



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΑΣ, ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ



Ερευνητική Ομάδα Γνώσης και Αβεβαιότητας

Πτυχιακή εργασία

Εκτίμηση και οπτικοποίηση της ερευνητικής περιοχής

Γαρυφαλλιά Δημητρίου
2024201100036

Επιβλέπων:

Μανόλης Γουάλλες
Λέκτορας

Τρίπολη, Ιούνιος 2015

Εγκρίθηκε από την εξεταστική επιτροπή την 10η Ιουνίου 2015.

Γεώργιος Λέπουρας Νικόλαος Τσελίκας
Αναπληρωτής καθηγητής Επίκουρος καθηγητής

.....

Γαρυφαλλιά Δημητρίου
Τμήμα Πληροφορικής και Τηλεπικοινωνιών
Πανεπιστήμιο Πελοποννήσου

Copyright © Γαρυφαλλιά Δημητρίου, 2015
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Πελοποννήσου.

Η παρούσα εργασία εκπονήθηκε σε συνεργασία με την Ερευνητική Ομάδα Γνώσης και Αβεβαιότητας (ΓΑΒ LAB)

Περίληψη

Οι δημοσιεύσεις είναι επιστημονικά άρθρα σε επιστημονικά περιοδικά και συνέδρια με σκοπό τη παρουσίαση μίας νέας καινοτομίας σε κάποιο επιστημονικό κλάδο. Κάθε μία από αυτές, και κατ' επέκταση και ο συγγραφέας της, επηρεάζει ανάλογα έναν ή και περισσότερους επιστημονικούς τομείς. Οι λίστες παραθέσεων και οι μετρικές που εφαρμόζονται σε αυτές αποτελούν μία σημαντική ένδειξη της επιστημονικής επίδρασης των ερευνητών. Το μέσο στο οποίο δημοσιεύεται ένα άρθρο δηλώνει και την αντίστοιχη επιστήμη που επηρεάζει. Αξιοποιώντας αυτή την πληροφορία επιτυγχάνουμε μία βαθύτερη κατανόηση της επιστημονικής επιρροής ενός ερευνητή. Στην παρούσα εργασία σχεδιάζουμε και αναπτύσσουμε λογισμικό που ενώνει και αποθηκεύει τα θέματα των γραφημάτων, τους συνδέσμους των περιοδικών μαζί με τα θέματα των γραφημάτων και τις επιρροές των ερευνητών, όλα μαζί σε μία ασαφή οντολογική δομή. Τα αποτελέσματα αποκτώνται στη μορφή των διανεμημένων μετρήσεων των επιπτώσεων, και έπειτα οπτικοποιούνται με έναν κατάλληλο και διαισθητικό τρόπο.

Abstract

References are scientific articles published in scientific journals and conferences with a view to presenting a new innovation in a scientific discipline. Each one of these papers, and by extension, the author of, affects one or more disciplines. The citation lists and metrics applied on them are considered a leading indication of a researchers scientific impact. The medium in which published states and the corresponding science influences. Utilizing this information to achieve a deeper understanding of a researcher's scientific impact. In this work we design and we develop software that connects and stores topics' graph, journal associations with the topics' graph and researchers' impacts, together in a fuzzy ontological structure. The results are obtained in the form of distributed impact metrics, which we then visualize in a suitable and intuitive manner.

*Αφιερώνεται στην οικογένειά μου
& στους φίλους που είναι πάντα δίπλα μου.*

Περιεχόμενα

Πρόλογος	ix
1 Εισαγωγή	1
2 Θεωρητικό υπόβαθρο	3
2.1 Web-based εργαλεία	3
2.2 Εισαγωγή στην οπτικοποίηση	7
2.3 Τεχνικές οπτικοποίησης	7
3 Ανάλυση αναφορών	13
4 Επιστήμες	17
4.1 Η ιεραρχία των επιστημών	23
4.2 Βαθμός συσχέτισης / Διεπιστήμες	28
5 Αξιοποίηση των δεδομένων	31
6 Οπτικοποίηση	35
6.1 Πλεονεκτήματα	39
7 Υλοποίηση	41
7.1 Αρχιτεκτονική	41
7.2 Λίστες	44
7.3 Αποτελέσματα	47
8 Παραδείγματα	49
9 Συμπεράσματα	57
Βιβλιογραφία	63

Πρόλογος

Ευχαριστίες

Πρώτα απ' όλα θα ήθελα να ευχαριστήσω τους γονείς μου που με στήριξαν στη μέχρι τώρα πορεία μου, καθώς και τον κ. Εμμανουήλ Γουάλλες για την εμπιστοσύνη, την κατανόηση και τη βοήθεια που μου προσέφερε κατά τη διάρκεια επίβλεψης της πτυχιακής αυτής. Επίσης, ένα μεγάλο ευχαριστώ στον συμφοιτητή μου Νικόλα Μπαμπέτα, για τις λίστες με τις δημοσιευμένες αναφορές που μου προσέφερε, οι οποίες αποτέλεσαν και τις κύριες εισόδους της εφαρμογής.

Κεφάλαιο 1

Εισαγωγή

Ο συσχετισμός των επιστημονικών εργασιών έχει χρησιμοποιηθεί για μεγάλο χρονικό διάστημα ως μέτρο της ερευνητικής επιρροής, που στηρίζεται στην παραδοχή ότι τα αναφερόμενα έργα αποτελούν την επιστημονική βάση πάνω στην οποία άλλα έργα έχουν χτιστεί. Ως εκ τούτου, κάθε μοναδική αναφορά προϋποθέτει την ύπαρξη της προσπάθειας ενός ακόμη, νέου ερευνητή που το αρχικό έργο είχε αντίκτυπο σε αυτή.

Τα αντεπιχειρήματα σε αυτή την προσέγγιση είναι πολλά, μερικά σαφώς είναι αρκετά έγκυρα. Για παράδειγμα, οι περισσότερες αναφορές βρίσκονται στην εισαγωγή ή σε συναφή τμήματα εργασίας, δείχνοντας πως τα αναφερόμενα έγγραφα είναι παρόμοια και όχι κατ' ανάγκη πως τα πορίσματά τους έχουν χρησιμοποιηθεί για την επίτευξη νέων αποτελεσμάτων. Εξάλλου, δεν είναι σπάνιο για ένα επιτυχημένο τίτλο να συμβάλει σημαντικά στον αριθμό των αναφορών που έχουν επιτευχθεί, κατά καιρούς ίσως περισσότερο από την ίδια τη ποιότητα του περιεχομένου, και πολλοί συγγραφείς έχουν παραδεχτεί ότι έχουν αναφέρει έργα που δεν έχουν καν διαβάσει, απλά και μόνο επειδή άλλοι τα έχουν αναφέρει σε κάποιο παρόμοιο πλαίσιο. Και φυσικά, υπάρχουν και εκείνες οι δημοσιεύσεις που απαριθμούν πολλές αναφορές, πολλές εκ των οποίων δεν έχουν πραγματικά ληφθεί υπόψιν. Ενώ αντίστοιχα υπάρχουν εκείνες που απαριθμούν λίγες αναφορές, μερικές φορές παραλείποντας έργα που έχουν ληφθεί υπόψιν έμμεσα ή ακόμη και άμεσα.

Ακόμα, μέχρι σήμερα, ο κατάλογος των παραπομπών θεωρείται ως κορυφαία ένδειξη της επιστημονικής επίπτωσης, και μετρικές όπως το πλήθος των αναφορών χρησιμοποιούνται ακόμη και στην αξιολόγηση των ερευνητών για τις επί σύμβαση διαλέξεις και τις τακτικές θέσεις. Οι προσπάθειες γίνονται, φυσικά, για τη βελτιστοποίηση της ποιότητας των πληροφοριών που προέρχονται από τη λίστα των αναφορών. Το φιλτράρισμα των αναφορών για την αφαίρεση των αναφορών στα τελευταία άρθρα των αναφερόμενων ερευνητών, καθώς και αναφορές από τα άρθρα συνεργαζόμενων συγγραφέων, είναι ένα πρώτο βήμα (αφαίρεση αυτοαναφορών). Στη συνέχεια, σε αυτή την ενισχυμένη/φιλτραρισμένη λίστα των αναφορών, διαφορετικές μετρικές μπορούν να εφαρμοστούν με στόχο την εξαγωγή μίας πιο δίκαιης και ακριβέστερης μέτρησης του αντικτύπου. Αυτές οι μετρικές, κάθε μία από τις οποίες έχουν τα πλεονεκτήματα και τις αδυναμίες τους, εξετάζουν αποκλειστικά και μόνο την ύπαρξη επιστημονικών αναφορών. Αυτό που αποτυγχάνουν να κάνουν είναι να εξετάσουν το περιεχόμενο της αναφοράς.

Σε αυτή την εργασία λαμβάνουμε υπόψιν το πλαίσιο των αναφορών με σκοπό να εξάγεται μία βαθύτερη κατανόηση της επιστημονικής επιρροής. Συγκεκριμένα, χρησιμοποιούμε το μέσο δημο-

σίευσης (συγκεκριμένο περιοδικό, συνέδριο κλπ) ως ένδειξη του επιστημονικού κλάδου στο οποίο ανήκει το κάθε δημοσιευμένο άρθρο. Επιπλέον, χρησιμοποιούμε αυτή την πληροφορία για να διαχωρίσουμε το χώρο των παραπομπών. Στη συνέχεια, είμαστε σε θέση να εξετάσουμε τον τρόπο με τον οποίο οι μετρήσεις των αναφορών διανέμονται στο χώρο των επιστημονικών κλάδων.

Προκειμένου να διατηρηθεί η δυνατότητα να αξιολογήσει κανείς την επιρροή ενός ερευνητή με μία μόνο ματιά, προτείνεται η κατάλληλη προσέγγιση απεικόνισης για τις “πλουσιότερες” μετρήσεις των επιπτώσεων που υπολογίσαμε. Τέλος, έχουμε περιγράψει πώς αυτό θα οδηγήσει τόσο σε μία ολοκληρωμένη πλατφόρμα λογισμικού απο μόνο του, όσο και σε μικρές εφαρμογές που μπορούν να προστεθούν σε εμπορικές ή προσωπικές ιστοσελίδες. Η εφαρμογή βασίζεται στη σχεδίαση και δημιουργία των κατάλληλων ασαφών οντολογιών που χρησιμοποιούνται για τη μοντελοποίηση των θεματικών ιεραρχιών, των θεμάτων των περιοδικών/συνεδρίων/βιβλίων και του πεδίου εφαρμογής της επιρροής των ερευνητών.

Σε αυτό το σημείο θα πρέπει να επισημάνουμε πως η προσέγγισή μας δεν αποτελεί μία εναλλακτική μέτρηση των παραπομπών. Αυτό που κάνουμε είναι να προσδιορίσουμε σημασιολογικά διαφορετικές διαμερίσεις του χώρου της παραπομπής, πάνω στα οποία οποιαδήποτε από τις συμβατικές μετρικές παραπομπών μπορεί να εφαρμοστεί. Για παράδειγμα, μπορεί κανείς να υπολογίσει το βαθμό επιρροής ενός ερευνητή ανά επιστημονικό πεδίο. Ως αποτέλεσμα μπορούμε να δούμε πάνω σε ποιο πεδίο κάθε ερευνητής έχει επιρροή, και ακόμη μπορούμε να κάνουμε διακρίσεις μεταξύ διαφορετικών τύπων επιρροών, όπως οι επικεντρωμένες (inter-discipline) ή οι ευρύτερες (intra-discipline).

Η δομή που θα ακολουθήσει η συγκεκριμένη πτυχιακή εργασία θα έχει ως ακολούθως. Στο κεφάλαιο 2 θα μιλήσουμε για το θεωρητικό υπόβαθρο που χρειάζεται κανείς για να κατανοήσει αυτό το έργο. Συγκεκριμένα, θα εστιάσουμε σε ορισμούς και συστήματα που υπάρχουν ήδη για την οπτικοποίηση δεδομένων και για την διευκόλυνση της ανάλυσης των αναφορών. Στο κεφάλαιο 3 θα αναφέρουμε τον τρόπο και τα κριτήρια που επιλέξαμε για να αναλύσουμε τις αναφορές. Στο κεφάλαιο 4 θα παρουσιάσουμε την ιεραρχία των επιστημών που επιλέξαμε, θα γνωστοποιήσουμε τα κριτήρια επιλογής της καθώς και το βαθμό συσχέτισης των διεπιστημονικών πεδίων με τις υπόλοιπες επιστήμες. Στο κεφάλαιο 5 θα εξηγήσουμε τον τρόπο με τον οποίο συνδέσαμε τα δεδομένα με το κομμάτι της οπτικοποίησης. Στο κεφάλαιο 6 θα παρουσιάσουμε τη νέα μέθοδο οπτικοποίησης που προτείνουμε καθώς και τα πλεονεκτήματα αυτής. Στο κεφάλαιο 7 θα εστιάσουμε στην αρχιτεκτονική της υλοποίησης, τις λίστες ως εισόδους και τα αποτελέσματα στα οποία καταλήξαμε. Στο κεφάλαιο 8 θα χρησιμοποιήσουμε μερικά ενδεικτικά παραδείγματα με σκοπό την καλύτερη κατανόηση μέσω της ανάδειξης κάποιων περιπτώσεων που παρουσιάζουν ιδιαίτερο ενδιαφέρον. Τέλος, στο κεφάλαιο 9 θα συνοψίσουμε το έργο αυτό και τα συμπεράσματα που μπορεί να εξάγει κανείς μελετώντας το.

