθμό των δημοσιευμένων εργασιών. Ο τρόπος με τον οποίο ορίζεται, ο G-index δεν έχει προσαρμοστεί στην παρούσα κατάσταση.

3.3.4 Google Scholar

Η πιο διαδεδομένη αναζήτηση άρθρων σήμερα μάλλον αποτελεί το [google scholar](#) και κατά συνέπεια δεν αποτελεί έκπληξη ότι χρησιμοποιείται και για την βαθμολόγησή των καθώς και την καταγραφή των αναφορών τους.

3.3.4.1 Google Scholar Citations

Το Google Scholar Citations παρέχει έναν απλό τρόπο για τους συγγραφείς να παρακολουθούν τις αναφορές στα άρθρα τους.

Πιο συγκεκριμένα μπορεί κάποιος:

- i) να ενημερωθεί αυτόματα ποιοι αναφέρουν τις δημοσιεύσεις του, να δει το γράφημα αναφορών του και να υπολογίζει αρκετούς δείκτες παραπομπών (όπως h-index).
- ii) να κάνει δημόσιο το προφίλ του, έτσι ώστε να μπορεί να εμφανίζεται στο Google Scholar όταν κάποιος ψάχνει για το όνομά του.
- iii) να προσθέσει ομάδες συναφών άρθρων, όχι μόνο ένα άρθρο τη φορά.
- iv) οι μετρήσεις των παραπομπών του να υπολογίζεται και ενημερώνεται αυτόματα καθώς το Google Scholar βρίσκει νέες αναφορές στο έργο του στο διαδίκτυο.
- v) να ενισχύσει την παγκόσμια προβολή και την προσβασιμότητα των δημοσιεύσεών του. Το Google Scholar συνεργάζεται με εκδότες περιοδικών από όλο τον κόσμο για να συμπεριλαμβάνει στα αποτελέσματα των αναζητήσεών του papers, διατριβές, προδημοσιεύσεις, περιλήψεις και τεχνικές εκθέσεις από όλους τους κλάδους της έρευνας.

3.3.4.2 Google Scholar Metrics

Το Google Scholar Metrics παρέχει έναν εύκολο τρόπο για τους συγγραφείς να μετρήσουν γρήγορα την προβολή και την επιρροή των πρόσφατων άρθρων τους σε επιστημονικές δημοσιεύσεις. Επίσης συνοψίζουν τις πρόσφατες αναφορές σε πολλές δημοσιεύσεις για να βοηθήσει τους συγγραφείς να αποφασίσουν που να δημοσιεύσουν μια νέα έρευνα τους. Κάποιος μπορεί να δει τις 100 πιο δημοφιλείς δημοσιεύσεις, να ταξινομήσει δημοσιεύσεις με βάση τους h-index και h-median δείκτες των τελευταίων 5 ετών ή να ψάξει άρθρα συγκεκριμένου κλάδου.

Διαθέσιμοι δείκτες που έχει το Google Scholar είναι οι h-index, h-core και h-median και h5-index, h5-core, h5-median αντίστοιχα (δηλαδή άρθρα δημοσιευμένα τα τελευταία 5 χρόνια).

Το Google Scholar Metrics καλύπτει σήμερα άρθρα που δημοσιεύτηκαν μεταξύ 2009 και 2013. Οι μετρήσεις αυτές βασίζονται σε αναφορές από άρθρα που καταγράφηκαν από το Google Scholar μέχρι τον Ιούνιο του 2014.

3.4 Βαθμολόγηση περιοδικών

Η κατάταξη περιοδικών χρησιμοποιείται ευρέως σε ακαδημαϊκούς κύκλους στην αξιολόγηση τους και την ποιότητα ενός ακαδημαϊκού περιοδικού. Η βαθμολόγηση προορίζεται να αντικατοπτρίσει τη θέση ενός περιοδικού στο πεδίο του, η σχετική δυσκολία για να δημοσιευτεί ένα άρθρο στο περιοδικό αυτό, και το κύρος που συνδέονται με αυτό. Έχει εισαχθεί πρόσφατα ως επίσημο εργαλείο αξιολόγησης έρευνας σε ορισμένες χώρες, όπως η Νορβηγία, η Αυστραλία και η Γαλλία.

Πιο γνωστές μέθοδοι υπολογισμού είναι:

- i) Impact Factor - Αντανακλώντας το μέσο αριθμό των παραπομπών σε άρθρα που δημοσιεύτηκαν στην επιστήμη και τις κοινωνικές επιστήμες περιοδικά.
- ii) Eigenfactor - Η αξιολόγηση της συνολικής σημασίας ενός επιστημονικού περιοδικού ανάλογα με τον αριθμό των εισερχόμενων αναφορών, με αναφορές από μεγαλύτερα περιοδικά να συμβάλουν περισσότερο στο δείκτη από μικρότερα.
- iii) SCImago Journal Rank (SJR) - Ένα μέτρο επιρροής επιστημονικών περιοδικών που λαμβάνει υπόψη τόσο τον αριθμό των αναφορών που λαμβάνονται από ένα περιοδικό αλλά και τη σημασία ή το κύρος των περιοδικών όπου προέρχονται οι αναφορές.
- iv) Δείκτης h (βλ. 3.3.2) - Συνήθως χρησιμοποιείται ως ένα μέτρο της επιστημονικής παραγωγικότητας επιστημονικής επίδρασης ενός μεμονωμένου επιστήμονα, αλλά μπορεί επίσης να χρησιμοποιηθεί και για να ταξινομήσει περιοδικά.

ν) Altmetrics - Βαθμολόγηση περιοδικών με βάση τις αναφορές αυτών στα κοινωνικά μέσα δικτύωσης.

3.4.1 Impact Factor (παράγοντας επιρροής)

Ο παράγοντας επιρροής (IF) του επιστημονικού περιοδικού είναι ένα μέτρο που αντικατοπτρίζει το μέσο αριθμό παραπομπών σε πρόσφατα άρθρα που δημοσιεύονται σε ένα περιοδικό. Συχνά χρησιμοποιείται ως προσεγγιστικός παράγοντας για τη σχετική σημασία ενός περιοδικού στο κλάδο του, όπου περιοδικά με υψηλότερο IF θεωρείται ότι είναι πιο σημαντικό από ό, τι περιοδικά με χαμηλότερο. Ο παράγοντας επιρροής επινοήθηκε από τον Eugene Garfield, ιδρυτή του Institute for Scientific Information. Οι παράγοντας αυτός υπολογίζεται ετησίως από το 1975 για τα περιοδικά που περιλαμβάνονται στο ευρετήριο του Journal Citation Reports.

3.4.1.1 Υπολογισμός

Σε οποιοδήποτε δεδομένο έτος, ο παράγοντας επιρροής ενός περιοδικού είναι ο μέσος αριθμός των αναφορών που λαμβάνει ανά paper που δημοσιεύεται στο περιοδικό αυτό κατά τη διάρκεια των δύο προηγούμενων ετών. Για παράδειγμα, εάν ένα περιοδικό έχει ένα συντελεστή επιπτώσεων 3 το 2008, τότε τα paper που δημοσιεύθηκαν το 2006 και το 2007 έλαβαν 3 αναφορές το καθένα κατά μέσο όρο το 2008. Ο παράγοντας επιρροής το 2008 του περιοδικού θα υπολογίζεται ως εξής:

2008 impact factor = A/B

όπου,

A = ο αριθμός των φορών όλων των στοιχείων που δημοσιεύτηκαν στο εν λόγω περιοδικό το 2006 και το 2007 που αναφέρθηκαν από δημοσιεύσεις κατά τη διάρκεια του 2008.

B = ο συνολικός αριθμός των "αναφερθέντων στοιχείων" που δημοσιεύτηκαν από το εν λόγω περιοδικό, το 2006 και το 2007. (αναφερθέντα στοιχεία για τον υπολογισμό αυτό είναι συνήθως άρθρα, reviews, πρακτικά ή σημειώσεις, όχι editorials ή γράμματα για τον συντάκτη).

Σημειώνεται ότι τα IF του 2008 δημοσιεύονται το 2009. Δεν μπορούν να υπολογιστούν έως ότου όλες οι δημοσιεύσεις του 2008 έχουν ληφθεί υπόψη.

Νέα επιστημονικά περιοδικά, τα οποία λαμβάνονται υπόψη από το πρώτο τεύχος τους, θα λάβουν έναν IF μετά από δύο χρόνια. Περιοδικά που δεν λαμβάνονται υπόψη από το πρώτο τεύχος τους δεν θα πάρουν έναν IF μέχρι να περάσουν τρία χρόνια. Ο παράγοντας επιρροής σχετίζεται με ένα συγκεκριμένο χρονικό διάστημα. Είναι δυνατόν να υπολογιστεί για οποιοδήποτε επιθυμητό χρονικό διάστημα όπως και το Journal Citation Reports (JCR) περιλαμβάνει επίσης έναν παράγοντα υπολογισμού πέντε ετών. Το JCR δείχνει κατατάξεις περιοδικών σύμφωνα με τον παράγοντα επιρροής και αν είναι επιθυμητό ανά κλάδο.

3.4.1.2 Μειονεκτήματα

Ένα μειονέκτημα είναι ότι δε μπορούν να συγκριθούν περιοδικά διαφορετικών κλάδων με αυτό το δείκτη γιατί κάθε κλάδος διαφέρει σε πολλούς τομείς αλλά κυρίως γιατί έχουν διαφορετικό ρυθμό δημοσιεύσεων και αναφορών. Π.χ. Στην επιστήμη των μαθηματικών και της φυσικής υπάρχει 1-3% ρυθμός αναφορών σε σχέση με 5-8% στην επιστήμη της βιολογίας όσο αναφορά μια καινούργια δημοσίευση στα πρώτα 2 χρόνια.

Ένα άλλο μειονέκτημα είναι ότι ένα περιοδικό μπορεί να υιοθετήσει δικές του πολιτικές σύνταξης για την αύξηση του IF του. Για παράδειγμα μπορεί να δημοσιεύει περισσότερα reviews που λαμβάνουν πιο πολλές αναφορές απ' ό,τι δημοσιεύματα που αφορούν μια έρευνα.