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

Επιστημονικό άρθρο (Paper) είναι η παρουσίαση και η τεκμηρίωση μιας πρωτότυπης επιστημονικής συνεισφοράς. Τα επιστημονικά άρθρα παρουσιάζονται σε συνέδρια και δημοσιεύονται σε επιστημονικά περιοδικά. Σκοπός τους είναι η έγκυρη επιστημονική ενημέρωση, που επιτρέπει την ανταλλαγή πληροφοριών μεταξύ των ερευνητών, καθώς και η γνωστοποίηση της επιστημονικής εργασίας, ώστε να μην υπάρξουν επαναλήψεις παρά μόνο βελτιώσεις σε αυτή.

2.1 Web-based εργαλεία

Κατά καιρούς καθώς ο αριθμός των επιστημονικών άρθρων αυξανόταν δημιουργήθηκε η ανάγκη να ταξινομηθούν οι δημοσιεύσεις ανάλογα με το θέμα, τον συγγραφέα, με το συνέδριο/περιοδικό και γενικότερα το μέσο στο οποίο δημοσιεύτηκαν. Σκοπός αυτού ήταν η διευκόλυνση της αναζήτησης τους από άλλους ερευνητές που θα ήθελαν να τα χρησιμοποιήσουν ως σημείο αναφοράς για κάποιο νέο έργο.

Όμως τα μόνα εργαλεία που υπάρχουν προς το παρόν, τα οποία μπορούν να προσφέρουν σημαντικές πληροφορίες για τις δημοσιευμένες αναφορές, είναι βάσεις δεδομένων που κρατούν αρχεία καταγραφών πληροφοριών των αναφορών για κάθε δημοσιευμένη αναφορά. Αποτελούν επίσης εύχρηστες διαδικτυακές μηχανές αναζήτησης με εύκολη πρόσβαση από το ευρύ κοινό σε επιστημονικά άρθρα, διατριβές, βιβλία από διάφορους ακαδημαϊκούς εκδότες, επαγγελματικές ενώσεις, πανεπιστήμια και άλλους διαδικτυακούς τόπους. Οι πιο γνωστές αυτών είναι το Google scholar, το Research Gate, το Scopus, το Web of Science, το Academic search, το CiteSeerX κ.α.

Google scholar Το Google Scholar είναι μια υπηρεσία του Google, με ελεύθερη πρόσβαση σε βιβλιογραφικά στοιχεία και αναφορές άρθρων, βιβλίων και γενικότερα επιστημονικών τεκμηρίων στο σύνολο των επιστημονικών πεδίων. Το Google Scholar παρέχει ένα απλό τρόπο αναζήτησης στην ακαδημαϊκή βιβλιογραφία. Μπορεί να αναζητήσει κανείς σε πολλούς επιστημονικούς κλάδους και πηγές (άρθρα, διατριβές, βιβλία, περιλήψεις και αποφάσεις δικαστηρίων) και σε πολλούς ακαδημαϊκούς εκδότες, επαγγελματικές ενώσεις, ιδρυματικά και θεματικά αποθετήρια, πανεπιστήμια και άλλους δικτυακούς τόπους. Το Google Scholar βοηθά να αναζητήσει κανείς όλη την ακαδημαϊκή βιβλιογραφία από ένα μόνο σημείο, να εξερευνήσει συναφή έργα, αναφορές, συγγραφείς, και δημοσιεύσεις, να εντοπίσει το πλήρες κείμενο μιας εργασίας μέσω βιβλιοθηκών ή μέσω του παγκόσμιου ιστού, να ελέγξει ποιος αναφέρεται σε ποιές δημοσιεύσεις, να δημιουργήσει ένα δημόσιο προφίλ συγγραφέα, κλπ.

Σχήμα 2.1: Google Scholar



Research gate Η πύλη έρευνας ResearchGate είναι ένα ελεύθερο κοινωνικό δίκτυο καθώς και ένα εργαλείο συνεργασίας που απευθύνεται σε ερευνητές από όλες τις επιστήμες. Παρέχει διαδικτυακές εφαρμογές συμπεριλαμβανομένης της εννοιολογικής αναζήτησης (αναζήτηση περιλήψεων), από κοινού χρήση και δημοσίευση ηλεκτρονικών αρχείων – database, forums τόπους δημόσιων μεθοδολογικών συζητήσεων, ομάδων κλπ. Επίσης, τα μέλη του δικτύου μπορούν να δημιουργήσουν το δικό τους blog και τις δικές τους ανοιχτές ή ιδιωτικές ομάδες. Το ResearchGate έχει δημιουργήσει μια μηχανή αναζήτησης εννοιών τόσο από το εσωτερικό ηλεκτρονικό αρχείο όσο και από εξωτερικές πηγές αναζήτησης αρχείων όπως PubMed, CiteSeer, arXiv, και τη βιβλιοθήκη της NASA. Η μηχανή αναζήτησης δημιουργήθηκε με στόχο την ανάλυση μεγαλύτερου αριθμού εννοιών από αυτόν που παρατηρείται σε συμβατικές μηχανές αναζήτησης – μπορεί να αναλύσει κείμενα περιλήψεων – δίνοντας έτσι τη δυνατότητα για ακριβή αποτελέσματα.

Σχήμα 2.2: Research Gate



Scopus Το Scopus είναι η υπηρεσία αναζήτησης βιβλιογραφίας και αναφορών (citations) του Elsevier και παρέχεται μέσα από την πλατφόρμα SciVerse. Θεωρείται η μεγαλύτερη στον κόσμο βάση δεδομένων αναφορών και περιλήψεων από έγκριτη διεθνή βιβλιογραφία με έξυπνα εργαλεία που βοηθούν να αναζητήσει κανείς, ανακτήσει, αναλύσει και οπτικοποιήσει τμήματα της πληροφορίας που τον ενδιαφέρουν. Περιλαμβάνει πάνω από 20.500 τίτλους από 5.000 εκδότες σε όλο τον κόσμο, 49 εκατομμύρια άρθρα (το 78% με περιλήψεις), πάνω από 5,3 εκατομμύρια άρθρα συνεδρίων, και το 100% της κάλυψης Medline.

Σχήμα 2.3: Scopus



Web of Science® Το Web of Science® είναι μια ευρέως γνωστή υπηρεσία εντοπισμού βιβλιογραφικών αναφορών (citations), αλλά και αρθρογραφίας, και αποτελεί μέρος της πλατφόρμας Web of Knowledge της Thomson Reuters. Το Web of Science® παρέχει στους ερευνητές, καθηγητές, φοιτητές και διοικητικά μέλη γρήγορη και δυναμική πρόσβαση σε κορυφαίες βάσεις δεδομένων αναφορών του κόσμου. Πρόκειται για αξιόπιστη διεπιστημονική πληροφορία που καλύπτει πάνω από 12.000 από τα υψηλότερα παγκοσμίως σε αντίκτυπο περιοδικά, συμπεριλαμβανομένων περιοδικών Ανοικτής Πρόσβασης, και πάνω από 150.000 πρακτικά συνεδρίων. Το Web of Science® χωρίζεται σε 3 βάσεις δεδομένων που καλύπτουν περισσότερες από 250 ειδικότητες:

1. Science Citation Index Expanded, με στοιχεία από περίπου 6.500 περιοδικά, σε θέματα φυσικών & εφαρμοσμένων επιστημών και επιστημών υγείας με κάλυψη από το 1970 έως σήμερα
2. Social Sciences Citation Index, με στοιχεία από περίπου 1.950 περιοδικά, σε θέματα κοινωνικών επιστημών, από το 1970 έως σήμερα
3. Arts & Humanities Citation Index, με στοιχεία από περίπου 1.160 περιοδικά, σε θέματα ανθρωπιστικών επιστημών και τεχνών, από το 1975 έως σήμερα

Σχήμα 2.4: Web of Science®



Academic Search Η Academic Search είναι μια δωρεάν υπηρεσία που αναπτύχθηκε από τη Microsoft για να βοηθήσει ερευνητές, καθηγητές, φοιτητές και επαγγελματίες να βρουν εύκολα και γρήγορα ακαδημαϊκό περιεχόμενο. Η Academic Search ευρετηριάζει όχι μόνο εκατομμύρια επιστημονικές εργασίες, αλλά αναδεικνύει τις βασικές σχέσεις μεταξύ των εργασιών και μεταξύ των συγγραφέων με ευνόητο, οπτικό τρόπο που υπογραμμίζει τις κρίσιμες συνδέσεις που συμβάλουν στον καθορισμό της επιστημονικής έρευνας.

Σχήμα 2.5: Academic Search



CiteSeerX Το CiteSeer ήταν μία δημόσια μηχανή αναζήτησης και μία ψηφιακή βιβλιοθήκη για τα επιστημονικά και τα ακαδημαϊκά άρθρα, κυρίως στους τομείς των ηλεκτρονικών υπολογιστών και της επιστήμης της πληροφορίας, που πλέον έχει αντικατασταθεί από το CiteSeerX. Έγινε γνωστό το 1998 και έχει πολλά κοινά χαρακτηριστικά με τις ακαδημαϊκές μηχανές αναζήτησης. Συγκεκριμένα δημιουργείται αυτόνομα ένα ευρετήριο παραπομπών που μπορεί να χρησιμοποιηθεί

για αναζήτηση και αξιολόγηση. Επιπλέον, τα στατιστικά των παραπομπών και τα σχετικά έγγραφα έχουν υπολογιστεί για όλα τα άρθρα που αναφέρονται στη βάση δεδομένων και όχι μόνο για αυτά που έχουν καταχωρηθεί στο ευρετήριο. Οι συνδέσεις των αναφορών επιτρέπουν την περιήγηση στη βάση δεδομένων, χρησιμοποιώντας τις παραπομπές. Επιτρέπει σε έναν ερευνητή να δει εύκολα και γρήγορα, ελέγχοντας το πλαίσιο των παραπομπών μίας συγκεκριμένης αναφοράς, ποιοί άλλοι ερευνητές έχουν ασχοληθεί με αυτό το έργο. Τέλος, προτείνονται σχετικά έγγραφα που χρησιμοποιούν παραπομπές και μέτρα βασισμένα σε λέξεις, ενώ παράλληλα παρουσιάζεται μία συνεχώς ενημερωμένη βιβλιογραφία για κάθε έγγραφο. Στόχος της είναι να βελτιωθεί η διάδοση και η πρόσβαση της ακαδημαϊκής και επιστημονικής βιβλιογραφίας. Ως υπηρεσία μη κερδοσκοπικού χαρακτήρα που μπορεί να χρησιμοποιηθεί ελεύθερα από όλους, έχει θεωρηθεί ως μέρος του κινήματος για ανοιχτή πρόσβαση, που στοχεύει να αλλάξει την ακαδημαϊκή και επιστημονική προσέγγιση στη δημοσίευση για να επιτρέψει μεγαλύτερη πρόσβαση στην επιστημονική βιβλιογραφία.

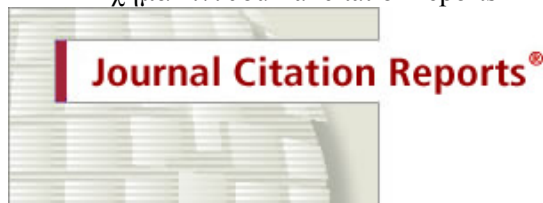
Σχήμα 2.6: CiteSeerX



Journal citation reports Εκτός από τις παραπάνω υπηρεσίες για την εύρεση αναφορών (citations) που αναφέρθηκαν, ιδιαίτερο ενδιαφέρον παρουσιάζει και η υπηρεσία Journal Citation Reports που σχετίζεται με τους δείκτες απήχησης των επιστημονικών περιοδικών (impact factors). Η υπηρεσία Journal Citation Reports είναι ένα μοναδικό εργαλείο που επιτρέπει στον χρήστη να αξιολογεί και να συγκρίνει πάνω από 7.500 διεθνή επιστημονικά περιοδικά από περισσότερους από 3.300 εκδότες με τη χρήση σχετικών δεδομένων αναφορών. Οι αναφορές (citations) άρθρων είναι σημαντικός δείκτης για την συχνότητα της χρήσης κάθε περιοδικού από τους ερευνητές. Τα στοιχεία αναφορών της υπηρεσίας JCR εκδίδονται σε ετήσια βάση και σε δύο εκδόσεις:

1. JCR Science Edition, με στοιχεία από περίπου 5.900 περιοδικά στις φυσικές και εφαρμοσμένες επιστήμες, στην τεχνολογία και τις επιστήμες υγείας.
2. JCR Social Sciences Edition, με στοιχεία από περίπου 1.700 περιοδικά στις κοινωνικές, ανθρωπιστικές και οικονομικές επιστήμες.