3.4.1.3 Αλγόριθμος PageRank

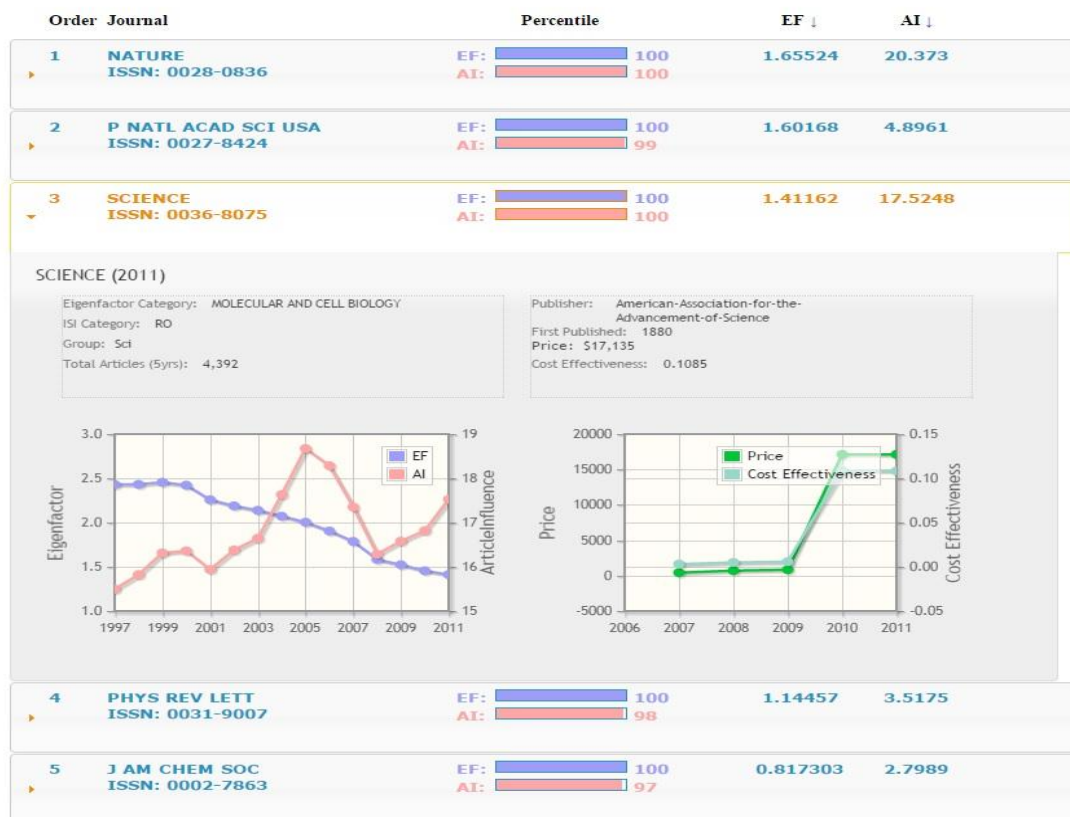
Το 1976 προτάθηκε ένας αναδρομικός παράγοντας επιρροής που δίνει μεγαλύτερο βάρος σε αναφορές από περιοδικά με υψηλό αντίκτυπο σε σχέση με αναφορές από περιοδικά με χαμηλότερο. Παρόμοιο αλγόριθμο χρησιμοποιεί και η Google (Google's PageRank) αν και ο αρχικός αλγόριθμος προβλέπει υψηλότερο βαθμό σε περιοδικά που αναφέρονται συχνά από άλλα, αλλά δεν αναφέρει το ίδιο άλλα περιοδικά. Αρκετοί ερευνητές έχουν προτείνει την πλήρη αντικατάσταση του IF από τον PageRank.

3.4.2 Eigenfactor

Το σκορ Eigenfactor, που αναπτύχθηκε από τους Jevin West και Carl Bergstrom στο Πανεπιστήμιο της Ουάσιγκτον, είναι η βαθμολογία της συνολικής σημασίας ενός επιστημονικού περιοδικού. Τα περιοδικά βαθμολογούνται ανάλογα με τον αριθμό των εισερχόμενων αναφορών, με αναφορές από υψηλότερης θέσης περιοδικά να συμβάλουν περισσότερο στο βαθμό από χαμηλότερης θέσης περιοδικά.

3.4.2.1 Υπολογισμός

Ο βαθμός Eigenfactor και ο βαθμός επιρροής ενός άρθρου υπολογίζονται στο <http://eigenfactor.org/>. Το σκορ Eigenfactor προορίζεται για τη μέτρηση της σημασίας ενός περιοδικού για την επιστημονική κοινότητα, εξετάζοντας την προέλευση των εισερχόμενων αναφορών, και θεωρείται ότι αντανακλά το πόσο συχνά ένας μέσος ερευνητής θα θέλει δει περιεχόμενο από αυτό το περιοδικό. Ωστόσο, η βαθμολογία Eigenfactor επηρεάζεται από το μέγεθος του περιοδικού, έτσι ώστε η βαθμολογία να διπλασιάζεται όταν το περιοδικό είναι διπλάσιο σε μέγεθος (αυτό μετριέται ως δημοσιευμένα άρθρα ανά έτος). Η βαθμολογία επιρροής ενός άρθρου μετρά τη μέση επίδραση των άρθρων στο περιοδικό.



Εικόνα 3.3 Υπολογισμός του Eigenfactor για το περιοδικό Science.

3.4.2.2 Πλεονεκτήματα

Ο Eigenfactor θεωρείται πιο εύρωστος αλγόριθμος συγκριτικά με τον impact factor γιατί λαμβάνει υπόψη του και την πηγή των αναφορών και όχι μόνο τον αριθμό τους και αν χρησιμοποιηθεί σε συνδυασμό με τον h-index μπορεί να αξιολογήσει το έργο ενός μεμονωμένου ερευνητή.

3.4.3 Ένδειξη SJR

Η ένδειξη SJR (SCImago Journal Rank) μετράει τόσο τον αριθμό των αναφορών που λαμβάνονται από ένα περιοδικό αλλά και τη σημασία ή το κύρος των περιοδικών όπου προέρχονται τέτοιες αναφορές. Ο δείκτης SJR είναι μια παραλλαγή του ιδιοδιανύσματος της κεντρικότητας που χρησιμοποιείται στη θεωρία των δικτύων. Η λογική είναι ότι η σημασία ενός κόμβου σε ένα δίκτυο βασίζεται στην αρχή ότι οι συνδέσεις με κόμβους υψηλής κίνησης συμβάλλουν περισσότερο στην αξία του κόμβου. Ο δείκτης SJR, ο οποίος είναι εμπνευσμένος από τον αλγόριθμο PageRank, έχει αναπτυχθεί για να χρησιμοποιείται σε εξαιρετικά μεγάλα και ετερογενή δίκτυα αναφορών από περιοδικά. Είναι μια ένδειξη η οποία είναι ανεξάρτητη του μεγέθους ενός άρθρου και καθορίζει περιοδικά με ένα "μέσο κύρος ανά άρθρο" και μπορεί να χρησιμοποιηθεί και για συγκρίσεις περιοδικών.

Ο δείκτης SJR είναι ένας μετρητής που μπορεί να υπολογιστεί δωρεάν και παρέχει μια εναλλακτική λύση αντί του παράγοντα επιπτώσεων (Impact Factor), ο οποίος βασίζεται σε στοιχεία από το Science Citation Index. Πάλι όμως υπολογίζεται με τον ίδιο τρόπο όπως και ο Impact Factor.

3.4.3.1 Υπολογισμός

Ο δείκτης SJR δίνει διαφορετικές τιμές σε αναφορές, ανάλογα με τη σημασία των περιοδικών που προέρχονται. Με αυτό τον τρόπο, αναφορές που έρχονται από πολύ σημαντικά περιοδικά θα είναι πιο πολύτιμες και ως εκ τούτου θα προσφέρουν μεγαλύτερο κύρος στα περιοδικά που τις δέχονται. Ο υπολογισμός του δείκτη SJR πραγματοποιείται χρησιμοποιώντας έναν επαναληπτικό αλγόριθμο που διανέμει αξίες "κύρους" μεταξύ των περιοδικών μέχρι να επιτευχθεί μια "σταθερή" κατάσταση λύσης. Αρχίζει θέτοντας το ίδιο βαθμό σε κάθε περιοδικό και στη συνέχεια, χρησιμοποιώντας μια επαναληπτική διαδικασία, αυτό το κύρος αναδιανέμεται μεταξύ περιοδικών μέσα από αναφορές μεταξύ τους.

Η διαδικασία αναπτύσσεται σε δύο φάσεις, (α) ο υπολογισμός του Prestige SJR (PSJR) για κάθε περιοδικό: ένα μέγεθος που αντικατοπτρίζει το σύνολο του κύρους του περιοδικού, και (β) ο δείκτης SJR δηλαδή η εξομάλυνση του μέτρου αυτού για να επιτευχθεί ένα μέγεθος ανεξάρτητο από το πόσο μεγάλο είναι το περιοδικό.

3.5 Altmetrics

Altmetrics είναι μη-παραδοσιακοί δείκτες που προτείνονται ως εναλλακτική λύση ως προς τις πιο παραδοσιακές μετρήσεις των επιπτώσεων των άρθρων και αναφορών, όπως ο impact factor και h-index. Ο όρος altmetrics προτάθηκε το 2010 και το μεγάλο του πλεονέκτημα είναι ότι μπορεί να εφαρμοστεί σε ανθρώπους, περιοδικά, βιβλία, σύνολα δεδομένων, παρουσιάσεις, βίντεο, source code repositories, ιστοσελίδες κ.α. Τα altmetrics καλύπτουν όχι μόνο τις παραπομπές, αλλά και άλλες πτυχές των επιπτώσεων ενός έργου, όπως πόσες βάσεις δεδομένων αναφέρονται σε αυτό, article views, downloads ή πόσες φορές αναφέρεται στα social media. Πολλές ιστοσελίδες υπολογίζουν altmetrics όπως ImpactStory (Wikipedia), Altmetric.com, Plum Analytic και CitedIn.

3.5.1 Κατηγορίες παραγόντων

Οι παράγοντες που λαμβάνονται υπόψη είναι οι εξής:

- i) **Viewed** - Παραδοσιακά, ένας συγγραφέας θα ήθελε να δημοσιεύσει σε ένα περιοδικό με ένα υψηλό ποσοστό συνδρομής ώστε πολλοί άνθρωποι να έχουν πρόσβαση στην έρευνά του. Με την εισαγωγή των τεχνολογιών του διαδικτύου κατέστη δυνατό να αναγνωστούν έρευνες δωρεάν οπότε αυτός ο παράγοντας μετράει πόσες φορές αναγνώστηκε μια σελίδα.
- ii) **Discussed** - Αυτός ο δείκτης μετράει πόσες φορές συζητήθηκε σε ένα blog/forum μια έρευνα όπως facebook, twitter, google+ κ.α.
- iii) **Saved** - Εδώ λαμβάνεται υπόψη πόσες φορές έγινε bookmarked μια ιστοσελίδα έρευνας ή πόσες φορές "κατεβάστηκε".
- iv) **Cited** - Εδώ δεν μετριέται μόνο πόσες φορές αναφέρθηκε ένα paper αλλά και ποια είναι η πηγή της αναφοράς, αν είναι για παράδειγμα το wikipedia ή κάποια άλλη λιγότερο σημαντική πηγή.

3.5.2 Μειονεκτήματα

Όπως όλοι οι δείκτες έτσι και τα altmetrics είναι ευάλωτα σε πλάγιες μεθόδους που σκοπό έχουν την προσωπική προβολή χωρίς απαραίτητα το έργο τους να έχει άμεση επίδραση στην έρευνα. Για παράδειγμα μπορούν να γίνουν σκόπιμα υπερβολικά πάρα πολλές αναφορές από τον ίδιο λογαριασμό σε μια έρευνα από ένα μέσο κοινωνικής δικτύωσης, πράξη που θα αυξήσει δραματικά το δείκτη.

Πανεπιστήμιο Πειραιώς

Κεφάλαιο 4^ο : DBLP (*Digital Bibliography & Library Project*)

4.1 Ανάγκη για μια συνολική βάση αναφοράς στην πληροφορική

Η χρηματοδότηση της επιστήμης χρειάζεται προσεκτικές αποφάσεις, αλλά αντιμετωπίζοντας κάποιος την αυξανόμενη πολυπλοκότητα και εξειδίκευση της έρευνας, πολλοί έχουν βρει αυτό το έργο δύσκολο. Κοιτάζοντας τις δημοσιεύσεις των ερευνητών και των ερευνητικών ομάδων από μια καθιερωμένη βάση αποτελεί πολύ βασικό για τον προσδιορισμό γόνιμης έρευνας. Ωστόσο, η φύση των δημοσιεύσεων στην πληροφορική είναι μοναδική και δεν μπορούν να εφαρμοστούν παραδοσιακές μεθόδους. Για την πληροφορική η βάση δεδομένων DBLP παρέχει μια δημοφιλή πύλη ανοικτού κώδικα για τη βιβλιογραφία. Σε ένα κοινό έργο, το Leibniz Κέντρο Πληροφορικής Schloss Dagstuhl - ένας παγκοσμίου φήμης τόπος συνάντησης για διεθνείς ερευνητικές ομάδες - και το DBLP έχουν ξεκινήσει μια συνεργασία για την ενίσχυση της αξίας του DBLP.