Σχήμα 2.7: Journal citation reports



Ιδιαίτερα χρήσιμο βέβαια, θα ήταν εάν υπήρχαν εργαλεία τα οποία θα μπορούσαν να βοηθήσουν στην εξόρυξη πληροφοριών από την ανάλυση κάποιας δημοσιευμένης αναφοράς με οποιοδήποτε

δυνατό τρόπο. Έτσι, στα πλαίσια της εργασίας αυτής θα αναπτύξουμε μία ολοκληρωμένη εφαρμογή που θα παίρνει ως είσοδο μία λίστα από δημοσιευμένες αναφορές (αρχείο xml) μαζί με τους συγγραφείς τους και το μέσο δημοσίευσής τους, και θα δίνει ως έξοδο ένα σύνολο γραφικών αναπαραστάσεων του βαθμού επιρροής τους, ανά επιστήμη για κάθε συγγραφέα.

2.2 Εισαγωγή στην οπτικοποίηση

Ως οπτικοποίηση δεδομένων (Data Visualization) ορίζεται η οπτική αναπαράσταση δεδομένων, τα οποία έχουν εξαχθεί είτε από κείμενο είτε από άλλου τύπου πηγές, με τη μορφή σχηματικών δομών και γραφημάτων. Σύμφωνα με τον Friendman (2008) «ο κύριος στόχος της οπτικοποίησης δεδομένων είναι να διαδώσει πληροφορία ξεκάθαρα και αποτελεσματικά με τη χρήση γραφικών μέσων».

2.3 Τεχνικές οπτικοποίησης

Υπάρχουν πολλές τεχνικές οπτικοποίησης και διάφοροι τρόποι που έχουν προταθεί κατά καιρούς από σημαντικές προσωπικότητες στο χώρο των επιστημών. Μία από τις πιο γνωστές που επιλέξαμε εμείς να παρουσιάσουμε βασίζεται στην ταξινόμηση των Ankerst και Keim.

Η Οπτική Εξόρυξη Πληροφοριών (Visual Data Mining) αποτελεί ένα ταχέως αναπτυσσόμενο πεδίο στην εξερευνητική ανάλυση και στο Data Mining. Βασίζεται στην συνένωση εννοιών από διάφορα επιστημονικά πεδία όπως μέθοδοι οπτικής παρουσίασης, οπτικοποίηση της πληροφορίας (Information Visualization), της επιστήμης της αντίληψης (Cognitive Science), της μηχανικής μάθησης κ.α..

Αντικειμενικός στόχος του Visual Data Mining είναι η αποτελεσματική παρουσίαση των αποτελεσμάτων έτσι ώστε ο αναλυτής να επωφελείται από την οπτική αναπαράσταση και να διακρίνει την πολύτιμη πληροφορία που ενδεχομένως να παραβλέπουν οι αυτοματοποιημένες μέθοδοι.

Σύμφωνα με ορισμό του Ankerst το Visual Data Mining αποτελεί ένα βήμα στην διαδικασία Knowledge Discovery in Databases που χρησιμοποιεί την οπτικοποίηση σαν δίαυλο επικοινωνίας μεταξύ του υπολογιστή και του αναλυτή (χρήστη) για την εξαγωγή νέων προτύπων που μπορούν να ερμηνευτούν καλύτερα.

Ο Ankerst πρότεινε μια ταξινόμηση των προσεγγίσεων του Visual Data Mining βασιζόμενος στην ισορροπία και στην ακολουθία μεταξύ του αυτόματου και του αλληλεπιδραστικού μέρους στη διαδικασία Data Mining. Έτσι κατέληξε στις ακόλουθες 3 προσεγγίσεις:

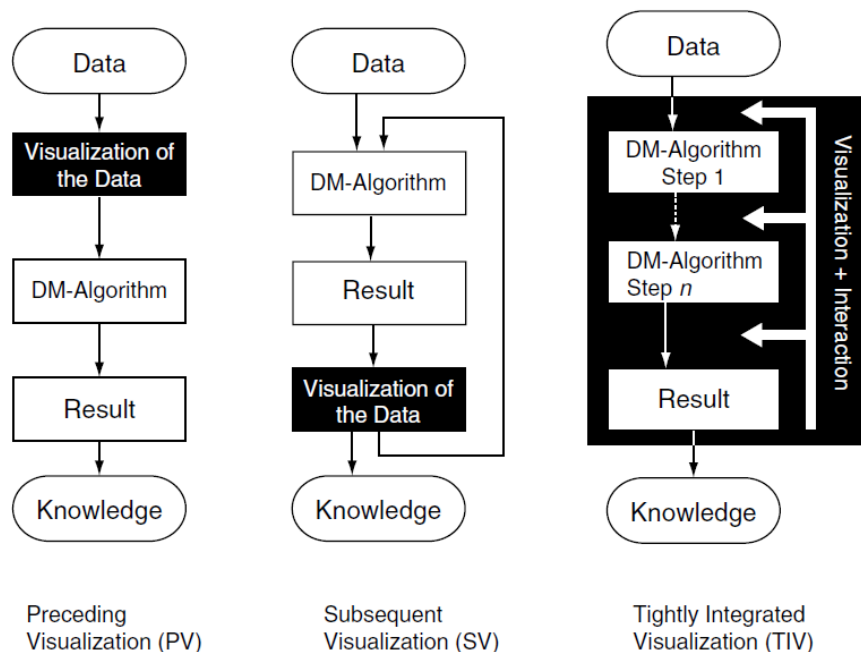
1. Οπτικοποίηση των δεδομένων
2. Οπτικοποίηση του αποτελέσματος [των κανόνων]
3. Οπτικοποίηση των ενδιάμεσων βημάτων

Στην προσέγγιση αυτή, η χρήση τεχνικών οπτικοποίησης γίνεται στα αρχικά δεδομένα και έχει σαν στόχο την μείωση των δεδομένων βάσει της ανθρώπινης αντίληψης. Η οπτικοποίηση αυτή παρέχει στον αναλυτή τη δυνατότητα να ελέγχει την αναζήτηση από την αρχή της διαδικασίας καθώς και την δυνατότητα της ανατροφοδότησης.

Στην προσέγγιση αυτή υλοποιείται ο αλγόριθμος και εξάγονται αυτόματα τα πρότυπα. Οι τεχνικές οπτικής παρουσίασης χρησιμοποιούνται για την παρουσίαση των αποτελεσμάτων στον αναλυτή. Σαν στόχο έχει να υποβοηθήσει τον αναλυτή στην ερμηνεία των αποτελεσμάτων και όταν αυτά δεν κρίνονται ικανοποιητικά να τον βοηθήσει στον καθορισμό διαφορετικών παραμέτρων στον αλγόριθμο.

Αφορά στην ταυτόχρονη εφαρμογή της διαδικασίας Data Mining με οπτικοποίηση των ενδιάμεσων βημάτων για καλύτερη παρακολούθηση και έλεγχο της διαδικασίας. Ένα ενδιάμεσο αποτέλεσμα του Data Mining αλγόριθμου παρουσιάζεται οπτικά και αν το αποτέλεσμα του δεν είναι ικανοποιητικό υλοποιείται ξανά ο αλγόριθμος με βάση τα νέα στοιχεία που έχει στην διάθεση του ο αναλυτής. Βασικός στόχος της προσέγγισης είναι η ανεξαρτητοποίηση της υλοποίησης του αλγορίθμου από την εκάστοτε εφαρμογή. Ένας αυτοματοποιημένος αλγόριθμος μπορεί να είναι αρκετά αποδοτικός σε μια συγκεκριμένη εφαρμογή αλλά σε κάποια άλλη να παρουσιάζει σημαντικές αδυναμίες. Καθώς δεν υπάρχει αλγόριθμος κατάλληλος για όλες τις εφαρμογές, τα αποτελέσματα από τα ενδιάμεσα βήματα λειτουργούν σαν βάση για περαιτέρω ανάλυση κατευθυνόμενη από τον χρήστη.

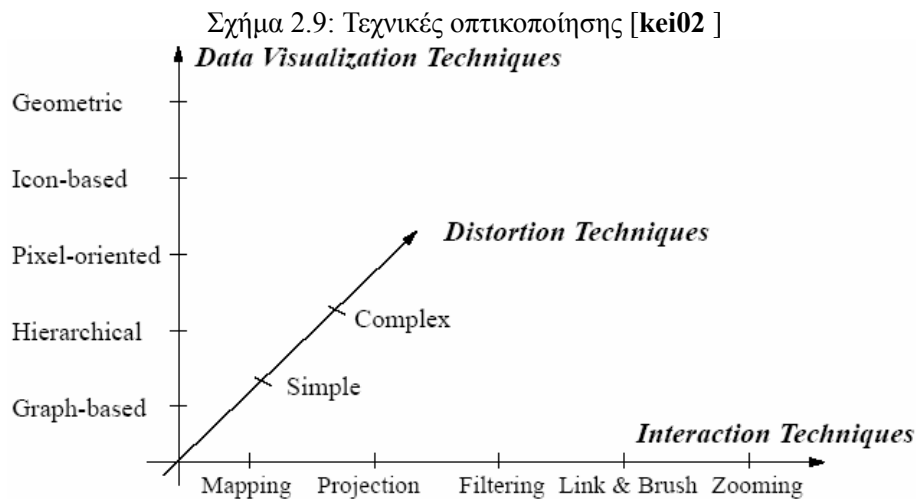
Σχήμα 2.8: Επισκόπηση των διαφορετικών προσεγγίσεων της ανθρώπινης ανάμειξης [Dan05]



Υπάρχουν πέντε ευρύτερες ομάδες ταξινόμησης σύμφωνα με την ταξονομία (taxonomy) των Keim και Ankerst για τις τεχνικές οπτικοποίησης δεδομένων.

1. Γεωμετρικές (Geometric)

2. Βασιζόμενες σε εικόνες (Icon-based)
3. Ιεραρχικές (Hierarchical)
4. Βασιζόμενες σε γράφους (Graph based)
5. Βασιζόμενες σε εικονοστοιχεία (Pixel Oriented)
6. Τεχνικές παραμόρφωσης εικόνας (Distortion Techniques)



Τεχνικές μείωσης διαστάσεων

Οι περισσότερες τεχνικές οπτικοποίησης χρησιμοποιούν μεθόδους μείωσης διαστάσεων (Dimensionality Reduction Techniques) με σκοπό τη μείωση των μεταβλητών στις πιο σημαντικές μέσω μίας κατάλληλης διαδικασίας. Αντιπροσωπευτικές τεχνικές της κατηγορίας αυτής είναι οι επόμενες:

1. Projection Pursuit / Grand Tour Method
2. Ανάλυση κύριων συνιστωσών (principal components analysis)
3. Biplots
4. Πολυδιάστατη κλιμάκωση (multidimensional scaling)
5. Χάρτες του Sammon (Sammon's mapping)
6. Χάρτες αυτοοργάνωσης (self-organizing maps)

Γεωμετρικές τεχνικές

Οι γεωμετρικές τεχνικές (Geometric Techniques) έχουν σαν στόχο την εύρεση προβολών πολυδιάστατων δεδομένων στο δισδιάστατο χώρο και αναπαριστούν τα δεδομένα ως γραμμές, επιφάνειες, ραβδογράμματα, ή διαγράμματα διασποράς. Οι βασικότερες τεχνικές αυτής της κατηγορίας είναι οι scatter-plots πίνακες, το τοπογράφημα, οι τεχνικές προβολής, η τεχνική hyperslice, η τεχνική parallel coordinates, η τεχνική Circular Parallel Coordinates και η τεχνική Radial Coordinate Visualization.

Τεχνικές βασισμένες σε εικόνες

Η βασική ιδέα αυτών των τεχνικών είναι η παρουσίαση κάθε εγγραφής ως ένα συγκεκριμένο γραφικό αντικείμενο όπου οι τιμές των μεταβλητών ανατίθενται σε συγκεκριμένα χαρακτηριστικά των αντικειμένων. Η ανάθεση είναι τέτοια ώστε η συνολική εμφάνιση των αντικειμένων να μεταβάλλεται ως συνάρτηση ρύθμισης των τιμών. Κάθε μονάδα αντιπροσωπεύεται από κάποιο εικονίδιο το οποίο έχει επιλεγεί με βάση συγκεκριμένους κανόνες ώστε να εκφράζει τα χαρακτηριστικά της κάθε εγγραφής. Οι πιο γνωστές τεχνικές αυτής της κατηγορίας είναι τα πρόσωπα Chernoff, τα γραμμικά σχήματα, η τεχνική shape coding, η έγχρωμη μορφοποίηση σχήματος και η τεχνική TileBars.

Ιεραρχικές τεχνικές

Στις τεχνικές αυτές η οπτικοποίηση των δεδομένων επιτυγχάνεται με χρήση μιας ιεραρχικής κατηγοριοποίησης της οθόνης σε υπό-μήματα. Οι σημαντικότερες μέθοδοι που εμπίπτουν σε αυτή την κατηγορία είναι οι εξής:

1. Dimensional Stacking
2. Ενθυμακωμένα διαγράμματα (World within Worlds)
3. Δενδρική ενθυλακωμένη παράθεση (Treemap)
4. Κωνικά δένδρα (Cone Tree)
5. Πληροφοριακός κύβος (InfoCube)

Τεχνικές βασισμένες σε γράφους

Οι πιο γνωστές τεχνικές βασισμένες σε γράφους είναι ο ορθογώνιος γράφος, ο συμμετρικός γράφος, ο γράφος βασισμένος σε ομάδες, ο κατευθυνόμενος ακυκλικός γράφος και οι υπεργράφοι.