Ερευνητές της πληροφορικής προτιμούν να δημοσιεύουν τα αποτελέσματά τους σε συνέδρια, ενώ ερευνητές στα περισσότερα άλλα πεδία τείνουν να δημοσιεύουν σε επιστημονικά περιοδικά. Αυτός είναι ο λόγος για τον οποίο η βιβλιογραφία του DBLP κυρίως αναφέρει περιοδικά, αφήνοντας δημοσιεύσεις σε πρακτικά συνεδρίων να υποεκπροσωπούνται. Στη Γερμανία, μια πρόσφατη αξιολόγηση για την ποιότητα των υπηρεσιών της πληροφορικής κατέληξε στο συμπέρασμα ότι μια συνολική βάση αναφοράς ως προς την έρευνα λείπει.

4.2 Τι είναι το DBLP

Η βιβλιογραφία της επιστήμης των υπολογιστών του DBLP εξελίχθηκε από έναν πρώιμο μικρό πειραματικό Web server σε μια δημοφιλή υπηρεσία για την κοινότητα της επιστήμης των υπολογιστών. Αναπτύχθηκε κατά τα τελευταία 15 χρόνια από τον Δρ Michael Ley στο Πανεπιστήμιο του Trier και είναι διεθνώς σεβαστό από ερευνητές πληροφορικής για την ακρίβεια των δεδομένων του όπως επίσης μπορεί να χρησιμοποιηθεί χωρίς χρέωση.

Για τους ερευνητές της επιστήμης της πληροφορικής, το DBLP είναι ένα δημοφιλές εργαλείο για να αναζητήσουν τη δουλειά συναδέλφων τους και να εξάγουν βιβλιογραφικές λεπτομέρειες όταν συντάσσουν μια λίστα αναφοράς για μια νέα έρευνα. Βαθμολόγηση και δημιουργία προφίλ ενός ερευνητή, πανεπιστημίου, περιοδικών, συνεδρίων είναι μια άλλη πιθανή χρήση του DBLP.

Τα δεδομένα του DBLP μπορούν να ληφθούν σε ένα τεράστιο αρχείο μορφής xml από [εδώ](#).

4.3 Αρχιτεκτονική του DBLP

Παρακάτω ακολουθεί μια περιγραφή των δεδομένων που περιέχει το αρχείο dblp.xml. Να σημειωθεί ότι αυτό το αρχείο συνοδεύεται από το dblp.dtd που περιέχει τον ορισμό των τύπων δεδομένων και χρειάζεται για να μπορέσει κάποιος να κάνει parsing αυτού.

4.3.1 Άρθρα περιοδικών

Εδώ βλέπουμε πως περιγράφεται ένα άρθρο δημοσιευμένο σε ένα περιοδικό στο xml:

```
<article mdate="2009-01-26" key="journals/comgeo/HershbergerS08">
<author>John Hershberger</author>
<author>Subhash Suri</author>
<title>Adaptive sampling for geometric problems over data streams.</title>
<pages>191-208</pages>
<year>2008</year>
<volume>39</volume>
<journal>Comput. Geom.</journal>
<ee>http://dx.doi.org/10.1016/j.comgeo.2006.10.004</ee>
<url>db/journals/comgeo/comgeo39.html#HershbergerS08</url>
</article>
```

Το mdate δηλώνει την ημερομηνία της τελευταίας τροποποίησης της εγγραφής του άρθρου και το key το id του για το dblp. Τα author tags τα ονόματα που έκαναν την έρευνα για το άρθρο και το title τον τίτλο του. Το pages δηλώνει σε ποιές σελίδες στο τεύχος του περιοδικού είναι το άρθρο, το year σε ποια χρονιά δημοσιεύτηκε και το volume ποιο τεύχος είναι. Το journal το όνομα του περιοδικού και τέλος το ee (electronic edition) δηλώνει την ιστοσελίδα όπου μπορεί να βρει κάποιος το συγκεκριμένο άρθρο και το url είναι ο τοπικός σύνδεσμος ή την επέκταση που πρέπει να προσθέσει κάποιος στην αρχική σελίδα του dblp για να το βρει σε αυτή την σελίδα.

Π.χ. για το συγκεκριμένο παράδειγμα:

<http://www.informatik.unitrier.de/~ley/db/journals/comgeo/comgeo39.html#HershbergerS0>

4.3.2 Άρθρα συνεδρίων

Εδώ βλέπουμε πως περιγράφεται ένα άρθρο δημοσιευμένο σε ένα συνέδριο στο xml και χωρίζεται σε δυο μέρη: σε inproceedings και σε proceedings tags.

```
<inproceedings mdate="2011-06-23" key="conf/lcc/KhoussainovN94">
<author>Bakhadyr Khoussainov</author>
<author>Anil Nerode</author>
<title>Automatic Presentations of Structures.</title>
<pages>367-392</pages>
<year>1994</year>
<crossref>conf/lcc/1994</crossref>
<booktitle>LCC</booktitle>
<url>db/conf/lcc/lcc94.html#KhoussainovN94</url>
<ee>http://dx.doi.org/10.1007/3-540-60178-3_93</ee>
</inproceedings>
```

Το inproceeding tag περιέχει πληροφορίες για το δημοσιευμένο άρθρο σε ένα συνέδριο. Επειδή τα άρθρα που παρουσιάστηκαν σε συνέδρια δημοσιεύονται αρκετά αργότερα βλέπουμε ότι η ημερομηνία δημοσίευσής του είναι πιο μετά από την ημερομηνία διεξαγωγής του συνεδρίου. Το key λειτουργεί πάλι ως αναγνωριστικό. Τα author tags αναφέρουν τα ονόματα των ερευνητών του άρθρου και το title τον τίτλο του. Το pages ποιες σελίδες είναι, το year πότε εκδόθηκε το βιβλίο με τα πρακτικά του συνεδρίου και το crossref στο αναγνωριστικό της εγγραφής που αναφέρεται στο dblp σε ποιο συνέδριο ανήκει (βλ. proceedings). Το booktitle πως λέγεται το βιβλίο και πάλι το url είναι ο τοπικός σύνδεσμος και το ee ο παγκόσμιος σύνδεσμος.

```

<proceedings mdate="2003-01-13" key="conf/lcc/1994">
<editor>Daniel Leivant</editor>
<title>Logical and Computational Complexity. Selected Papers. Logic and Computational Complexity,
International Workshop LCC '94, Indianapolis, Indiana, USA, 13-16 October 1994</title>
<booktitle>LCC</booktitle>
<volume>960</volume>
<publisher>Springer</publisher>
<year>1995</year>
<isbn>3-540-60178-3</isbn>
<url>db/conf/lcc/lcc94.html</url>
</proceedings>

```

Το `proceedings` tag περιέχει πληροφορίες για το ίδιο το συνέδριο. Το `key` είναι το αναγνωριστικό του (εδώ όπου αναφέρεται και ένα `crossref` tag από ένα άρθρο), το `editor` αναφέρει τους κύριους διοργανωτές του συνεδρίου, το `title` τον τίτλο του, όπου εδώ αναφέρεται πέρα από το όνομα του, ο τόπος διεξαγωγής του και η διάρκειά του. Τα `booktitle`, `volume`, `publisher`, `year`, `isbn` αναφέρουν πάλι τον τίτλο του βιβλίου, το τεύχος του, την εταιρεία εκδόσεως, την χρονιά έκδοσης και τον μοναδικό αναγνωριστικό κωδικό του βιβλίου ISBN (International Serial Book Number). Το `url` πάλι είναι ο τοπικός σύνδεσμος.

4.3.3 Συγγραφείς - Ερευνητές

Εδώ βλέπουμε πως περιγράφεται ένας ερευνητής στο αρχείο xml:

```

<www mdate="2009-06-23" key="homepages/36/586">
<author>Thilo Streichert</author>
<title>Home Page</title>
<url>http://www12.informatik.uni-erlangen.de/people/streichert/</url>
</www>

```

Το `mdate` αναφέρει την ημερομηνία της τελευταίας τροποποίησης της εγγραφής του ερευνητή και το `key` το αναγνωριστικό `id` του. Το `author` το όνομά του και το `title` αναφέρει πάντα ότι έχει προσωπική ιστοσελίδα ο ερευνητής που ακολουθεί στο `url` tag αλλά πολλές φορές το `url` λείπει από πολλούς καθηγητές στο `xml`.

Πανεπιστήμιο Πειραιώς

Κεφάλαιο 5^ο : Αρχιτεκτονική Εργασίας

Αυτή η εργασία χωρίζεται κυρίως σε δυο μέρη:

- α) Συλλογή των δεδομένων
- β) Κατασκευή της ιστοσελίδας

5.1 Συλλογή των δεδομένων

Η συλλογή των δεδομένων ξεκινάει από το parsing του dblr.xml αρχείου γιατί όλες οι πληροφορίες που αναφέρθηκαν στο τρίτο κεφάλαιο αντλούνται από εκεί. Η ανάλυση της δομής του dblr αρχείου γίνεται στο κεφάλαιο 4. Ο Parser που χρησιμοποιείται είναι ο SAX και είναι γραμμένος σε PHP (βλ. κεφάλαιο 6). Ταυτόχρονα με το parsing γίνεται και το edit των δεδομένων όπου χρειάζεται για να εισαχθούν πιο εύκολα στη βάση ή να αφαιρεθεί πληροφορία που δεν χρειάζεται για την ιστοσελίδα και κατόπιν αποθηκεύονται σε τρία csv τύπου αρχεία. Ο λόγος που διαλέχτηκε αυτός ο τύπος αρχείου είναι γιατί είναι ένας από τους πιο "ελαφρύς" και κατα συνέπεια πιο γρήγορος τύπος αρχείου για εισαγωγή πολλών χιλιάδων εγγραφών μετά στη βάση δεδομένων. Η δομή της βάσης περιγράφεται αναλυτικά στο κεφάλαιο 7.

5.2 Κατασκευή της ιστοσελίδας

5.2.1 Αναζήτηση ερευνητών

Η κατασκευή της ιστοσελίδας ξεκινάει από τη δημιουργία ενός μηχανισμού αναζήτησης καθηγητών-ερευνητών ώστε να μπορεί ο χρήστης να αναζητεί όποιους καθηγητές επιθυμεί πριν τους αξιολογήσει (αναλυτικά βλ. κεφάλαιο 8).

Αυτό γίνεται με δυο τρόπους:

α) Απλή αναζήτηση

Με την απλή αναζήτηση μπορεί ο χρήστης να βρει έναν καθηγητή γράφοντας το όνομα του (ή μέρος αυτού) και να του εμφανίσει τα αντίστοιχα αποτελέσματα. Μετά θα μπορεί να επιλέγει πάνω στο όνομα του για να βλέπει το βιογραφικό των δημοσιεύσεών του και να τον αξιολογεί.

β) Σύνθετη αναζήτηση

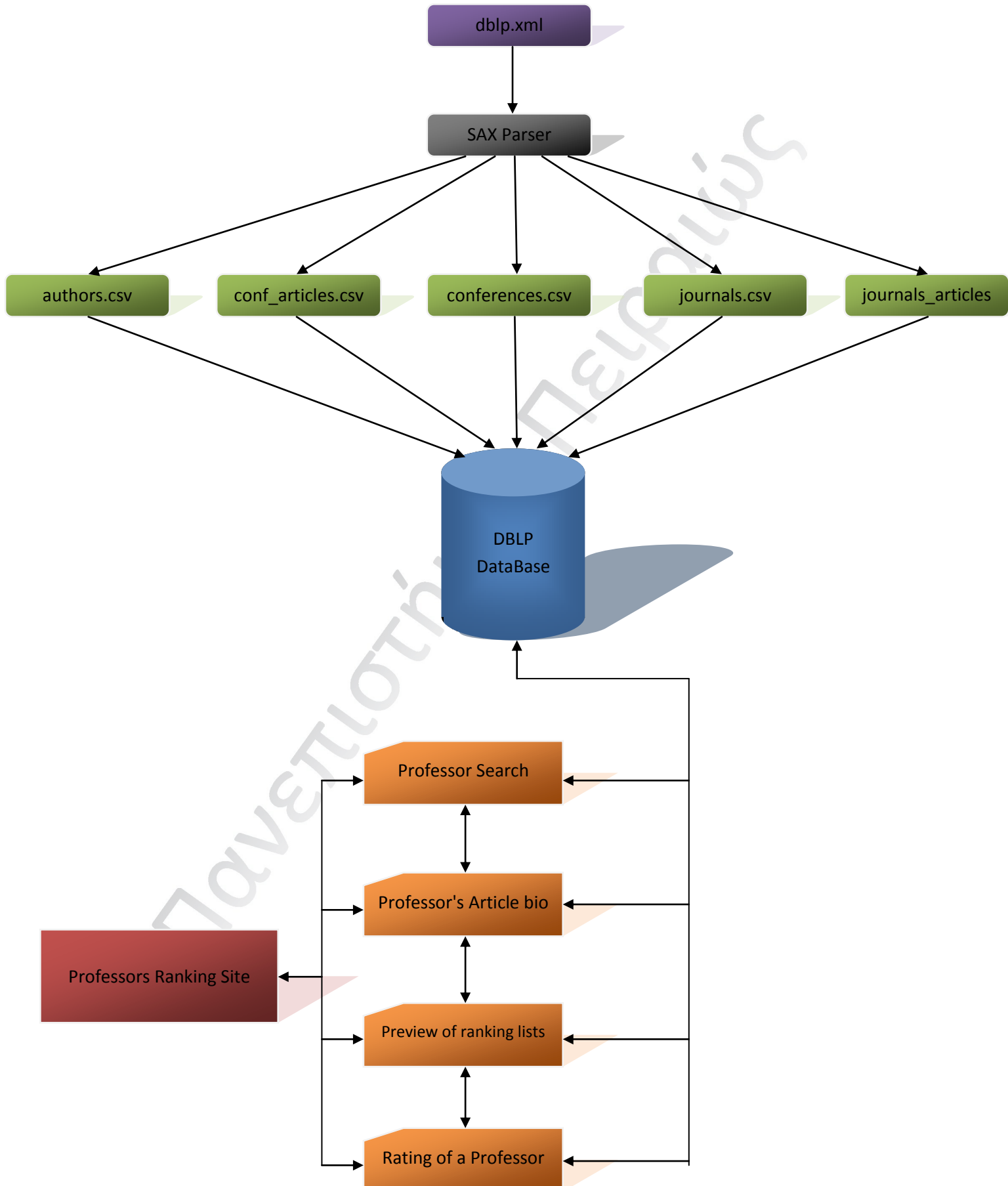
Με την σύνθετη αναζήτηση ο χρήστης μπορεί να αναζητήσει καθηγητές με 4 διαφορετικούς τρόπους:

- i) Με βάση τα περιοδικά που έχουν δημοσιεύσει άρθρα (μπορεί να επιλέξει μέχρι 3)
- ii) Με βάση τα συνέδρια που έχουν δημοσιεύσει άρθρα (μέχρι 3)
- iii) Με βάση την πρώτη χρονολογία δημοσίευσής άρθρου τους (είτε σε περιοδικό είτε σε συνέδριο)
- iv) Με βάση το πλήθος των δημοσιεύσεών τους (αρθροιστικά σε συνέδρια και περιοδικά)

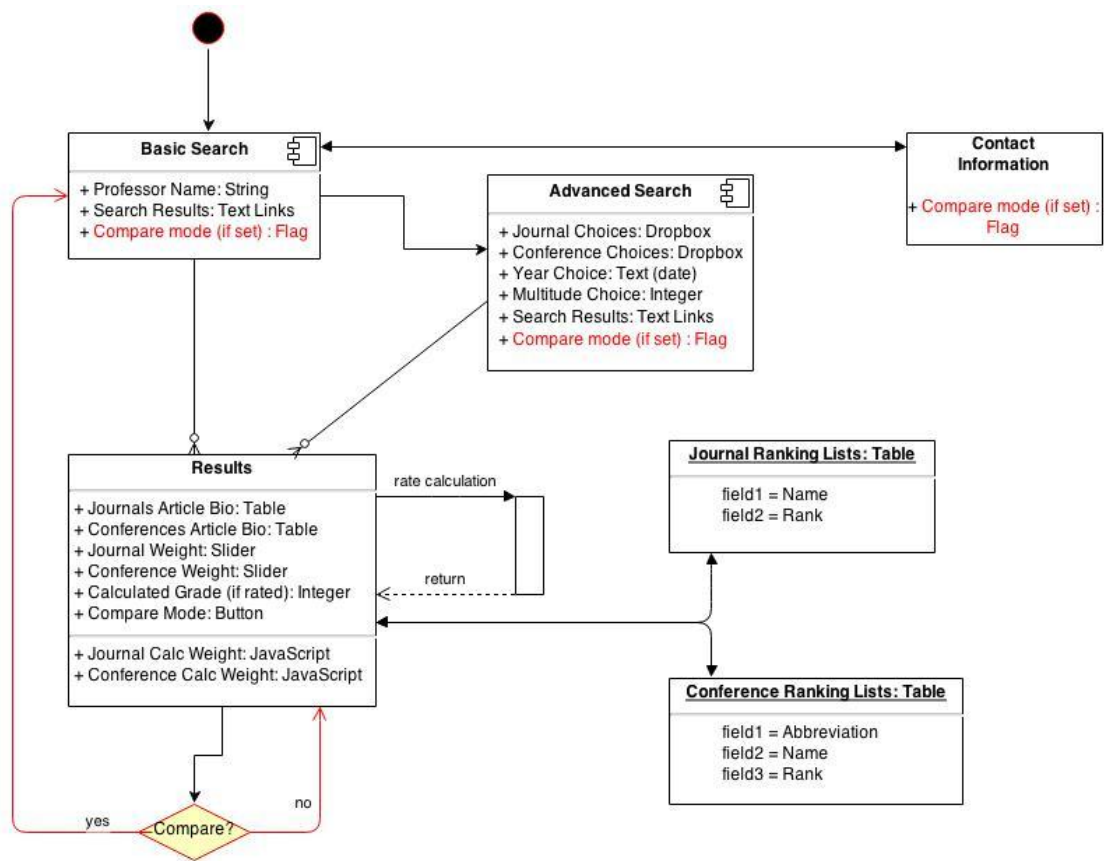
5.2.2 Αξιολόγηση ερευνητών

Ο στόχος της αξιολόγησης είναι να γίνεται όσο το δυνατόν πιο αντικειμενικά. Γι αυτό το σκοπό ο χρήστης έχει την επιλογή να διαλέγει μέσα από έτοιμες βαθμολογίες συνεδρίων/περιοδικών ποιές από αυτές θα λάβει υπόψη του πιο πολύ κατά την βαθμολόγηση ενός καθηγητή. Επίσης υπάρχει και η δυνατότητα σύγκριση δυο καθηγητών καθώς και να δει φυσικά τις προαναφερθείσες έτοιμες βαθμολογίες. Αναλυτικά η βαθμολόγηση επεξηγείται στο κεφάλαιο 9.

5.3 Διάγραμμα ροής



5.4 Site Diagram



Εικόνα 5.1: Site Diagram

Κεφάλαιο 6^ο : Χρησιμοποιούμενες Τεχνολογίες

Παρακάτω γίνεται μια περιγραφή των τεχνολογιών που χρησιμοποιήθηκαν για την κατασκευή του site και πως έγινε η συλλογή πληροφοριών. Οι τεχνολογίες είναι οι εξής: XAMPP, Apache server, MySQL, PHP, SAX Parser.

6.1 XAMPP

Το **XAMPP** είναι ένα πακέτο προγραμμάτων ελεύθερου λογισμικού, λογισμικού ανοικτού κώδικα και ανεξαρτήτου πλατφόρμας το οποίο περιέχει το εξυπηρετητή ιστοσελίδων http Apache, την βάση δεδομένων MySQL και ένα διερμηνέα για κώδικα γραμμένο σε γλώσσες προγραμματισμού PHP και Perl.

Το XAMPP είναι ακρωνύμιο και αναφέρεται στα παρακάτω αρχικά:

- **X** (αναφέρεται στο «cross-platform» που σημαίνει λογισμικό ανεξάρτητο πλατφόρμας, Windows στη δικιά μας περίπτωση)
- **A**pache HTTP εξυπηρετητής
- **M**ySQL
- **P**HP
- **P**erl (δεν χρησιμοποιήθηκε στη συγκεκριμένη εργασία)

Το XAMPP είναι ένα ελεύθερο λογισμικό το οποίο περιέχει ένα εξυπηρετητή ιστοσελίδων το οποίο μπορεί να εξυπηρετεί και δυναμικές ιστοσελίδες τεχνολογίας PHP/MySQL. Είναι ανεξάρτητο πλατφόρμας και τρέχει σε Microsoft Windows, Linux, Solaris, and Mac OS X και χρησιμοποιείται ως πλατφόρμα για την σχεδίαση και ανάπτυξη ιστοσελίδων με την τεχνολογίες όπως PHP, JSP και Servlets.

Το XAMPP προϋποθέτει μόνο τα λογισμικά συμπίεσης αρχείων zip, tar, 7z ή exe κατά την διάρκεια της εγκατάστασης. Το XAMPP έχει δυνατότητα αναβάθμισης σε νέες εκδόσεις του εξυπηρετητή ιστοσελίδων http Apache, της βάσης δεδομένων MySQL, της γλώσσας PHP και Perl. Το XAMPP συμπεριλαμβάνει επίσης τα πακέτα OpenSSL και το phpMyAdmin.

Επίσης οι σχεδιαστές του XAMPP προόριζαν το λογισμικό ως εργαλείο ανάπτυξης και δοκιμής ιστοσελίδων τοπικά στον υπολογιστή χωρίς να είναι απαραίτητη η σύνδεση στο διαδίκτυο. Για να είναι δυνατή η χρήση του, πολλές σημαντικές λειτουργίες ασφάλειας

έχουν απενεργοποιηθεί. Στην πράξη το XAMPP ορισμένες φορές χρησιμοποιείται και για την φιλοξενία ιστοσελίδων. Υπάρχει ειδικό εργαλείο το οποίο περιέχεται στο XAMPP για την προστασία με κωδικό των σημαντικών μερών. Το XAMPP υποστηρίζει την δημιουργία και διαχείριση βάσεων δεδομένων τύπου MySQL και SQLite.

Όταν το XAMPP εγκατασταθεί στον τοπικό υπολογιστή διαχειρίζεται τον localhost ως ένα απομακρυσμένο κόμβο, ο οποίος συνδέεται με το πρωτόκολλο μεταφοράς αρχείων FTP. Η σύνδεση στον localhost μέσω του FTP μπορεί να γίνει με το όνομα χρήστη «newuser» και το κωδικό «wampp». Για την βάση δεδομένων MySQL υπάρχει ο χρήστης «root» χωρίς κωδικό πρόσβασης.

6.2 Apache HTTP

Ο **Apache HTTP** γνωστός και απλά σαν Apache είναι ένας εξυπηρετητής του παγκόσμιου ιστού (web). Όποτε ένας χρήστης επισκέπτεται ένα ιστότοπο το πρόγραμμα πλοήγησης (browser) επικοινωνεί με έναν διακομιστή (server) μέσω του πρωτοκόλλου HTTP, ο οποίος παράγει τις ιστοσελίδες και τις αποστέλλει στο πρόγραμμα πλοήγησης. Ο Apache είναι ένας από τους δημοφιλέστερους εξυπηρετητές ιστού, εν μέρει γιατί λειτουργεί σε διάφορες πλατφόρμες όπως τα Windows, το Linux, το Unix και το Mac OS X. Κυκλοφόρησε υπό την άδεια λογισμικού Apache και είναι λογισμικό ανοιχτού κώδικα. Συντηρείται από μια κοινότητα ανοικτού κώδικα με επιτήρηση από το Ίδρυμα Λογισμικού Apache (Apache Software Foundation).

Ο Apache χρησιμοποιείται και σε τοπικά δίκτυα σαν διακομιστής συνεργαζόμενος με συστήματα διαχείρισης Βάσης Δεδομένων π.χ. Oracle, MySQL.