Τεχνικές σχεδίασης με εικονοστοιχεία

Στις μεθόδους αυτές η αναπαράσταση κάθε τιμής ενός χαρακτηριστικού ενός πίνακα αναπαρίσταται από ένα χρωματισμένο εικονοστοιχείο (pixel). Ο χρωματισμός επιλέγεται από ένα σταθερό χάρτη χρωμάτων και παρουσιάζει την σχετικότητα (relevance) των δεδομένων. Κάθε χαρακτηριστικό παρουσιάζεται σε διαφορετικά υπό-παράθυρα στην οθόνη. Με την χρήση των εικονοστοιχείων μπορούν να αναπαρασταθούν ως και 1.000.000 τιμές χαρακτηριστικών. Οι διάφορες τεχνικές που έχουν αναπτυχθεί κατά καιρούς διαφοροποιούνται ως προς τον τρόπο τοποθέτησης των

pixels. Οι τεχνικές αυτές διαχωρίζονται σε δύο κύριες υποκατηγορίες, αυτές που είναι ανεξάρτητες από ερωτήσεις και αυτές που είναι εξαρτώμενες από ερωτήσεις.

Τεχνικές παραμόρφωσης εικόνας

Η βασική ιδέα των τεχνικών παραμόρφωσης εικόνας (Distortion Techniques) είναι η παρουσίαση μέρους των δεδομένων με μεγάλη λεπτομέρεια, μειώνοντας αντίστοιχα τον χώρο απεικόνισης για τα υπόλοιπα δεδομένα που εμφανίζονται σε χαμηλότερα επίπεδα λεπτομέρειας. Οι διασημότερες είναι η τεχνική Fisheye View, η τεχνική Perspective Wall, η τεχνική Table Lens, η τεχνική δένδρων υπερβολικής γεωμετρίας και η τεχνική υπέρ-κύβου.

Κεφάλαιο 3

Ανάλυση αναφορών

Σε κάθε επιστημονικό άρθρο που δημοσιεύεται καθοριστικό ρόλο παίζουν οι παραπομπές του. Μέσω αυτών γίνεται ο συσχετισμός της ερευνητικής επιρροής και χρησιμοποιείται και ως ένα αρκετά αξιόπιστο μέτρο για την εκτίμηση των ερευνητικών επιπτώσεων. Ένας σημαντικός παράγοντας για αυτό είναι ότι είναι πλήρεις, ποσοτικές και αντικειμενικές και μπορούν να υπολογιστούν με αυτόματο τρόπο, με λίγη ή και καθόλου ανθρώπινη παρέμβαση. Για να ξεπεραστούν οι αδυναμίες που προέκυψαν για την παραγωγική μέτρηση των παραπομπών, και τη χρήση τους ως μέτρο, έχει προστεθεί μία λίστα από διεξοδικά μέτρα. Τα πιο γνωστά από αυτά, που χρησιμοποιούνται και από τις υπηρεσίες έυρεσης αναφορών, είναι τα εξής:

1. Μέσος αριθμός παραθέσεων ανά επιστημονικό άρθρο

Αυτό το μέτρο στοχεύει στο να επανορθώσει το γεγονός, ότι μερικοί συγγραφείς δημοσιεύουν περισσότερο, και ίσως φαίνεται να έχουν υψηλότερη επιρροή, όταν μόνο ο συνολικός αριθμός των παραπομπών λαμβάνεται υπόψιν, όπου στην πραγματικότητα κάθε ένα από τα έργα τους παρατίθεται σποραδικά. Ταυτόχρονα, κατά ένα μέρος αντισταθμίζει τις διαφορές στον αριθμό των χρόνων στους οποίους έχουν δραστηριοποιηθεί οι ερευνητές.

2. Μέσος αριθμός παραθέσεων ανά συγγραφέα

Αυτή η μέτρηση έχει ως στόχο να αντισταθμίσει το γεγονός ότι τα επιστημονικά άρθρα με ένα συγγραφέα και αυτά που αποτελούν συνεταιριστικά έργα (έχουν παραπάνω από έναν συγγραφέα), δεν δείχνουν τον ίδιο βαθμό ιδιοκτησίας/συμβολής και για αυτό δεν πρέπει να συνδέονται με το ίδιο βαθμό της προσωπικής επιρροής. Έτσι, καταμέτρηση των παραθέσεων για το κάθε επιστημονικό άρθρο κατανέμεται εξίσου σε όλους τους συγγραφείς που έχουν συμβάλει σε αυτό. Διάφορες παραλλαγές είναι δυνατές, όπως η πίστωση μεγαλύτερου μέρους συνεισφοράς για τον ηγετικό συγγραφέα, ή και η σταδιακή μείωση του ποσοστού της συνεισφοράς που βασίζεται στη θέση που κατέχει ο συγγραφέας στη λίστα του.

3. Μέσος αριθμός παραθέσεων ανά έτος

Αυτό το μέτρο στοχεύει στο να επανορθώσει το γεγονός ότι οι ερευνητές που έχουν δραστηριοποιηθεί για μεγαλύτερο χρονικό διάστημα αναπόφευκτα έχουν παράγει και περισσότερο έργο. Υπολογίζοντας λοιπόν το μέσο όρο για πολλά χρόνια, αυτός ο τρόπος μέτρησης υπολογίζει μία πιο δίκαιη τιμή των παραθέσεων, που επιτρέπει τη σύγκριση των έμπειρων και των νέων ερευνητών.

4. h-δείκτης και παρόμοιοι δείκτες

Ένας επιστήμονας έχει δείκτη h, εάν τα N_p επιστημονικά άρθρα αυτού ή αυτής έχουν τουλάχιστον h αναφορές η κάθε, μία και τα άλλα N_p-h επιστημονικά άρθρα έχουν h αναφορές η κάθε μία. Έχουν γίνει διάφορες προτάσεις για την τροποποίηση του h-δείκτη, με σκοπό να δοθεί έμφαση στα διαφορετικά χαρακτηριστικά. Για παράδειγμα, υπάρχει μία διακύμανση που λαμβάνει υπόψιν των αριθμό των συγγραφέων σε κάθε επιστημονικό άρθρο, και άλλη μία που υπολογίζει το μέσο όρο των h-δεικτών, για όλα τα χρόνια της δραστηριότητάς τους. Ο h-δείκτης διαχωρίζεται σε H1, H2 και H3, κάνοντας διάκριση έτσι μεταξύ των διαφόρων τύπων των συγγραφέων, όπως οι τελειομανείς και οι μαζικοί παραγωγοί.

Ο g-δείκτης αποτελεί μία πολύ γνωστή επέκταση του h-δείκτη του Hirsch, και καταμετρά τις αναφορές κατά μέσο όρο. Ο e-δείκτης δίνει την τεταγωνική ρίζα των παραπομπών στο h-set, πέραν του h₂, δηλαδή πέραν του θεωρητικού ελαχίστου που απαιτείται για την απόκτηση ενός h-δείκτη, η οποία το καθιστά χρήσιμο για τη σύγκριση των συγγραφέων με το ίδιο h-δείκτη.

5. s-δείκτης

Ο s-δείκτης βασίζεται στην έννοια της εντροπίας και φαίνεται να περιέχει μία καλύτερη ποσοτικοποίηση της επιρροής των ερευνητών από τον h-δείκτη, αλλά μόνο για τους ερευνητές με μεγάλο αριθμό δημοσιευμένων άρθρων και καταλόγων παραπομπών.

6. i-10 δείκτης

Ο i-10 δείκτης χρησιμοποιείται στο Google Scholar, και ισοδυναμεί με το πλήθος των άρθρων που έχουν το λιγότερο δέκα παραπομπές.

Οι παραπάνω δείκτες όμως αποτυγχάνουν να εξετάσουν το περιεχόμενο / πλαίσιο της κάθε παραπομπής.

Όπως είναι λογικό, δεν έχουν όλες οι αναφορές, που παρουσιάζονται στην ενότητα βιβλιογραφία κάποιου έργου, τον ίδιο ρόλο. Το CiteSeerX, για παράδειγμα, παρέχει μία επιπλέον πληροφορία για το πλαίσιο των παραπομπών, αλλά αυτό δεν λαμβάνεται υπόψιν στον υπολογισμό των μετρήσεων. Για να λειτουργήσει βέβαια αυτό θα απαιτούσε ανθρώπινη παρέμβαση, και για την ακρίβεια την παρέμβαση ενός εμπειρογνώμονα, που θα ήταν σε θέση να αξιολογήσει το αντίκτυπο της κάθε παραπομπής. Για να είναι αυτό δυνατό όμως χρειάζονται πόροι, οι οποίοι δεν είναι εύκολα διαθέσιμοι και παράλληλα αλλοιώνουν τον αντικειμενικό χαρακτήρα του όλου έργου. Συνεπώς, μία τέτοια βαθιά μελέτη δεν είναι πρακτικά εφικτή.

Άλλη μία προσέγγιση στο θέμα αυτό, είναι η εξέταση του τμήματος του κειμένου στο οποίο αναφέρονται οι παραπομπές. Αναλυτικότερα, παρατηρείται ότι οι παραπομπές στην εισαγωγή του κειμένου τυπικά έχουν μικρή ή και καθόλου επιρροή στο έργο. Ενώ οι παραπομπές στα αποτελέσματα, που συνήθως χρησιμοποιούνται ως σημεία αναφοράς, καθώς και οι παραπομπές στο κομμάτι περιγραφής της προτεινόμενης μεθοδολογίας, που πιθανότητα έχουν χρησιμοποιηθεί ως μεθοδολογική βάση, είναι αυτές που έχουν και την υψηλότερη επιρροή. Δυστυχώς αυτό είναι επίσης ανεφάρμοστο, αφού απαιτεί όλες οι παραπομπές να περιλαμβάνονται στο επιστημονικό άρθρο και τα τμήματα να προσδιορίζονται ξεκάθαρα, πράγμα το οποίο δεν αποτελεί συχνή περίπτωση.

Αφού αναφερθήκαμε στους πιο γνωστούς τρόπους για την μέτρηση των παραπομπών και για το ρόλο των αυτών και του περιεχομένου τους, θα παρουσιάσουμε μία περισσότερο ενοιολογική προσέγγιση που σχετίζεται με τη σύνδεση της κάθε παραπομπής με το αντίστοιχο επιστημονικό πεδίο.

Κάθε δημοσιευμένο έργο συνδέεται στενά με ένα ή περισσότερα επιστημονικά πεδία. Στη δεύτερη περίπτωση, λέγεται πως το έργο έχει διεπιστημονικό χαρακτήρα, ή απλά μπορούμε να πούμε πως βασίζεται σε ιδέες από διαφορετικούς επιστημονικούς τομείς. Εάν και είναι συζητήσιμο το πόσο σημαντικό είναι να γνωρίζει κανείς το επιστημονικό πεδίο του κάθε επιστημονικού άρθρου, το ανάποδο παρουσιάζει σαφώς περισσότερο ενδιαφέρον. Ειδικότερα, εξετάζοντας το επιστημονικό πεδίο ενός αναφερόμενου επιστημονικού άρθρου να μπορούμε να ανακαλύψουμε ποια πεδία της επιστήμης έχουν επηρεαστεί από το δεδομένο άρθρο. Με τον τρόπο αυτό δίνεται η δυνατότητα σε κάποιον ερευνητή, ο οποίος θα θέλει να ασχοληθεί με ένα συγκεκριμένο πεδίο των επιστημών, να εστιάσει κατευθείαν στις δημοσιεύσεις που είχαν τη μεγαλύτερη επιρροή και έτσι θα γλιτώσει πολύ χρόνο και κόπο.

Το ζήτημα που τίθεται τώρα λοιπόν είναι το πώς θα προσδιοριστεί το επιστημονικό πεδίο μίας αναφοράς. Για να γίνει όμως αυτό βασική προϋπόθεση είναι να αναλυθεί η αναφορά για να εξακριβωθεί το περιεχόμενό της.

Υπάρχουν διάφοροι τρόποι να αναλύσει κανείς μία αναφορά. Μπορεί να ελέγχει απλά τον τίτλο ή την περίληψη του κειμένου ή ακόμα να βασίζεται σε λέξεις-κλειδιά. Ωστόσο, ο τίτλος μπορεί μερικές φορές να είναι παραπλανητικός, ή η περίληψη να μη δίνει όλα τα απαραίτητα στοιχεία που χρειάζεται κανείς για να αποφασίσει σε ποιον επιστημονικό κλάδο εντάσσεται, και οι λέξεις-κλειδιά μπορεί να μην υπάρχουν ή να είναι δύσκολο να εντοπιστούν. Η κειμενική ανάλυση δεν είναι ακόμα αρκετά ώριμη, ώστε να εγγυάται σχεδόν τέλεια αποτελέσματα και ο τίτλος και οι λέξεις-κλειδιά σίγουρα δεν είναι αρκετά αξιόπιστα.