Η πρώτη του έκδοση, γνωστή ως NCSA HTTPd, δημιουργήθηκε από τον Robert McCool και κυκλοφόρησε το 1993. Θεωρείται ότι έπαιξε σημαντικό ρόλο στην αρχική επέκταση του παγκόσμιου ιστού. Ήταν η πρώτη βιώσιμη εναλλακτική επιλογή που παρουσιάστηκε απέναντι στον εξυπηρετητή http της εταιρίας Netscape και από τότε έχει εξελιχθεί στο σημείο να ανταγωνίζεται άλλους εξυπηρετητές βασισμένους στο Unix σε λειτουργικότητα και απόδοση. Από το 1996 ήταν από τους πιο δημοφιλείς όμως από τον Μάρτιο του 2006 έχει μειωθεί το ποσοστό της εγκατάστασής του κυρίως από τον Microsoft Internet Information Services και την πλατφόρμα .NET. Τον Οκτώβριο του 2007 το μερίδιο του ήταν 47.73% από όλους τους ιστοτόπους.

6.3 MySQL

Η **MySQL** είναι ένα σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων που μετρά περισσότερες από 11 εκατομμύρια εγκαταστάσεις. Έλαβε το όνομά της από την κόρη του Μόντυ Βιντένιους, τη Μάι (αγγλικά My). Το πρόγραμμα τρέχει έναν εξυπηρετητή (server) παρέχοντας πρόσβαση πολλών χρηστών σε ένα σύνολο βάσεων δεδομένων.

Ο κωδικός του εγχειρήματος είναι διαθέσιμος μέσω της GNU General Public License, καθώς και μέσω ορισμένων ιδιόκτητων συμφωνιών. Ανήκει και χρηματοδοτείται από μία και μοναδική κερδοσκοπική εταιρία, τη σουηδική MySQL AB, η οποία σήμερα ανήκει στην Oracle.

Η MySQL είναι δημοφιλής βάση δεδομένων για διαδικτυακά προγράμματα και ιστοσελίδες. Χρησιμοποιείται σε κάποιες από τις πιο διαδεδομένες διαδικτυακές υπηρεσίες, όπως το Flickr, το YouTube, η Wikipedia, το Google, το Facebook και το Twitter.

6.4 PHP

Η **PHP** είναι μια γλώσσα προγραμματισμού για τη δημιουργία σελίδων web με δυναμικό περιεχόμενο. Μια σελίδα PHP περνά από επεξεργασία από ένα συμβατό διακομιστή του Παγκόσμιου Ιστού (π.χ. Apache), ώστε να παραχθεί σε πραγματικό χρόνο το τελικό περιεχόμενο, που θα σταλεί στο πρόγραμμα περιήγησης των επισκεπτών σε μορφή κώδικα HTML.

Ένα αρχείο με κώδικα PHP θα πρέπει να έχει την κατάλληλη επέκταση(π.χ. .php, .php4, .phtml κ.ά.). Η ενσωμάτωση κώδικα σε ένα αρχείο επέκτασης .html δεν θα λειτουργήσει και θα εμφανίσει στον browser τον κώδικα χωρίς καμία επεξεργασία, εκτός αν έχει γίνει η κατάλληλη ρύθμιση στα MIME types του server. Επίσης ακόμη κι όταν ένα αρχείο έχει την επέκταση .php, θα πρέπει ο server να είναι ρυθμισμένος για να επεξεργάζεται και να μεταγλωττίζει τον κώδικα PHP σε HTML που καταλαβαίνει το πρόγραμμα πελάτη. Ο διακομιστής Apache, που χρησιμοποιείται σήμερα ευρέως σε συστήματα με τα λειτουργικά συστήματα GNU/Linux, Microsoft Windows, Mac OS X υποστηρίζει εξ ορισμού την εκτέλεση κώδικα PHP, είτε με την χρήση ενός πρόσθετου (mod_php) ή με την αποστολή του κώδικα προς εκτέλεση σε εξωτερική διεργασία CGI ή FCGI ή με την έλευση της php5.4 υποστηρίζονται η εκτέλεση σε πολυάσχολους ιστοχώρους με FastCGI Process Manager (FPM).

Ο συνδυασμός Linux/Apache/MySQL/PHP, που είναι η πιο δημοφιλής πλατφόρμα εκτέλεσης ιστοσελίδων είναι γνωστός και με το ακρωνύμιο LAMP. Παρόμοια, ο συνδυασμός */Apache/MySQL/PHP ονομάζεται *AMP, όπου το πρώτο αρχικό αντιστοιχεί στην πλατφόρμα, στην οποία εγκαθίστανται ο Apache, η MySQL και η PHP (π.χ. Windows, Mac OS X).

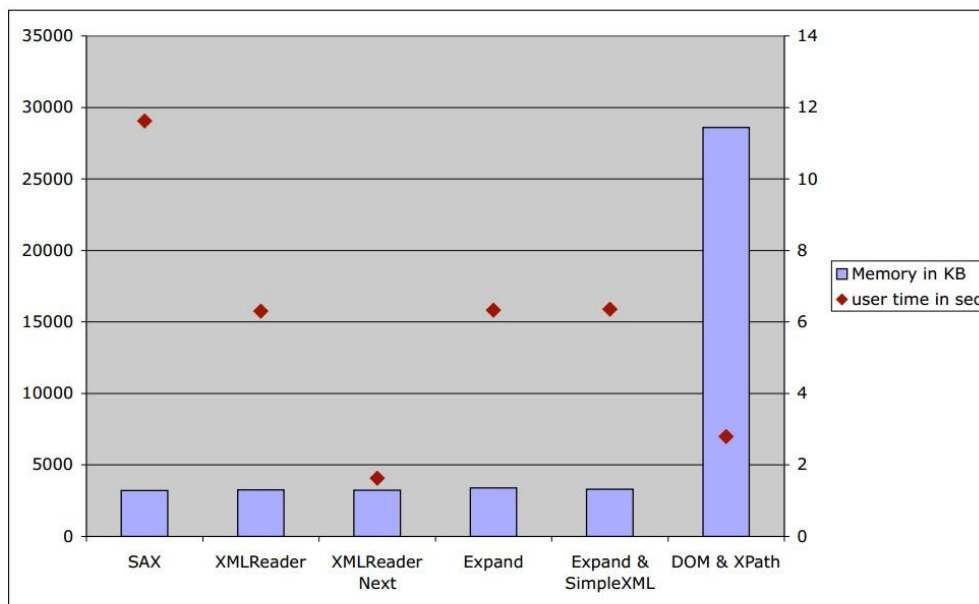
Ο LAMP συνήθως εγκαθίσταται και ρυθμίζεται στο Linux με τη βοήθεια του διαχειριστή πακέτων της εκάστοτε διανομής. Στην περίπτωση άλλων λειτουργικών συστημάτων, επειδή το κατέβασμα και η ρύθμιση των ξεχωριστών προγραμμάτων μπορεί να είναι πολύπλοκη, υπάρχουν έτοιμα πακέτα προς εγκατάσταση, όπως το XAMPP και το WAMP για τα Windows και το MAMP για το Mac OS X.

Η ιστορία της PHP ξεκινά από το 1994, όταν ένας φοιτητής, ο Rasmus Lerdorf δημιούργησε χρησιμοποιώντας τη γλώσσα προγραμματισμού Perl ένα απλό script με όνομα php.cgi, για προσωπική χρήση. Το script αυτό είχε σαν σκοπό να διατηρεί μια λίστα στατιστικών για τα άτομα που έβλεπαν το online βιογραφικό του σημείωμα. Αργότερα αυτό το script το διέθεσε και σε φίλους του, οι οποίοι άρχισαν να του ζητούν να προσθέσει περισσότερες δυνατότητες. Η γλώσσα τότε ονομαζόταν PHP/FI από τα αρχικά Personal Home Page/Form Interpreter. Το 1997 η PHP/FI έφθασε στην έκδοση 2.0, βασιζόμενη αυτή τη φορά στη γλώσσα C και αριθμώντας περισσότερους από 50.000 ιστότοπους που τη χρησιμοποιούσαν, ενώ αργότερα την ίδια χρονιά οι Andi Gutmans και Zeev Suraski ξαναέγραψαν τη γλώσσα από την αρχή, βασιζόμενοι όμως αρκετά στην PHP/FI 2.0. Έτσι η PHP έφθασε στην έκδοση 3.0 η οποία θύμιζε περισσότερο τη σημερινή μορφή της. Στη συνέχεια, οι Zeev και Andi δημιούργησαν την εταιρεία Zend (από τα αρχικά των ονομάτων τους), η οποία συνεχίζει μέχρι και σήμερα την ανάπτυξη και εξέλιξη της γλώσσας PHP. Ακολούθησε το 1998 η έκδοση 4 της PHP, τον Ιούλιο του 2004 διατέθηκε η έκδοση 5, ενώ αυτή τη στιγμή έχουν ήδη διατεθεί και οι πρώτες δοκιμαστικές εκδόσεις της επερχόμενης PHP 6, για οποιονδήποτε προγραμματιστή θέλει να τη χρησιμοποιήσει. Οι περισσότεροι ιστότοποι επί του παρόντος χρησιμοποιούν κυρίως τις εκδόσεις 4 και 5 της PHP.

6.5 SAX Parser

SAX (Simple API for XML) είναι ένας σειριακής πρόσβασης parser που αναπτύχθηκε από την XML-DEV για έγγραφα XML.

Ο SAX παρέχει ένα μηχανισμό για την ανάγνωση δεδομένων από ένα έγγραφο XML που είναι μια εναλλακτική λύση σε σχέση με το Document Object Model (DOM). Όταν το DOM λειτουργεί στο έγγραφο ως σύνολο, ο SAX λειτουργεί σε κάθε κομμάτι του XML διαδοχικά. Σε αντίθεση με το DOM, δεν υπάρχει καμία επίσημη προδιαγραφή για το SAX. Στο παρακάτω διάγραμμα βλέπουμε τις διαφορές μεταξύ των διαφορετικών σειριακών parsers (με εξαίρεση το DOM που είναι tree-based) όσο αναφορά την μνήμη και το χρόνο που χρειάζονται.



Εικόνα 6.1: Σύγκριση διαφορετικών parser

Ο SAX είναι event-driven API δηλαδή ο χρήστης ορίζει μια σειρά από μεθόδους επανάκλησης που θα κληθούν όταν συμβαίνουν συγκεκριμένα γεγονότα κατά τη διάρκεια του parsing. Κατά συνέπεια με αυτές τις μεθόδους ο χρήστης μπορεί να επιλέξει ότι πληροφορίες θέλει από το XML (Data Mining). Τα γεγονότα SAX περιλαμβάνουν (μεταξύ άλλων):

- Κόμβοι κειμένων
- Κόμβοι εκκίνησης/τέλους στοιχείων
- Κόμβοι οδηγιών επεξεργασίας
- Κόμβοι σχολίων

Ορισμένα γεγονότα αντιστοιχούν σε αντικείμενα XML που είναι εύκολο να επιστραφούν όλα με τη μία, όπως τα σχόλια. Ωστόσο, τα στοιχεία XML μπορεί να περιέχουν πολλά άλλα XML αντικείμενα, και έτσι ο SAX τα εκπροσωπεί όπως το κάνει η ίδια XML: με ένα γεγονός στην αρχή και ένα άλλο στο τέλος. Δηλαδή ο SAX δεν ασχολείται με τα στοιχεία, αλλά σε γεγονότα που αντιστοιχούν σε μεγάλο βαθμό σε κόμβους. Επίσης ο SAX είναι μονής κατεύθυνσης: προηγούμενα αναλυθέντα δεδομένα δεν μπορούν να ξαναδιαβαστούν χωρίς την έναρξη του parsing πάλι.

Το βασικό πλεονέκτημα του SAX είναι ότι δε χρειάζεται να αποθηκεύσει (άρα και να διαβάσει ολόκληρο πρώτα) το XML όπως κάνει το DOM που δεσμεύει μνήμη για

αποθήκευση του αρχείου σε μορφή δέντρου (tree form). Αυτό σε μεγάλα αρχεία όπως στη περίπτωση μας θα ήταν αρκετά χρονοβόρο ενώ ο SAX διαβάζει σειριακά δεσμεύοντας μνήμη μόνο για το πακέτο δεδομένων που διαβάζει κάθε φορά, κάτι το οποίο είναι και παραμετροποιήσιμο.

Παρακάτω ακολουθεί η βασική δομή του SAX Parser σε κώδικα PHP:

```
$parser = xml_parser_create();  
  
xml_set_element_handler($parser, 'start_element', 'end_element');  
  
xml_set_character_data_handler($parser, 'character_data');  
  
function start_element($parser, $element_name, $element_attrs) {  
  
    //do something when a desired tag starts  
  
}  
  
function character_data($parser, $data) {  
  
    //do something with the data inside a desired tag  
  
}  
  
function end_element($parser, $element_name) {  
  
    //do something when a desired tag closes  
  
}  
  
$fp = fopen('dblp.xml', 'r')  
    or die ("Cannot open dblp.xml!");  
  
while ($data = fread($fp, 4096)) {  
  
    xml_parse($parser, $data)  
  
    or die(sprintf('XML ERROR: %s at line %d',
```

```

xml_error_string(xml_get_error_code($parser)),

xml_get_current_line_number($parser));

}

```

```
xml_parser_free($parser);
```

Με την εντολή `xml_parser_create()` γίνεται η αρχικοποίηση ενός parser αντικειμένου.

Ο SAX λειτουργεί με χειριστές στοιχείων/δεδομένων για κάθε γεγονός. Χρειάζεται να οριστούν τρεις χειριστές(handlers):

- α) Τι κώδικας θα εκτελείται όταν ο parser θα συναντάει την αρχή ενός tag, δηλαδή το σύμβολο '<' .
- β) Τι κώδικας θα εκτελείται όταν ο parser θα συναντάει το περιεχόμενο ενός tag.
- γ) Τι κώδικας θα εκτελείται όταν ο parser θα συναντάει το τέλος ενός tag, δηλαδή το σύμβολο '>' .

Την λειτουργία α) και γ) την εκτελούν οι συναρτήσεις function `start_element` και function `end_element` ενώ την β) η συνάρτηση function `character_data`. Η αρχικοποίηση αυτών γίνεται με τις εντολές `xml_set_element_handler` και `xml_set_character_data_handler` αντίστοιχα. Όπως βλέπουμε δεν μπορούν να οριστούν με την ίδια εντολή γιατί είναι δυο διαφορετικοί χειριστές γιατί χειρίζονται διαφορετικού είδους δεδομένα, ο ένας στοιχεία(σύμβολα) και ο άλλος δεδομένα(data).

Ο χειριστής στοιχείων δέχεται ως ορίσματα τις συναρτήσεις που θέλουμε να δημιουργήσει για τον χειρισμό αυτών των στοιχείων. Πιο συγκεκριμένα τον parser που αρχικοποιήθηκε πριν, την `start_element` που θα καλείται κάθε φορά που θα βρσκει ο parser ένα tag, και την `end_element` που θα καλείται κάθε φορά που θα τελειώνει ένα tag.

Ο χειριστής δεδομένων δέχεται ως όρισμα τη συνάρτηση που θέλουμε να δημιουργήσει η οποία θα χειρίζεται το περιεχόμενο των tags, την `character_data` και φυσικά τον parser.

Η συνάρτηση `start_element` δέχεται ως όρισμα πέρα από τον parser το `element_name` που είναι το όνομα του εκάστοτε tag και τα `element_attrs` που είναι τα attributes που μπορεί να έχει ένα tag σε περίπτωση που θέλουμε να εκτελεστεί κώδικας με βάση κάποιο συγκεκριμένο attribute. Εδώ γράφουμε κώδικας που θέλουμε να εκτελεστεί κάθε φορά που ξεκινάει κάποιο tag.

Η συνάρτηση `character_data` μαζί με τον `parser` δέχεται ως όρισμα φυσικά τα δεδομένα που είναι το περιεχόμενο του κάθε `tag` για επεξεργασία και καλείται μετά την `start_element` μόνο αν το `tag` δεν ήταν χωρίς περιεχόμενο.

Η συνάρτηση `end_element` δέχεται και αυτή ως όρισμα τον `parser` και το όνομα του `tag`. Καλείται όταν ο `parser` συναντάει το τέλος ενός `tag`.

Στη συνέχεια η εντολή που ξεκινάει το parsing είναι η `xml_parse` που δέχεται ως ορίσματα τον `parser` και τα δεδομένα (`data`) που θέλουμε να κάνουμε parsing κάθε φορά από το αρχείο. Το parsing δηλαδή σπάει σε κομμάτια γιατί όπως αναφέρθηκε γίνεται σειριακά. Γι αυτό το λόγο αυτή η εντολή μπαίνει σε μια επανάληψη `while($data = fread($fp, 4096))` όπου εδώ ορίζεται πόσα δεδομένα θα διαβάζονται (και θα υπόκεινται σε parsing) κάθε φορά, στη προκειμένη περίπτωση 4096 bytes.

Τέλος, όταν το parsing τελειώσει ο `parser` ελευθερώνεται με την εντολή `xml_parser_free($parser)` για να μη δεσμεύει μνήμη.

Κεφάλαιο 7^ο : Δημιουργία Βάσης Δεδομένων

Με βάση τον SAX Parser που αναφέραμε προηγουμένως θα κάνουμε data mining για να πάρουμε τα επιθυμητά data από το dblp.xml και να δημιουργήσουμε τη database του site.

7.1 Δομή της βάσης

Η βάση δεδομένων θα περιέχει τους εξής πίνακες:

i) 1 Πίνακας με στοιχεία καθηγητών από DBLP

όπου περιέχει τα πεδία:

- author_id = το αναγνωριστικό του
- author_name = το όνομα του
- author_url = η ιστοσελίδα του

ii) 1 Πίνακας με στοιχεία δημοσιευμένων άρθρων σε περιοδικά από το DBLP

όπου περιέχει τα πεδία:

- J_A_id = το αναγνωριστικό του άρθρου
- J_A_authors = τα ονόματα των συγγραφέων του
- J_A_title = το όνομα του άρθρου
- J_A_pages = σε ποιες σελίδες βρίσκεται στο περιοδικό
- J_A_year = ποια χρονιά δημοσιεύτηκε το περιοδικό
- J_A_volume = σε πιο τεύχος του περιοδικού είναι δημοσιευμένο
- J_A_journals_name = το όνομα του περιοδικού

iii) 1 Πίνακας με στοιχεία δημοσιευμένων άρθρων σε συνέδρια από το DBLP

όπου περιέχει τα πεδία:

- C_A_id = το αναγνωριστικό του άρθρου
- C_A_authors = τα ονόματα των συγγραφέων του
- C_A_title = ο τίτλος του άρθρου
- C_A_pages = πόσες σελίδες είναι

- C_A_year = ποιά χρονιά δημοσιεύτηκε
- abbrev = η συντομογραφία του συνεδρίου που δημοσιεύτηκε
- Cross_ref = αναγνωριστικό πεδίο αναφοράς ως προς το id του συνεδρίου (πίνακας iv)

iv) 1 Πίνακας με τις πλήρεις πληροφορίες κάθε συνεδρίου από το DBLP

όπου περιέχει τα πεδία:

- Cross_ref = αναγνωριστικό id του συνεδρίου
- C_name = πλήρες όνομα του συνεδρίου (όνομα, τοποθεσία, χρονιά)

v) 1 Πίνακας με τις συντομογραφίες του κάθε συνεδρίου

όπου περιέχει το πεδίο:

- abbrev = η συντομογραφία του κάθε συνεδρίου

vi) 1 Πίνακας με στοιχεία περιοδικών από το DBLP

όπου περιέχει το πεδίο:

- J_name = το όνομα του κάθε περιοδικού

vii) 2 Πίνακες με στοιχεία περιοδικών από άλλες πηγές

όπου περιέχει τα πεδία:

- J_name = το όνομα του κάθε περιοδικού
- J_rank = η βαθμολογία του

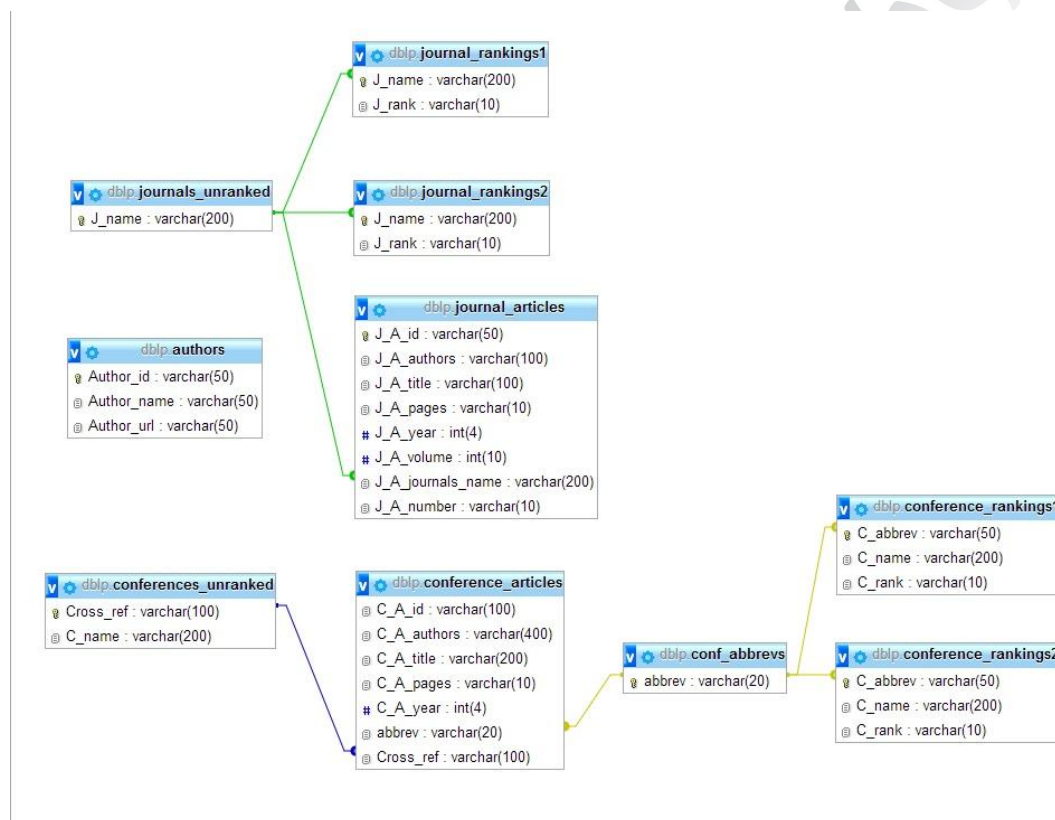
viii) 2 Πίνακες με στοιχεία συνεδρίων από άλλες πηγές

όπου περιέχει τα πεδία:

- C_abbrev = η συντομογραφία του κάθε συνεδρίου
- C_name = το όνομα του κάθε συνεδρίου
- C_rank = η βαθμολογία του

7.2 Database Schema

Εδώ παρατίθεται το Schema της βάσης.