Ωστόσο, το μέσο δημοσίευσης μπορούμε να πούμε ότι αποτελεί καλή ένδειξη για το επιστημονικό πεδίο μίας αναφοράς. Όταν ένα άρθρο παραλαμβάνεται είτε από κάποιο περιοδικό είτε από κάποιο συνέδριο, η σχετικότητά του εξετάζεται μαζί με την επιστημονική ποιότητα. Ως εκ τούτου, η συντακτική διαδικασία, εγγυάται ότι το θέμα της δημοσίευσης σχετίζεται στενά με τα θέματα που ασχολείται το αντίστοιχο περιοδικό ή συνέδριο. Για παράδειγμα, όλα τα άρθρα που έχουν δημοσιευτεί στην IEEE στο Image Processing έχουν να κάνουν με την επεξεργασία εικόνας. Αντίστοιχα, όλες οι αναφορές που έχουν παρουσιαστεί στο SMAP είναι σχετικές με τα θέματα που απαριθμεί το ίδιο το συνέδριο.

Σχεδόν όλα τα περιοδικά και συνέδρια έρχονται σαφώς με καθορισμένα πεδία και καταλόγους θεμάτων, και για αυτά που δεν το κάνουν είναι σχετικά εύκολο να το κάνει κανείς χειροκίνητα, αφού αυτό πρέπει να γίνεται μία φορά για κάθε μέσο δημοσίευσης και όχι ξεχωριστά για κάθε δημοσιευμένη εργασία. Συνεπώς, η αυτοματοποιημένη και αντικειμενική, χωρίς τη γνώμη κάποιου εμπειρογνώμονα ανθρώπου, εξέταση της επιστημονικής έκτασης μίας δεδομένης δημοσιευμένης αναφοράς είναι εφικτή.

Η πρότασή μας, στην οποία και βασιστήκαμε, είναι ότι μεγάλο ενδιαφέρον παρουσιάζει και το πεδίο εφαρμογής της ανάλυσης των παραπομπών, καθώς παρέχει μία πιο πλούσια εικόνα για το αντίκτυπο της κάθε δημοσίευσης.

Κεφάλαιο 4

Επιστήμες

Με τον όρο επιστήμη εννοούμε το οργανωμένο σώμα της εξακριβωμένης και τεκμηριωμένης γνώσης (Θεαίτητος, Πλάτωνας).

Οι επιστήμες χωρίζονται σε τέσσερις μεγάλες ομάδες:

1. Θετικές και φυσικές επιστήμες
2. Κοινωνικές επιστήμες
3. Ανθρωπιστικές επιστήμες
4. Εφαρμοσμένες επιστήμες

Συνήθως οι φυσικές επιστήμες ταξινομούνται μαζί με τα μαθηματικά, τη στατιστική και τη πληροφορική, συναποτελώντας έτσι τις θετικές επιστήμες. Εμείς επιλέξαμε να κρατήσουμε και τα δύο ονόματα θετικές και φυσικές επιστήμες για να τονίσουμε τη στενή σχέση των φυσικών επιστημών με τη φυσική, τη χημεία, το περιβάλλον, τη φύση και τη ζωή, ξεχωρίζοντας έτσι τον κλάδο των μαθηματικών, και δίνοντας έμφαση στο τομέα της πληροφορικής. Ταυτόχρονα, προσθέσαμε άλλη μία ομάδα επιστημών, τις επιστήμες σύμπαντος.

Φυσικές επιστήμες

Φυσικές επιστήμες (natural sciences) είναι οι επιστήμες των οποίων η μελέτη σχετίζεται με την περιγραφή, την πρόβλεψη και την κατανόηση των φυσικών φαινομένων, με βάση τις παρατηρησιακές και εμπειρικές ενδείξεις. Στις φυσικές επιστήμες, οι υποθέσεις πρέπει να ορίζονται επιστημονικά για να θεωρούνται επιστημονικές θεωρίες. Η εγκυρότητα, η ακρίβεια και κοινωνικοί μηχανισμοί διασφαλίζουν ποιοτικό έλεγχο, έχοντας, επίσης, ως ομότιμα κριτήρια την επαναληψιμότητα των ευρυμάτων, ανάμεσα σε άλλα κριτήρια και μεθόδους που χρησιμοποιήθηκαν για το σκοπό αυτό. Δίνουν κυρίως έμφαση στα ποσοτικά δεδομένα, που η έρευνά τους παράγει, δοκιμάζει και επιβεβαιώνει, μέσω της επιστημονικής μεθόδου, μερικές φορές αποκαλούνται «σκληρές επιστήμες».

Διαχωρίμε τις Φυσικές επιστήμες στους ακόλουθους τέσσερις κύριους κλάδους:

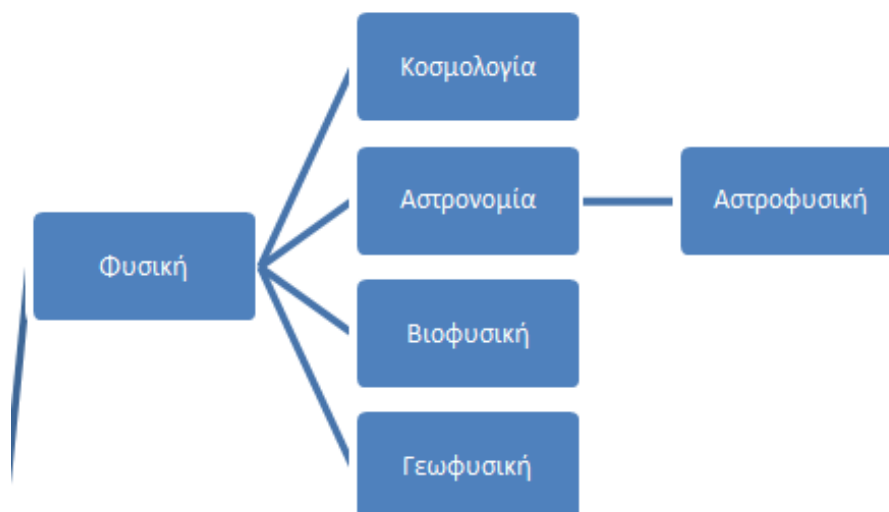
1. Φυσική

2. Χημεία
3. Επιστήμες γης και περιβάλλοντος
4. Επιστήμες ζωής / Βιολογία

Η διάκριση αυτή βασίζεται στον αρχικό διαχωρισμό των φυσικών επιστημών που ήταν σε Αστρονομία, Βιολογία, Γεωεπιστήμες (ή επιστήμες γης), Φυσική και Χημεία. Την Αστρονομία όμως επιλέξαμε να την εντάξουμε σαν υποκλάδο της Φυσικής με την οποία σχετίζεται άμεσα.

Στο κλάδο της Φυσικής, σχήμα 4.1, συμπεριλάβαμε όλα τα πεδία που ασχολούνται με τη μελέτη της ύλης (Κοσμολογία), της κίνησης της μέσα στο χώρο και το χρόνο, τις σχετικές ποσότητες όπως οι ενέργεια και η δύναμη, ό,τι σχετίζεται με την κατανόηση της συμπεριφοράς του σύμπαντος (Αστρονομία, Αστροφυσική), καθώς και την ανάπτυξη συγκεκριμένων τεχνολογικών ή πρακτικών εφαρμογών (Βιοφυσική, Γεωφυσική).

Σχήμα 4.1: Φυσική

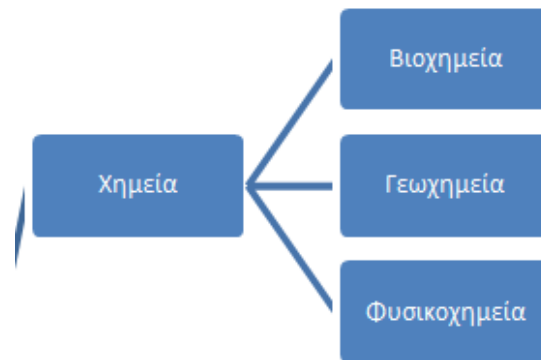


Στόν κλάδο της Χημείας, σχήμα 4.2, εντάξαμε τους κλάδους που δίνουν έμφαση στη μελέτη της ύλης, της σύνθεσής της, της δομής της, της ιδιότητές της και ιδιαίτερα τις μεταβολές της σύστασής της, δηλαδή τις χημικές αντιδράσεις. Η Χημεία ονομάζεται και κεντρική θετική επιστήμη γιατί παρεμβάλλεται και συνδέει τη Φυσική (Φυσικοχημεία) με τη Γεωλογία (Γεωχημεία) και τη Βιολογία (Βιοχημεία).

Στις επιστήμες γης και περιβάλλοντος, σχήμα 4.3, συγκατελέξαμε τα επιστημονικά πεδία τον πλανήτη Γη, συμπεριλαμβανομένου του εδάφους (Γεωλογία) του αλλά και της υδρόσφαιρας (Ωκεανογραφία) και της ατμόσφαιρας (Μετεωρολογία, Κλιματολογία).

Τέλος, για τις επιστήμες ζωής, σχήμα 4.4, λάβαμε υπόψιν ό,τι σχετίζεται με τη μελέτη της δομής και της λειτουργίας και της συμπεριφοράς των εμβρίων οργανισμών σε φυσιολογικές και παθολογικές καταστάσεις (Ζωολογία, Γενετική κ.α.).

Σχήμα 4.2: Χημεία



Σχήμα 4.3: Επιστήμες γης και περιβάλλοντος

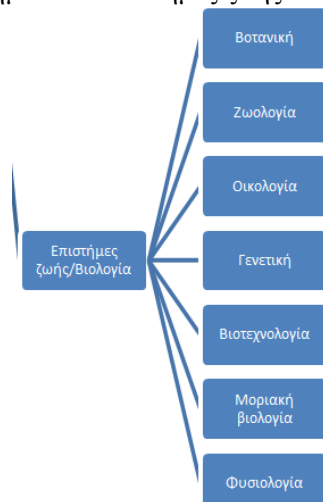


Θετικές επιστήμες

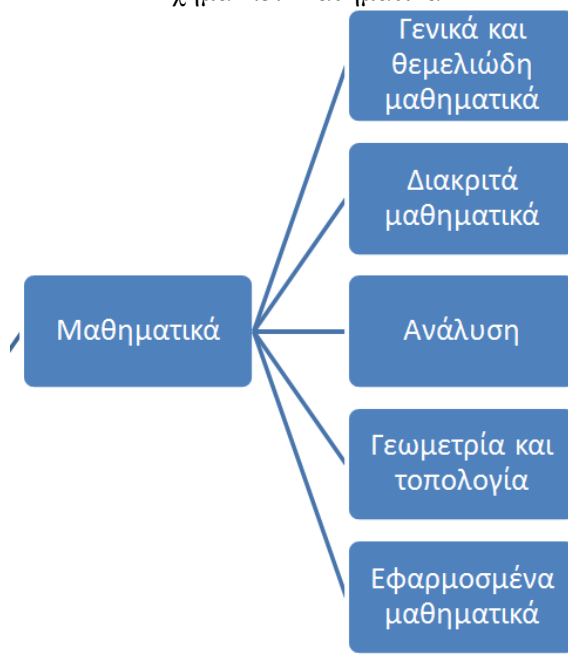
Οι θετικές επιστήμες, εάν αυτές μελετηθούν ξεχωριστά, συνίστανται από τρεις κλάδους, σημειώνοντας ιδιαίτερο ενδιαφέρον στον τομέα της πληροφορικής, στον οποίο και εστιάζουμε. Οι τριπλή διάκριση γίνεται ανάμεσα στα παρακάτω επιστημονικά πεδία:

1. Μαθηματικά
2. Στατιστική
3. Πληροφορική ή επιστήμη υπολογιστών

Σχήμα 4.4: Επιστήμες ζωής / Βιολογία



Σχήμα 4.5: Μαθηματικά



Η Στατιστική μπορεί να συγκαταλεχθεί είτε στον κλάδο των εφαρμοσμένων μαθηματικών, μιας και η στατιστική έρευνα βασίζεται στη χρήση της στατιστικής θεωρίας που είναι κλάδος των εφαρμοσμένων μαθηματικών, είτε να αποτελέσει έναν ξεχωριστό κλάδο των θετικών επιστημών. Εμείς επιλέξαμε τη δεύτερη πρόταση και μπορούμε να τη στηρίξουμε με το γεγονός ότι η Στατιστική έχει πολλούς εξειδικευμένους επιστημονικούς τομείς όπως η Βιοστατιστική, Επιχειρηματική, Οικονομική, Μηχανική, Φυσική Στατιστική, Δημογραφία, Ψυχολογική, Κοινωνική, Ανάλυση διαδικασιών, Χημειομετρία, Οικονομετρία, Τεχνομετρία κ.α.

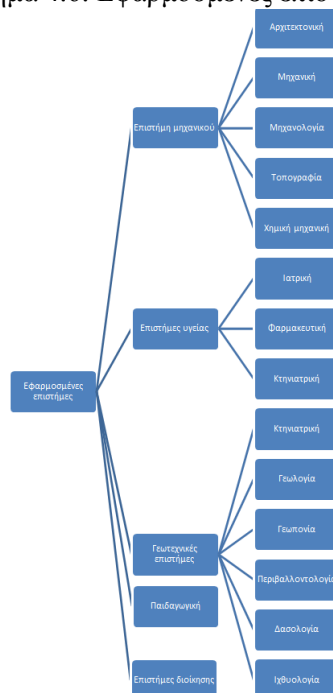
Για την επιστήμη της πληροφορικής επιλέξαμε να χρησιμοποιήσουμε το μοντέλο της ACM (Association for Computing Machinery) και συγκεκριμένα την τελευταία έκδοση του 2012. Η έκδοση αυτή διακρίνει την πληροφορική στους εξής κλάδους:

1. General Literature
2. Hardware
3. Computer Systems Organization
4. Software
5. Data
6. Theory of Computation
7. Mathematics of Computing
8. Information Systems
9. Computing Methodologies
10. Computer Applications
11. Computing Milieux

Εφαρμοσμένες επιστήμες

Στόχος τους είναι η εφαρμογή της επιστημονικής γνώσης για την επίλυση πρακτικών προβλημάτων και την ανάπτυξη συστηματικών μεθοδολογιών για αυτό το σκοπό. Οι εφαρμοσμένες επιστήμες, σε πρώτο επίπεδο, συμπεριλαμβάνουν την επιστήμη του μηχανικού, τις επιστήμες υγείας, τις γεωτεχνικές επιστήμες, την παιδαγωγική και τις επιστήμες διοίκησης.