Εικόνα 7.1: Database Schema

7.3 Έτοιμοι πίνακες βαθμολόγησης

Οι πίνακες vii και viii παρέχουν έτοιμη βαθμολόγηση συνεδρίων και περιοδικών με σκοπό την αναφορά σε αυτούς από τους χρήστες για την αξιολόγηση καθηγητών. Ο χρήστης μπορεί να τους δει και να επιλέξει ποιός θα ληφθεί πιο πολύ υπόψη. Για αρχή υπάρχουν 2 πίνακες συνεδρίων και 2 περιοδικών όπου είναι διαθέσιμοι από εδώ:

Για τα συνέδρια:

<https://lipn.univ-paris13.fr/~bennani/CSRank.html#A>

http://www-inf.it-sudparis.eu/~nguyen_n/conferences/conference_ranking

Για τα περιοδικά:

http://www-inf.it-sudparis.eu/~nguyen_n/conferences/journal_ranking

<https://people.creighton.edu/~pna06432/Journal.htm>

Πανεπιστήμιο Πειραιώς

Κεφάλαιο 8^ο : Αναζήτηση Καθηγητών

Για να γίνει αξιολόγηση ενός καθηγητή πρέπει να προφανώς ο χρήστης να διαλέξει πρώτα ποιον καθηγητή θέλει να βαθμολογήσει. Επίσης να σημειωθεί ότι σε οποιαδήποτε σελίδα κι αν βρίσκεται, μπορεί να επιστρέψει στην αρχική σελίδα πατώντας το home tab (που είναι η βασική αναζήτηση καθηγητών) και να δει πληροφορίες για το site στο contact tab.

8.1 Απλή αναζήτηση καθηγητών

Ο πρώτος τρόπος αναζήτησης γίνεται γράφοντας το όνομά του (ή μέρος του ονόματός του) στο textbox της μηχανής αναζήτησης και πατώντας search ή μπορεί να χρησιμοποιήσει την πιο λεπτομερή αναζήτηση πατώντας το σύνδεσμο *here*.

Τα αποτελέσματα επιστρέφονται πίσω ταξινομημένα αλφαβητικά και με δυνατότητα σελιδοποίησης αυτών σε περίπτωση που είναι πολλά. Επίσης επιστρέφονται στην ίδια σελίδα για να μπορεί ο χρήστης να κάνει καινούργια αναζήτηση χωρίς να αλλάξει σελίδα. Όταν βρει ποιον καθηγητή θέλει, απλά πατάει πάνω στο όνομα του και θα οδηγηθεί στην σελίδα προεπισκόπησης και βαθμολόγησης του καθηγητή γιατί το κάθε όνομα λειτουργεί και ως σύνδεσμος.

The screenshot shows the 'University of Piraeus' website interface. At the top, there are navigation tabs for 'Home' and 'Contact'. Below the header, a welcome message reads 'Welcome to our professor ranking site.' A search prompt asks the user to 'Please type the name of the professor you are looking for here:' followed by a text input field and a 'Search' button. A link for 'advanced professor search' is also present. Below the search section, a heading says 'Please click to rate a professor' followed by a list of professor names starting with 'A': A. Ali-Pacha, A. S. Salama, Albert Swart, Alejandro Rosete, Alex Hajnal, Alex See Kok Bin, Alexander N. Pak, Alexander Schwade, Alexandre Kruszewski, and Alexandros Bouganis. At the bottom, it indicates '1 of total 4 pages' and provides navigation links: 'First | Previous | Next | Last'.

Εικόνα 8.1: Αναζήτηση καθηγητών που περιέχουν τα γράμματα "a"

8.2 Σύνθετη αναζήτηση καθηγητών

Αυτό το είδος της αναζήτησης δε γίνεται με βάση το όνομα του καθηγητή αλλά με βάση τα εξής κριτήρια:

i) Με βάση τις δημοσιεύσεις σε συγκεκριμένα περιοδικά (μέχρι 3)

Ο χρήστης μπορεί να επιλέξει από αναδυόμενα κουτιά μέχρι 3 περιοδικά όπου γνωρίζει ότι έχει δημοσιεύσει ο καθηγητής που ψάχνει ή για να δει τους καθηγητές που πληρούν αυτό το κριτήριο. Για να προσθέσει ένα περιοδικό πρέπει να πατήσει το add journal button και αυτόματα θα προστεθεί ένα ακόμα κουτί επιλογών το οποίο δεν θα έχει τα προηγούμενα επιλεγμένα περιοδικά φυσικά. Για να αφαιρέσει ένα συνέδριο απλά πατάει το remove journal button.

Εικόνα 8.2: Αναζήτηση με βάση περιοδικά

ii) Με βάση τις δημοσιεύσεις σε συγκεκριμένα συνέδρια (μέχρι 3)

Πάλι όπως πριν, ο χρήστης μπορεί να επιλέξει από αναδυόμενα κουτιά μέχρι 3 συνέδρια όπου γνωρίζει ότι έχει δημοσιεύσει ο καθηγητής που ψάχνει ή για να δει τους καθηγητές που πληρούν αυτό το κριτήριο. Για να προσθέσει ένα συνέδριο πρέπει να πατήσει το add conference button και αυτόματα θα προστεθεί ένα ακόμα κουτί επιλογών το οποίο δεν θα έχει τα προηγούμενα επιλεγμένα συνέδρια φυσικά. Για να αφαιρέσει ένα συνέδριο απλά πατάει το remove conference button.

University of Piraeus Home Contact

Welcome to our professor ranking site.

Please choose what type of search interests you or you can go back to basic professor search

Choose on which journals the professor has published an article on (up to 3)

Please choose a journal Add Journal

Journal based Search

Choose on which conferences the professor has published an article on (up to 3)

ANSS Add Conference

ASST Add Conference Remove Conference

GEOINFORMATICS Remove Conference

Conference based Search

Choose from which year and onwards the articles will be included in the results (both journals/conferences)

Year based Search

Choose how many articles do you want the professor to have published overall at the minimum

Multitude based Search

Please click to rate a professor

Chatchai Khunboa
Robert Simon
Jonathan M. Huntley
David Tepper
David Wangerin
Chris DeCoro
Luis Miguel Campos
Hugo Coyote
Isaac D. Scherson
Dan Chen
Boon-Ping Gan
Nitupam Jukka
Stephen John Turner
Wentong Cai
Junhu Wei

12 of total 383 pages
First | Previous | Next | Last

Εικόνα 8.3: Αναζήτηση με βάση συνέδρια

iii) Με βάση τη χρονολογία

Με αυτόν τον τρόπο ο χρήστης μπορεί να επιλέξει να εμφανιστούν οι καθηγητές που έχουν δημοσιεύσει ένα άρθρο από μια συγκεκριμένη χρονολογία και μετά (είτε άρθρο συνεδρίου είτε περιοδικού).

Εικόνα 8.4: Αναζήτηση με βάση τη χρονολογία

iv) Με βάση το πλήθος δημοσιεύσεων

Με αυτόν τον τρόπο ο χρήστης μπορεί να επιλέξει τους καθηγητές που τον ενδιαφέρουν με βάση το ελάχιστο πλήθος των δημοσιεύσεών τους σε συνέδρια και περιοδικά (μαζί).

University of Piraeus Home Contact

Welcome to our professor ranking site.

Please choose what type of search interests you or you can go back to basic professor search

Choose on which journals the professor has published an article on (up to 3)

Please choose a journal

Choose on which conferences the professor has published an article on (up to 3)

Please choose a conference

Choose from which year and onwards the articles will be included in the results (both journals/conferences)

Choose how many articles do you want the professor to have published overall at the minimum

Please click to rate a professor

Adam Smyk
 Marek Tudruj
 Arend Rensink
 Bo Xing
 Wen-jing Gao
 Fuluwheo Vincent Nelwamondo
 Kimberly Battle
 Tshildzi Marwala
 Bruno Aiazzi
 Luciano Alparone
 Stefano Baronti

1 of total 14 pages
 First | Previous | Next | Last

Εικόνα 8.5: Με βάση πλήθος δημοσιεύσεων

Κεφάλαιο 9^ο : Αξιολόγηση Καθηγητών

9.1 Έτοιμες λίστες βαθμολόγησης

Όπως είναι λογικό, μια αξιολόγηση είναι πάντα υποκειμενική. Στη δικιά μας περίπτωση έχουμε τη δυνατότητα σύγκρισης μεταξύ δύο διαφορετικών έτοιμων λιστών με βαθμολογίες συνεδρίων και περιοδικών με τη χρήση ενός βαριδίου. Ο χρήστης θα μπορεί να μεταφέρει το δείκτη πιο κοντά στη λίστα που θέλει να λάβει πιο πολύ υπόψη ως προς τη βαθμολογία του καθηγητή. Έτσι η βαθμολόγηση θα γίνεται με βάση τους δύο δείκτες που όρισε ο χρήστης για τα συνέδρια/περιοδικά.

Πατώντας πάνω στα ονόματα των λιστών ο χρήστης μπορεί να τις δει πριν αποφασίσει ποιές θα λάβει υπόψη πιο πολύ.

University of Piraeus

Home Advanced Contact

Journal Ranking list A

Name	Rank
Journal of Visual Communication and Image Representation	A
Machine Learning	A
Multimedia Systems	A
Neural Computation	A
Operations Research	A
Performance Evaluation	A
Real-time Systems	A
SIAM Jnl on Computing	A
SIAM Jnl on Discrete Mathematics?	A
SIAM Jnl on Scientific and Statistical Computing	A
VLDB Intl Journal	A
VLSI Design	A
ACM Computing Surveys	B
ACM Trans on Mathematical Software	B
ACM Transactions on Asian Language Information Processing	B

4 of total 22 pages

First | Previous | Next | Last

You can return to the previous page here

Εικόνα 9.1: Λίστα βαθμολογίας περιοδικών

University of Piraeus Home Advanced Contact

Conference Ranking list B

Abbreviation	Name	Rank
WIDM	ACM Workshop on Web Information and Data Management	B
WIKISYM	International Symposiums on Wikis	B
WoLLIC	Workshop on Logic, Language, Information and Computation	B
WORDS	IEEE International Workshop on Object-Oriented Real-Time Dependable Systems	B
WSDM	ACM International Conference on Web Search and Data Mining	B
WWIC	International Conference on Wired / Wireless Internet Communications	B
XP	International Conference on extreme Programming and Flexible Processes in Software Engineering	B
AAA-IDEA	International Workshop on Advanced Architectures and Algorithms for Internet Delivery and Applications	C
AAAC	Symposium of Asian Association for Algorithms and Computation	C
AAIM	Conference on Algorithmic Aspects in Information and Management	C
ACAC	Athens Colloquium on Algorithms and Complexity	C
ACC	Applied Computing Conference	C
ACCMCC	Australasian Conference on Combinatorial Mathematics and Combinatorial Computing	C
ACG	Workshop on Algorithms, Combinatorics, and Geometry	C
ACHI	International Conference on Advances in Computer-Human Interactions	C

32 of total 76 pages
[First](#) | [Previous](#) | [Next](#) | [Last](#)

[You can return to the previous page here](#)

Εικόνα 9.2: Λίστα βαθμολογίας συνεδρίων

9.2 Βαθμολόγηση

Όταν ο χρήστης έχει διαλέξει το βαρίδιο στις λίστες δεν έχει παρά να πατήσει το κουμπί `rate this professor` και θα εμφανιστεί ο μέσος όρος βαθμολογίας που συγκεντρώνει ο καθηγητής όσο αναφορά τα περιοδικά, ο μέσος όρος των συνεδρίων και ο συνολικός μέσος όρος. Αν δεν υπάρχουν δημοσιεύσεις σε συνέδρια ή περιοδικά τότε αυτή η βαθμολογία δεν θα λαμβάνεται υπόψη στο συνολικό βαθμό. Επίσης υπάρχει η δυνατότητα να γίνει βαθμολόγηση λαμβάνοντας υπόψη άρθρα μόνο σε συνέδρια ή μόνο σε περιοδικά.