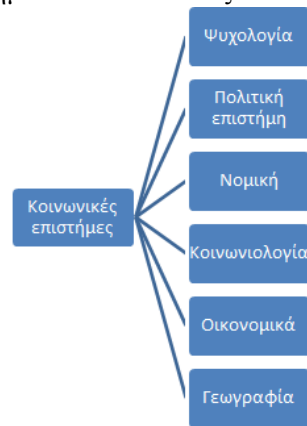
Σχήμα 4.6: Εφαρμοσμένες επιστήμες



Κοινωνικές επιστήμες

Μελετούν πτυχές της κοινωνικής ζωής των ανθρώπινων ομάδων και σε αυτές συγκαταλέγονται η Ψυχολογία, η Πολιτική επιστήμη, η Νομική, η Κοινωνιολογία, τα Οικονομικά και η Γεωγραφία.

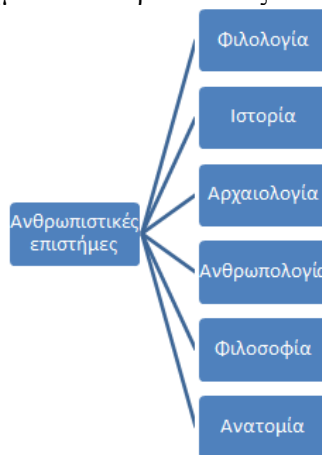
Σχήμα 4.7: Κοινωνικές επιστήμες



Ανθρωπιστικές επιστήμες

Εννοούνται οι επιστήμες που ασχολούνται με τον άνθρωπο και τη γενική συμπεριφορά του. Σε αντίθεση με τις φυσικές επιστήμες, βασίζονται στην ποιοτική έρευνα και από μερικούς αποκαλούνται και «μαλακές επιστήμες». Τέτοιες επιστήμες είναι η Φιλολογία, Ιστορία, Αρχαιολογία, Ανθρωπολογία, Φιλοσοφία και Ανατομία.

Σχήμα 4.8: Ανθρωπιστικές επιστήμες



Επιστήμες σύμπαντος

Στην εκτεταμένη ιεραρχία των επιστημών που βασιστήκαμε, διακρίναμε μία ακόμα κατηγορία επιστημών που σχετίζονται με το σύμπαν. Με τον τρόπο αυτόν στοχεύουμε να ξεκαθαρίσουμε ακόμα περισσότερο το τοπίο και να διαχωρίσουμε ξεκάθαρα τις λεπτές διαφορές μεταξύ των επιστημονικών πεδίων. Αυτή την κατηγορία συνιστούν η Κοσμολογία, η Μεταφυσική, η Θεολογία και η Αστρονομία.

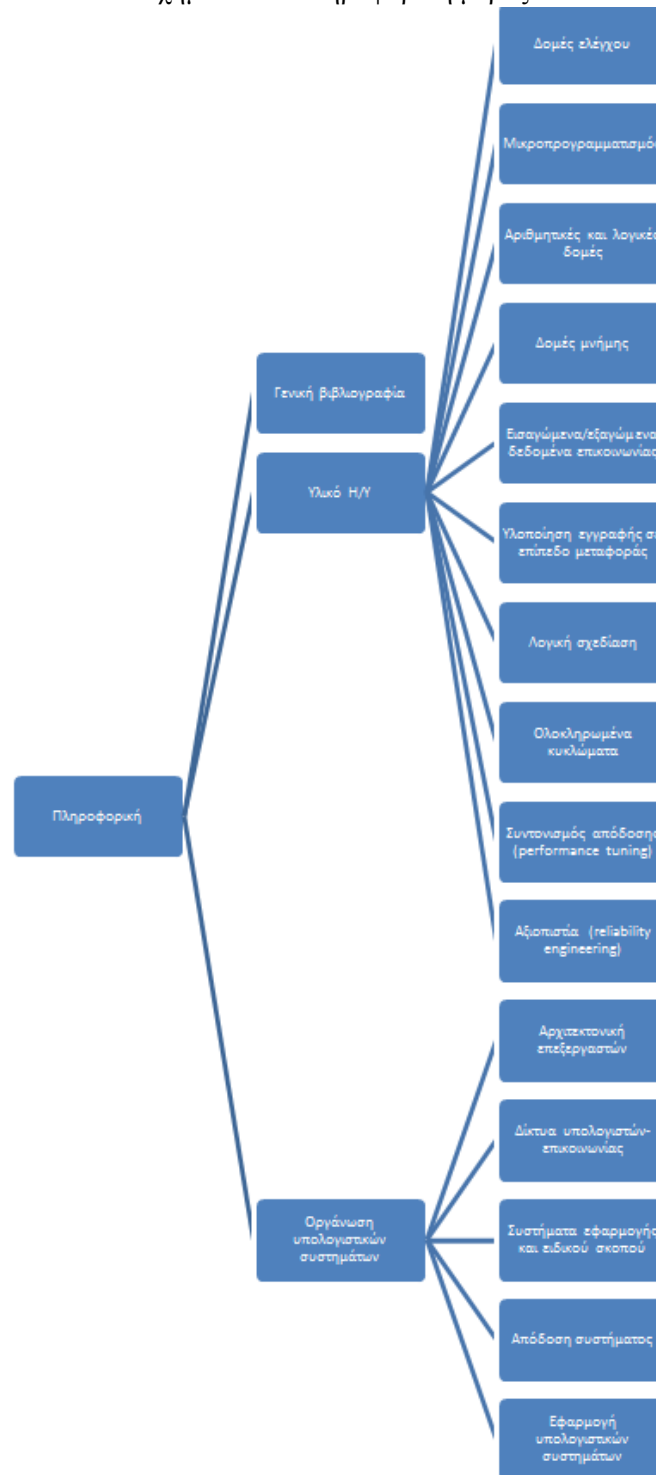
Σχήμα 4.9: Επιστήμες σύμπαντος



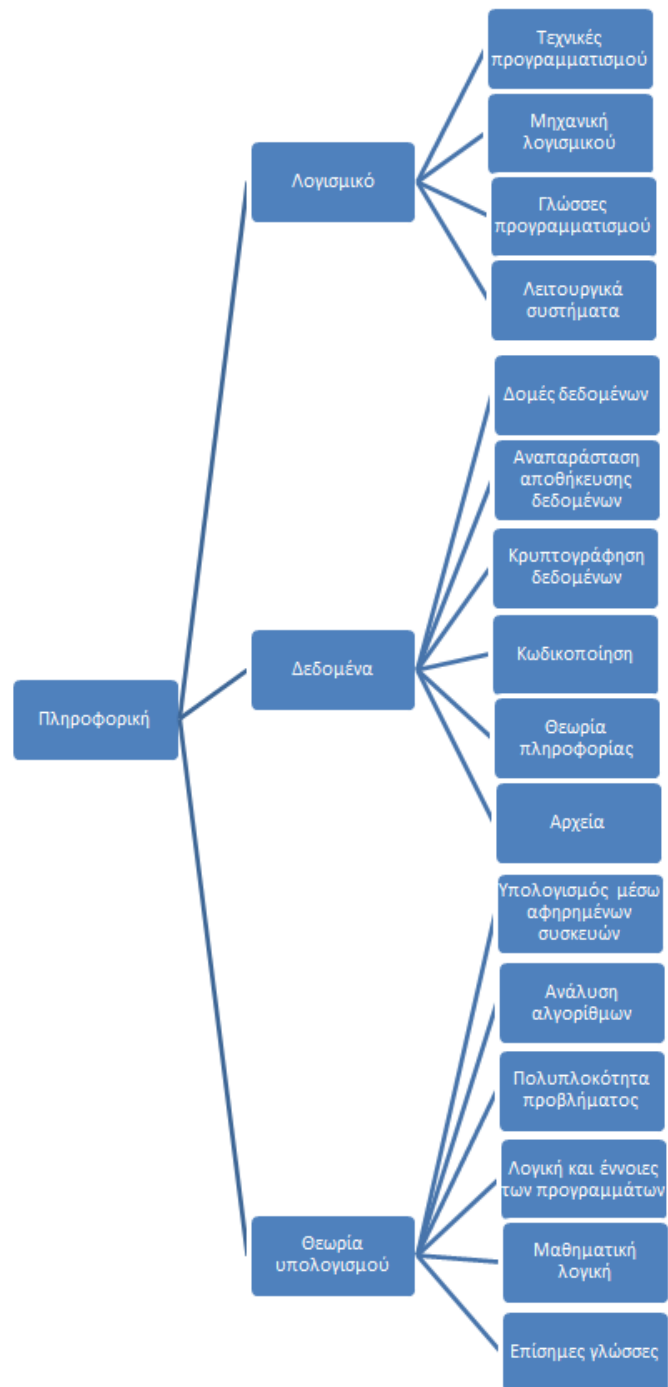
4.1 Η ιεραρχία των επιστημών

Για πρακτικούς λόγους σε αυτό το σημείο θα παρουσιάσουμε μόνο την ιεραρχία του κλάδου της πληροφορικής (η οποία δεν αναφέρθηκε παραπάνω) και ολόκληρη η ιεραρχία των επιστημών θα υπάρχει στο σχετικό παράρτημα.

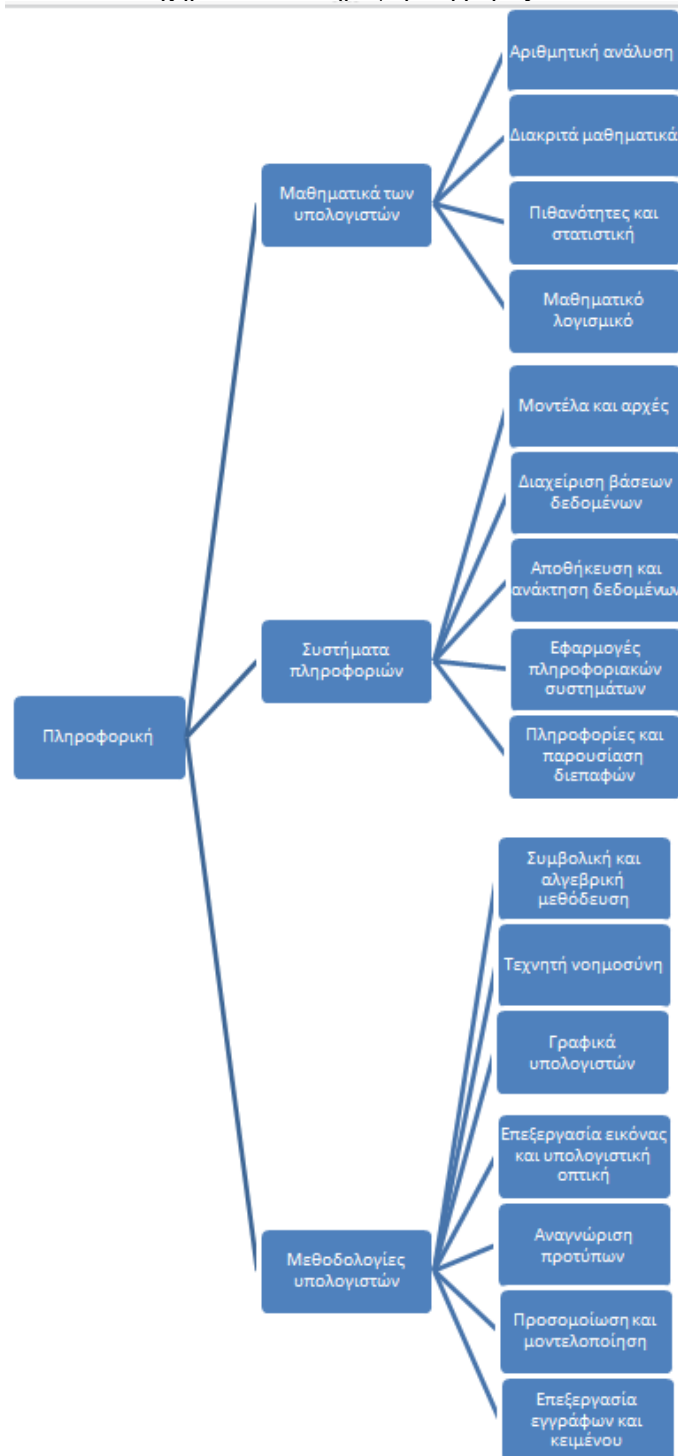
Σχήμα 4.10: Πληροφορική μέρος 1



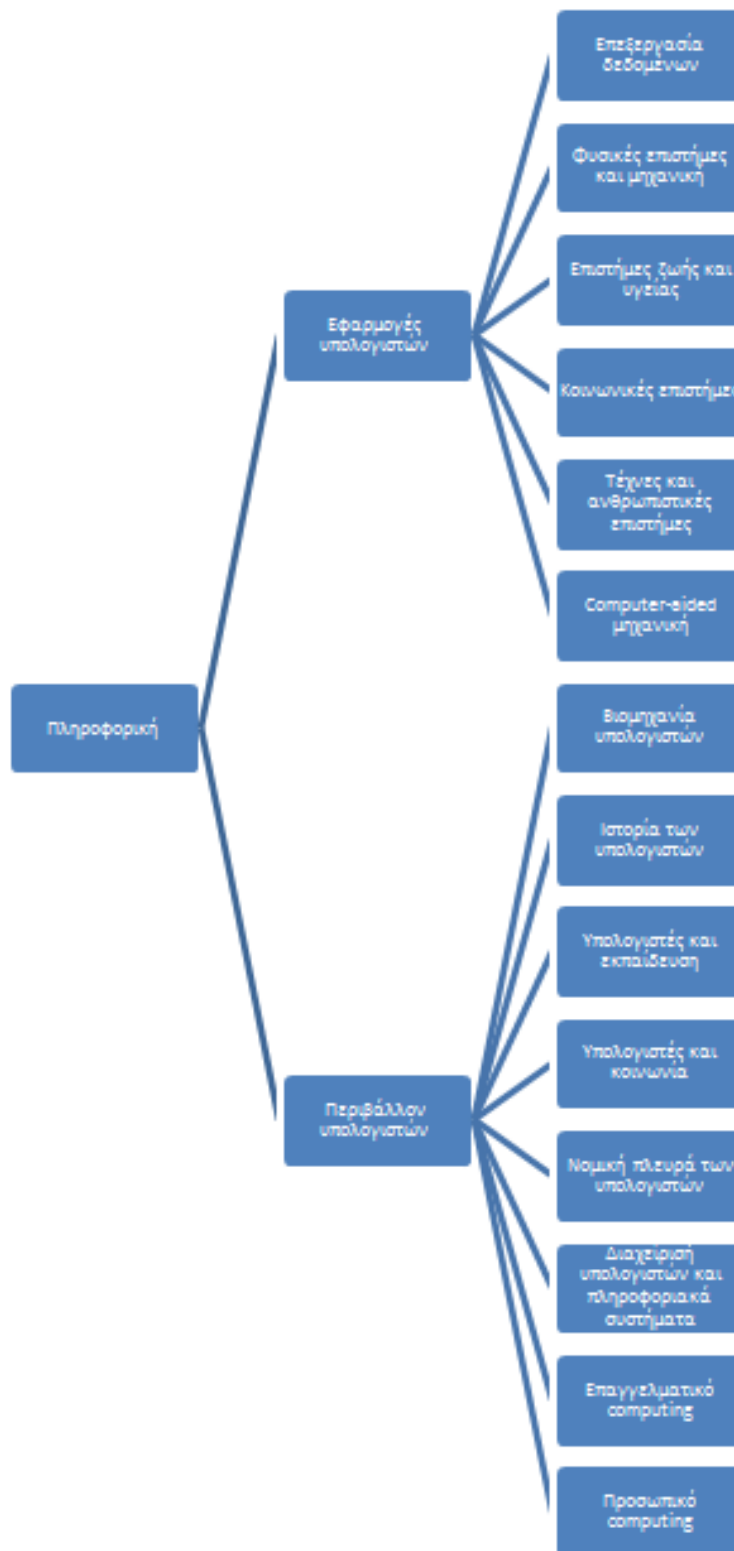
Σχήμα 4.11: Πληροφορική μέρος 2



Σχήμα 4.12: Πληροφορική μέρος 3



Σχήμα 4.13: Πληροφορική μέρος 4



4.2 Βαθμός συσχέτισης / Διεπιστήμες

Στην ιεραρχία που παρουσιάσαμε παραπάνω, με μία προσεκτική ματιά μπορεί κάποιος να διακρίνει ότι μερικά επιστημονικά πεδία υπάρχουν σε περισσότερες από μία κατηγορίες επιστημών. Για παράδειγμα, η κτηνιατρική εμφανίζεται τόσο ως επιστήμη υγείας αλλά και ως γεωτεχνική επιστήμη (και οι δύο υποκατηγορίες των εφαρμοσμένων επιστημών). Το ίδιο παρατηρείται και με τη Γεωλογία, την Αστρονομία, τη Γεωπονία και τα Διακριτά μαθηματικά. Αυτό συμβαίνει γιατί τα πεδία αυτά σχετίζονται με παραπάνω από έναν επιστημονικούς κλάδους, με αποτέλεσμα να μην είναι δυνατό κα σωστό να ενταχθούν μόνο σε μία κατηγορία, μιάς και περιέχουν ανάμεικτα στοιχεία. Τότε λέμε ότι αυτά τα πεδία έχουν διεπιστημονικό χαρακτήρα.

Ορίζοντας λοιπόν την παραπάνω πρόταση, διεπιστήμες ή διακλαδική επιστήμη (ή διεπιστημονικό πεδίο) ονομάζονται οι επιστήμες (ή το πεδίο) που συνθέτουν στοιχεία από διαφορετικούς κλάδους επιστημών.

Όμοια με το παράδειγμα της Κτηνιατρικής που αναφέρθηκε, αντίστοιχα ζητήματα προκύπτουν με την Κοσμολογία και την Αστρονομία που αμφιταλαντεύονται μεταξύ των επιστημών σύμπαντος και του κλάδου της Φυσικής. Τα Διακριτά μαθηματικά επίσης εμφανίζονται στα μαθηματικά των υπολογιστών (κλάδος της πληροφορικής) αλλά και στον κλάδο των μαθηματικών στις θετικές επιστήμες.

Πώς όμως θα αποφανθούμε για το με ποιόν από τους δύο τομείς των επιστημών συνδέονται πιο στενά οι διεπιστήμες που προέκυψαν;

Η λύση που χρησιμοποιήσαμε για να ξεπεράσουμε αυτό το ζήτημα είναι ο σχεδιασμός ενός κατευθυνόμενου γραφήματος με βάρη (λεγόμενο και δίκτυο). Έστω λοιπόν $G=(V,E)$, όπου V το σύνολο των κόμβων που αναπαριστούν τις διεπιστήμες και E το σύνολο των ακμών που δηλώνουν τη σχέση για ένα διατεταγμένο ζεύγος κορυφών. Το W είναι το βάρος που δείχνει το βαθμό συσχέτισης των διατεταγμένων κορυφών.

Η κλίμακα που επιλέξαμε για να εκφράζει το βάρος είναι από το 0 έως το 1. Άλλες ιδέες ήταν η κλίμακα από 1 έως 10 καθώς και τα ποσοστά επί τοις εκατό. Κρίναμε όμως καταλληλότερη και πιο ξεκάθαρη τη χρήση συντελεστών (π.χ. 0.5 για ίσα βάρη). Βέβαια αυτό αποτελεί καθαρά μία σχεδιαστική απόφαση.

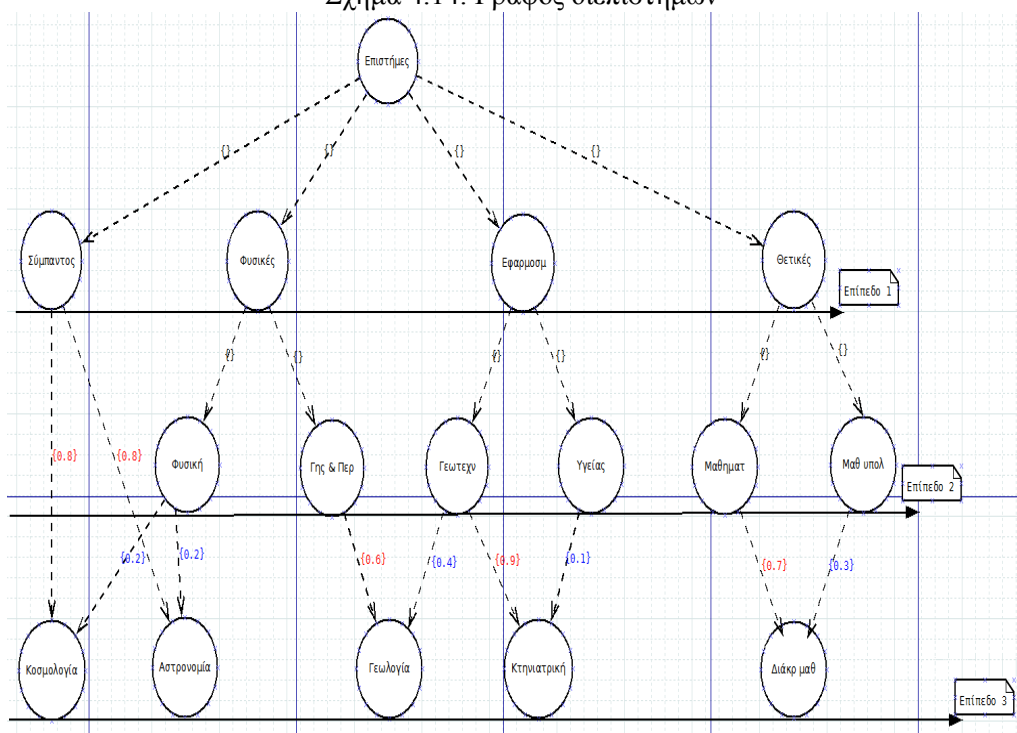
Μετά από έρευνα που πραγματοποιήσαμε καταλήξαμε στα εξής αποτελέσματα: Η Κτηνιατρική έχει συντελεστή βαρύτητας 0.1 για τις επιστήμες υγείας και 0.9 για τις γεωτεχνικές. Σε ένα δείγμα δημοσιεύσεων που εξετάσαμε παρατηρήσαμε ότι ένα μικρό ποσοστό σχετίζεται με τις επιστήμες υγείας. Η Γεωλογία σχετίζεται με τις επιστήμες γης και περιβάλλοντος με συντελεστή 0.6, και με τις γεωτεχνικές με συντελεστή 0.4. Αυτό σημαίνει πως το αντικείμενο της Γεωλογίας αφορά κατά ένα μεγάλο βαθμό τη μελέτη της γης (Εδαφολογία, φυσικοί πόροι) και την κατανόηση των

φυσικών φαινομένων (Ηφαιστειολογία, Σεισμολογία), ενώ το υπόλοιπο κομμάτι της (σύνδεση με τις γεωτεχνικές επιστήμες) δείχνει το βαθμό γνώσης που μας παρέχει η Γεωλογία για να κάνουμε τη γη (περιβάλλον) καλύτερη. Η Κοσμολογία και η Αστρονομία έχουν μεγαλύτερο βάρος 0.8 για τις επιστήμες σύμπαντος και μικρότερο 0.2 για τον κλάδο της Φυσικής. Αυτό συμβαίνει γιατί από τη μία πλευρά η φυσική με τους νόμους της βοηθά στην κατανόηση των φαινομένων αλλά από την άλλη η θέση τους είναι στις επιστήμες σύμπαντος αφού το σύμπαν είναι και το κομμάτι στο οποίο και εστιάζουν. Τέλος, στα διακριτά μαθηματικά αποδόθηκε 0.3 συντελεστής βαρύτητας για τον κλάδο της πληροφορικής και 0.7 για την επιστήμη των μαθηματικών. Βέβαια είναι ξεκάθαρο πως όσο σημαντικά κι αν είναι τα μαθηματικά για την πληροφορική δεν παύουν να αποτελούν ένα μικρό μέρος αυτής.

Γενικότερα, ο βαθμός συσχέτισης δείχνει το πόσο σχετίζεται το χαμηλότερο επίπεδο του δένδρου που θα παρουσιάσουμε με το υψηλότερο. Για να είμαστε πιο σαφείς, δηλώνει τη σχέση των κόμβων του χαμηλότερου επιπέδου με τους κόμβους του αμέσως προηγούμενου επιπέδου με τους οποίους συνδέεται με ακμές. Για παράδειγμα, φανερώνει το πόσο σχετίζεται η Κτηνιατρική με τις επιστήμες υγείας και πόσο με τις γεωτεχνικές επιστήμες.

Ο παρακάτω γράφος σχεδιάστηκε σε σχήμα δένδρου για να είναι πιο διακριτά τα επίπεδα, και παρουσιάζει το βαθμό συσχέτισης των διεπιστημών που εντοπίστηκαν.

Σχήμα 4.14: Γράφος διεπιστημών



Κεφάλαιο 5

Αξιοποίηση των δεδομένων

Σε αυτό το κεφάλαιο θα εξηγήσουμε τον τρόπο με τον οποίο συνδέσαμε την ιεραρχία των επισημών με τα μέσα οπτικοποίησης.