University of Piraeus Home Advanced Contact

William H. Sanders

Journals

Article	Pages	Year	Volume	Journal	Number
<i>This professor has no journal articles published</i>					

Conferences

Article	Pages	Year	Conference	Conference Details
A Connection Formalism for the Solution of Large and Stiff Models.	258-265	2001	ANSS	Proceedings 34th Annual Simulation Symposium (SS 2001), Seattle, WA, USA, 22-26 April 2001
Computation of the Asymptotic Bias and Variance for Simulation of Markov Reward Models.	173-182	1996	ANSS	Proceedings 29th Annual Simulation Symposium (SS '96), April 8-11, 1996, New Orleans, LA, USA
Passive Replication Schemes in Aqua.	125-130	2002	PRDC	9th Pacific Rim International Symposium on Dependable Computing (PRDC 2002), 16-18 December 2002, Tsukuba-City, Ibaraki, Japan
Specification-Based Intrusion Detection for Advanced Metering Infrastructures.	184-193	2011	PRDC	17th IEEE Pacific Rim International Symposium on Dependable Computing, PRDC 2011, Pasadena, CA, USA, December 12-14, 2011
Ferret: A Host Vulnerability Checking Tool.	389-394	2004	PRDC	10th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC 2004), 3-5 March 2004, Papeete, Tahiti
Formal Specification and Verification of a Group Membership Protocol for an Intrusion-Tolerant Group Communication System.	9-18	2002	PRDC	9th Pacific Rim International Symposium on Dependable Computing (PRDC 2002), 16-18 December 2002, Tsukuba-City, Ibaraki, Japan
Simultaneous Simulation of Alternative System Configurations.	41-48	2005	PRDC	11th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC 2005), 12-14 December, 2005, Changsha, Hunan, China
Detecting and Exploiting Symmetry in Discrete-state Markov Models.	26-38	2006	PRDC	12th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC 2006), 18-20 December, 2006, University of California, Riverside, USA

Εικόνα 9.3: Δημοσιεύσεις ενός καθηγητή

Do not include Journal rating Rate this Professor
 Do not include Conference rating

Journal ranking: 0.00 Conference ranking: 6.88

Total ranking: 6.88

Εικόνα 9.4: Βαθμολόγηση ενός καθηγητή (στην ίδια σελίδα)

Φυσικά στην ίδια σελίδα αξιολόγησης θα εμφανίζονται και όλες οι πληροφορίες για τα δημοσιευμένα άρθρα του καθηγητή καθώς επίσης και ο τίτλος της στήλης του κάθε πίνακα πληροφοριών μπορεί να λειτουργήσει ως κριτήριο ταξινόμησης αν πατηθεί.

Σημείωση: Υπάρχει η περίπτωση ένα περιοδικό ή συνέδριο να μην υπάρχει σε καμία από τις έτοιμες λίστες βαθμολόγησης. Σε αυτήν την περίπτωση ο βαθμός υπολογίζεται από τα ήδη υπάρχοντα περιοδικά/συνέδρια και αν δεν υπάρχει καμία αντιστοίχιση τότε προφανώς ο βαθμός θα είναι 0. Σε κάθε περίπτωση υπάρχει προειδοποιητικό μήνυμα που αναφέρει ότι δεν υπάρχει αντιστοίχιση σε συνέδρια ή περιοδικά. Αυτό το πρόβλημα θα εκλείψει με τον καιρό καθώς θα προστίθενται περισσότερες λίστες στην ιστοσελίδα καλύπτοντας όλο και περισσότερες περιπτώσεις (βλ. 9.4).

9.3 Σύγκριση

Ο χρήστης έχει τη δυνατότητα να συγκρίνει δύο καθηγητές ως προς την βαθμολογία τους. Πατώντας το κουμπί compare mode ο χρήστης μπαίνει σε λειτουργία σύγκρισης όπου οδηγείται πάλι στην αρχική σελίδα για να διαλέξει έναν δεύτερο καθηγητή για σύγκριση.

The screenshot shows the profile of Alexander Specker on the University of Piraeus website. The page is titled 'University of Piraeus' and has navigation links for 'Home', 'Advanced', and 'Contact'. The main content area is for Alexander Specker, with a sub-section for 'Journals' and a 'Conferences' section. The 'Journals' section indicates that the professor has no journal articles published. The 'Conferences' section lists one conference: 'An Agent-Based Telecooperation Framework' from 1998, published in COBUILD. Below the conference list, there are two horizontal sliders for ranking lists: 'Telecom SudParis Journals Ranking list' and 'Creighton University Journals Ranking list' (both at 50%), and 'CORE Conferences Ranking list' and 'Telecom SudParis Conferences Ranking list' (both at 50%). There are checkboxes for 'Do not include Journal rating' and 'Do not include Conference rating'. A 'Rate this Professor' button is visible. At the bottom, it shows 'Conference ranking: 6.25' and 'Total ranking: 6.25'. A 'Compare Mode' button is at the bottom left.

Εικόνα 9.5: Σύγκριση καθηγητών (α)

Όπως βλέπουμε ο χρήστης μπορεί να καταλάβει αν είναι σε λειτουργία σύγκρισης ανά πάσα στιγμή από τα κόκκινα γράμματα που το αναφέρουν, εμφανίζοντας και ποιος καθηγητής έχει οριστεί ως μέτρο σύγκρισης. Μπορεί να βγει από αυτήν την κατάσταση πατώντας απλά το κουμπί clear compare mode.



Εικόνα 9.6: Σύγκριση καθηγητών (β)

Σε αυτήν την εικόνα βλέπουμε ότι μπορεί να γίνει σύγκριση ως προς και τις τρεις βαθμολογίες: των άρθρων σε περιοδικά, σε συνέδρια και στον συνολικό βαθμό.

9.4 Προσθήκη νέων λιστών βαθμολόγησης

Όπως αναφέρθηκε προηγουμένως υπάρχει η πιθανότητα να μην γίνει αντιστοίχιση ενός συνεδρίου ή περιοδικού από ένα άρθρο ενός καθηγητή σε σχέση με τις λίστες βαθμολόγησης. Αυτό το πρόβλημα βελτιώνεται προσθέτοντας περισσότερες λίστες στη βάση δεδομένων και κατά συνέπεια στην ιστοσελίδα. Τα βήματα που πρέπει να ακολουθήσει κάποιος για να προσθέσει μια λίστα είναι τα εξής:

i) Export

Αφού έχει αποφασίσει ο διαχειριστής του συστήματος ποιά λίστα θα προσθέσει πρέπει να την εξαγάγει από την πηγή του σε μια μορφή αρχείου που να μπορεί να εισαχθεί αργότερα στη βάση δεδομένων (κατα προτίμηση .csv).

ii) Editing

Το επόμενο βήμα είναι να προσέξει τα πεδία της πληροφορίας να είναι ίδια με αυτά των προηγούμενων λιστών, δηλαδή να περιέχει μόνο το όνομα του περιοδικού και τη βαθμολογία του αν πρόκειται για περιοδικό και τη συντομογραφία του, το όνομά του και τη βαθμολογία του αν πρόκειται για συνέδριο.

Επίσης θα πρέπει να προσέξει η βαθμολόγηση να μετατραπεί σε A, B, C όπου:

A = 10

B = 10 > βαθμός >= 7,5

C = 7,5 > βαθμός >= 5

Η μετατροπή από γράμματα σε αριθμούς αργότερα για να υπολογιστεί ο βαθμός γίνεται από τον κώδικα αυτόματα.

iii) Import

Εισαγωγή του αρχείου στη βάση δεδομένων δημιουργώντας έναν καινούργιο πίνακα.

iv) Redesign

Επειδή μέχρι τώρα η ιστοσελίδα έχει μόνο 2 λίστες για επιλογή από τον χρήστη είτε σε συνέδρια ή περιοδικά αν προστεθούν παραπάνω, η λογική του βαριδίου (slider) δεν έχει νόημα. Για να μπορέσει ο χρήστης να επιλέξει με ποιο ποσοστό % θέλει να λάβει υπόψη του την κάθε λίστα ένας τρόπος είναι να το κάνει με τη χρήση αναδυόμενων κουτιών έτοιμων επιλογών (dropboxes). Δίπλα από κάθε λίστα να υπάρχει η επιλογή μέσω του dropdown να διαλέξει κατά πόσο ποσοστό (0% - 100%) θέλει να λάβει υπόψη του τη συγκεκριμένη λίστα. Φυσικά όταν διαλέγει ένα ποσοστό πρέπει μετά στις υπόλοιπες λίστες να μειώνεται ανάλογα και το ποσοστό τους διαθέσιμο προς επιλογή (π.χ. αν επιλεχτεί ποσοστό 50% στη λίστα A, το ποσοστό που θα μπορεί να επιλέξει ο χρήστης στις λίστες B και C να μην υπερβαίνει το υπόλοιπο 50%). Αυτές οι αλλαγές αφορούν το αρχείο rate.php.

v) Recalculate

Τέλος θα πρέπει να προσθέσει τις τιμές που θα πέρνουν οι καινούργιες λίστες στη συνάρτηση υπολογισμού του μέσου όρου των τριών βαθμολογιών (από άρθρα περιοδικών, συνεδρίων, συνολικός) στο αρχείο rate_calc.php.

Κεφάλαιο 10^ο : Επεκτασιμότητα - Επίλογος

10.1 Επεκτασιμότητα

Αυτή η ιστοσελίδα σχεδιάστηκε με σκοπό να είναι επεκτάσιμη. Πολλές ιδέες μπορούν να προστεθούν για να κάνουν την αξιολόγηση όσο πιο αντικειμενική γίνεται στο μέλλον.

Μία ιδέα είναι να αποθηκεύονται όλες οι αξιολογήσεις που γίνονται και κάθε φορά πριν αξιολογείται ένας καθηγητής να εμφανίζεται ο μέχρι τώρα μέσος όρος του με βάση τις προηγούμενες αξιολογήσεις του.

Μια ακόμα ιδέα είναι στην σύνθετη αναζήτηση καθηγητών να μπορεί να γίνει ακόμη πιο σύνθετη συνδυάζοντας και τα τέσσερα ήδη υπάρχον κριτήρια αντί για ξεχωριστή αναζήτηση που γίνεται αυτή τη στιγμή.

Επίσης, επειδή ο αριθμός των καθηγητών είναι υπερβολικά μεγάλος που επιστρέφεται πολλές φορές στην σύνθετη αναζήτηση, θα μπορούσε να προστεθεί μια αναζήτηση με βάση το όνομα του (ή μέρος αυτού) που θα αφορούσε μόνο τα επιστρεφόμενα αποτελέσματα για να μην χρειάζεται ο χρήστης να μετακινείται ανάμεσα σε άπειρες σελίδες εγγραφών.

10.2 Επίλογος

Ο στόχος της εργασίας αυτής ήταν η δημιουργία ενός site βαθμολόγησης ενεργών καθηγητών στην έρευνα γιατί υπάρχουν ελάχιστα site στο internet για αυτό το σκοπό. Ελπίζω η εκπόνηση αυτής της πτυχιακής να πέτυχε το στόχο της για μια όσο πιο γίνεται αντικειμενική και δίκαια βαθμολόγηση των καθηγητών/ερευνητών όπου το έργο τους είναι η κινητήρια δύναμη στην βελτίωση της ποιότητας ζωής της ανθρωπότητας πάντα μέσω της τεχνολογίας.

Βιβλιογραφία

<http://el.wikipedia.org/wiki/>

<http://dblp.uni-trier.de/xml/docu/dblp.xml.pdf>

<http://www.dblp.org/db/>

<http://www.harzing.com/pop.htm>

<http://subjectguides.uwaterloo.ca/content.php?pid=84805&sid=1885850>

<http://scholar.google.gr/>

<http://eigenfactor.org/>

http://blog.liip.ch/archive/2004/05/10/processing_large_xml_documents_with_php.html

Lists of Conferences used:

<https://lipn.univ-paris13.fr/~bennani/CSRank.html#A>

http://www-inf.it-sudparis.eu/~nquyen_n/conferences/conference_ranking

Lists of Journals used:

http://www-inf.it-sudparis.eu/~nquyen_n/conferences/journal_ranking

<https://people.creighton.edu/~pna06432/Journal.htm>

Programmes used:

[XAMPP](#)

[CSV Reader](#)

XML Editor: [Gvim](#)