Το πρώτο βήμα ήταν να συγκεντρώσουμε όλα τα μέσα δημοσίευσης που εμφανίζονταν στο αρχείο με τις λίστες των αναφορών που μας δόθηκε. Έπειτα, ελέγξαμε ένα ένα τα ονόματα αυτών και ερευνήσαμε τα θέματα με τα οποία ασχολούνται και δημοσιεύουν ανά χρόνους. Έτσι κρίναμε με ποιές επιστήμες και επιστημονικά πεδία σχετίζεται το κάθε μέσο δημοσίευσης. Στη συνέχεια, συμβουλευτήκαμε την ιεραρχία των επιστημών που δημιουργήσαμε και παρουσιάσαμε στο κεφάλαιο 4, για να προσαρμόσουμε τις αντιστοιχίες των επιστημών, που βρήκαμε στο προηγούμενο βήμα, στη δική μας ιεραρχία. Το τελευταίο στάδιο ήταν να δημιουργήσουμε ένα νέο δικό μας αρχείο xml με τις τελικές και προσαρμοσμένες αντιστοιχίες των μεσων δημοσίευσης και των επιστημών, στις οποίες καταλήξαμε ακολουθώντας τα βήματα που αναφέρθηκαν παραπάνω.

Το νέο αρχείο xml περιλαμβάνει όλες τις απαραίτητες αντιστοιχίες μέσα στις κατάλληλες ετικέτες που ορίζονται από τη γλώσσα xml (EXtensible Markup Language). Συγκεκριμένα, στην ετικέτα <journal> θα βρίσκεται το όνομα του μέσου δημοσίευσης που εξετάζουμε κάθε φορά. Στην επόμενη ετικέτα <science> στο χαρακτηριστικό (attribute) “category” θα περιέχεται το όνομα την ευρύτερης επιστήμης στην οποία εντάσσεται (π.χ. φυσικές & θετικές, κοινωνικές, εφαρμοσμένες, σύμπαντος και ανθρωπιστικές) και στην οποία έχει επιρροή. Όμοια, υπάρχουν και άλλες ετικέτες, “παιδιά” της ετικέτας <science>, όπως <field>, <subfield> και <subsubfield> οι οποίες και αναπαραστούν τα επιμέρους επιστημονικά πεδία των επιστημών που αναφέρθηκαν. Με αυτόν τον τρόπο έχουμε δημιουργήσει ουσιαστικά για κάθε μέσο δημοσίευσης ένα μονοπάτι για το δένδρο επιστημών με το οποίο συνδέεται. Όπως αναφέρθηκε και παραπάνω το μονοπάτι ακολουθεί πιστά το ιεραρχικό δένδρο των επιστημών που δημιουργήσαμε, και μπορεί να φτάσει μέχρι και τέσσερα επίπεδα κατηγοριοποίησης (fourth level field classification) των επιστημονικών πεδίων (υποπεδίων). Για λόγους ευκολίας όμως εμείς θα αξιοποιήσουμε και θα οπτικοποιήσουμε μέχρι και το δεύτερο επίπεδο κατηγοριοποίησης (science και fields). Με αυτόν τον τρόπο επιτυγχάνουμε ένα αρκετά αξιόπιστο και ομοιόμορφο αποτέλεσμα, μιας και σε πολλές περιπτώσεις η πληροφορία που μας δόθηκε κρίθηκε ανεπαρκής για να προχωρήσουμε βαθύτερα στο δένδρο. Ωστόσο το μονοπάτι, όπου αυτό ήταν εφικτό, υπάρχει στο αρχείο μας και με κάποια ιδανικά δεδομένα εισόδου προφανώς θα καταλήγαμε και στην καλύτερη δυνατή οπτική αναπαράστασή τους.

Αφού λοιπόν έχουμε ολοκληρώσει επιτυχώς τις αντιστοιχίες μέσω δημοσίευσης με επιστήμες και επιστημονικά πεδία, αυτό που μένει είναι να ενώσουμε τον κάθε συγγραφέα με τα μέσα στα οποία δημοσίευσε και κατ' επέκταση τις επιστήμες ή επιστημονικά πεδία με τα οποία συσχετίζεται. Για κάθε συγγραφέα λοιπόν κρατάμε το συνολικό αριθμό των μέσων δημοσίευσης, που ταυτόχρονα αποτελεί και το σύνολο των δημοσιευμένων αναφορών του. Εν συνεχεία, αφού έχουμε κρατήσει λίστες με τα ονόματα των μέσων δημοσίευσης μπορούμε να ενώσουμε τους συγγραφείς με τις αντίστοιχες επιστήμες και να βρούμε συχνότητες για το πόσες φορές εμφανίστηκε στην κάθε μία από αυτές. Με το τρόπο αυτό παίρνουμε ένα ενδιάμεσο αποτέλεσμα το οποίο και θα οπτικοποιήσουμε στο επόμενο βήμα.

Ωστόσο ένα ζήτημα που προκύπτει είναι η ορθότητα του να έχουν την ίδια βαρύτητα οι αναφορές που έχουν δημοσιευτεί σε μέσο δημοσίευσης με αντίκτυπο σε μία μόνο επιστήμη, και οι αναφορές που το μέσο δημοσίευσής τους επηρεάζει περισσότερες από μία επιστήμες. Η πρότασή μας είναι να αλλάζει ο συντελεστής βαρύτητας για τις δύο αυτές περιπτώσεις. Δηλαδή, για τα έργα που επηρεάζουν μόνο έναν επιστημονικό κλάδο ο συντελεστής βαρύτητας θα είναι 1, ενώ για τα έργα που επηρεάζουν ταυτόχρονα πολλούς επιστημονικούς κλάδους θα μειώνεται στο 0.7, ελαφρώς μικρότερος για ένα πιο δίκαιο υπολογισμό. Ο λόγος που επιλέξαμε αυτούς τους συντελεστές είναι για να δείξουμε ότι τα έργα που έχουν μόνο έναν τομέα επιρροής έχουν μεγαλύτερο αντίκτυπο στο αντίστοιχο επιστημονικό πεδίο από έργα που επηρεάζουν παράλληλα πολλά επιστημονικά πεδία.

Στην περίπτωση που μία δημοσίευση έχει παραπάνω από έναν επιστημονικό κλάδο με τον οποίο και σχετίζεται, τότε είναι σωστό να πούμε ότι επηρεάζει εξίσου τον καθένα από αυτούς. Ο λόγος που καταλήξαμε σε αυτή την προσέγγιση είναι γιατί μόνο από τον τίτλο του μέσου δημοσίευσης, χωρίς την μελέτη του περιεχομένου του έργου, δεν είναι δυνατό να κρίνουμε σε ποιόν από τους συσχετιζόμενους κλάδους έχει μεγαλύτερη επιρροή. Ως εκ τούτου, οι αναφορές αυτού του τύπου διαμοιράζουν το βαθμό επιρροής τους σε ίσα μέρη όσος και ο αριθμός των επηρεαζόμενων επιστημονικών κλάδων.

Σύμφωνα με τις παραπάνω προσεγγίσεις, ο υπολογισμός της επιρροής κάθε ερευνητή ανά επιστημονικό πεδίο προκύπτει ως εξής:

1. Βαθμός επιρροής αναφορών με αντίκτυπο σε μία επιστήμη:

$$b = \text{αριθμός δημοσιεύσεων με ένα επιστημονικό πεδίο} * 1$$

2. Βαθμός επιρροής για αναφορές με πολλαπλό αντίκτυπο ανά επιστήμη:

$$b = \frac{\text{αριθμός δημοσιεύσεων με πολλαπλά επιστημονικά πεδία} * 0.7}{\text{πλήθος επιστημονικών πεδίων}}$$

Έπειτα το μόνο που απομένει είναι να αθροίσουμε τα επιμέρους αποτελέσματα για κάθε επιστήμη, τα οποία θα αποτελούν και τα δεδομένα που θα παίρνουν ως εισόδους τα γραφήματα στο κομμάτι της οπτικοποίησης. Με βάση αυτά θα υπολογίζεται το ύψος για την κάθε μπάρα του ιστογράμματος (histogram), το εμβαδό του κάθε τριγώνου στην αναπαράσταση περιοχών (area representation) και το μήκος των ευθύγραμμων τμημάτων στην αναπαράσταση συγκεκριμένων πεδίων ανά περιοχή (second level field classification).

Στο σημείο αυτό θα είχε ενδιαφέρον να αναφέρουμε κάποιες ακραίες περιπτώσεις και τον τρόπο με τον οποίο τις διαχειριστήκαμε.

Για τις αναφορές

Μιάς και η λίστα των δεδομένων μας έχει δημιουργηθεί δυναμικά από διαδικτυακό ιστότοπο σε μερικές αναφορές παρατηρήθηκαν προβλήματα στην κωδικοποίηση, πράγμα που καθιστούσε αδύνατη την κατανόηση και την ανάγνωσή τους. Οι αναφορές αυτές δεν λήφθηκαν υπόψιν.

Άλλο ένα σενάριο με το οποίο βρεθήκαμε αντιμέτωποι ήταν να απουσιάζει το όνομα των συγγραφέων από την αναφορά, που σημαίνει ότι δεν μας προσφέρει αναγνώριση του προσώπου που θα θέλαμε να εξετάσουμε, και έτσι τέτοιου είδους αναφορές δεν είχαν νόημα να προσμετρηθούν στα αξιοποιήσιμα δεδομένα μας.

Εξίσου σημαντικό ήταν και να υπάρχει το όνομα του μέσου δημοσίευσης στο οποίο δημοσιεύτηκε η αναφορά. Στις περιπτώσεις που δεν υπήρχε, προφανώς δεν είμασταν σε θέση να δημιουργήσουμε αντιστοιχίες με τις κατάλληλες επιστήμες στις οποίες και είχε αντίκτυπο και ως αποτέλεσμα η πληροφορία ήταν ανεπαρκής και δεν συνυπολογίστηκε μαζί με τις άλλες.

Για την αντιστοίχιση των επιστημών

Την ίδια αντιμετώπιση είχαν και οι αναφορές στις οποίες υπήρχε το μέσο δημοσίευσης αλλά το όνομά του δεν προσέδιδε την απαραίτητη πληροφορία για να αξιοποιηθεί. Για παράδειγμα, μπορούμε να αποφανθούμε ότι το μέσο δημοσίευσης με τίτλο “International Journal of Computer , 2011 - Springer” επηρεάζει τις φυσικές & θετικές επιστήμες και συγκεκριμένα τον κλάδο της πληροφορικής αλλά δεν μπορούμε να ισχυριστούμε το ίδιο και για το “Springer” (χωρίς κάποιο συγκεκριμένο τίτλο) μιας και φαίνεται να επηρεάζει όλους τους επιστημονικούς τομείς με αποτέλεσμα η πληροφορία που δίνει να είναι πολύ γενική (μη αξιοποιήσιμη).

Κεφάλαιο 6

Οπτικοποίηση

Για να είναι όμως τα αποτελέσματα της ανάλυσης των αναφορών ευανάγνωστα, προσιτά και κατανοητά πρέπει να τα οπτικοποιήσουμε πριν τα παρουσιάσουμε.

Μία πιο προσεκτική ανάλυση των καταχωρήσεων των παραπομπών, που βασίζονται στο περιεχόμενό τους, και συγκεκριμένα στα μέσα δημοσίευσης των παραπομπών, μπορεί να αποκαλύψει διαφορές στο εύρος και στο πεδίο εφαρμογής του αντικτύπου των αναλυτών. Όμως με τον υπολογισμό πολλαπλών μετρήσεων αντί του ενός ανά συγγραφέα, χάνεται μία σημαντική δύναμη των συμβατικών μετρήσεων των παραπομπών η οποία είναι να μπορεί κανείς να αξιολογήσει τις επιπτώσεις ενός ερευνητή με ένα μόνο αριθμό.

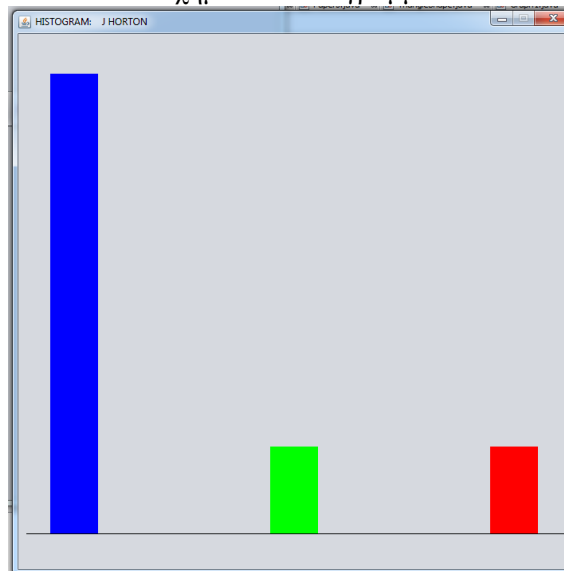
Για να ξεπεραστεί αυτό, θα χρησιμοποιήσουμε διαισθητικούς τρόπους για να απεικονίσουμε πολλές τιμές που υπολογίζουμε χρησιμοποιώντας ένα μόνο γράφημα. Έτσι, θα εξακολουθεί να είναι δυνατό να έχουμε μία σαφέστερη άποψη για την επιρροή του αναλυτή, με μία μόνο ματιά. Τυπικά, τέτοιοι κατάλογοι αριθμών θα εμφανίζονται χρησιμοποιώντας μπάρες. Όπως και ένα απλό συνηθισμένο γράφημα με μπάρες σχήμα 6.1. Αριστερά του σχήματος, η ύπαρξη μίας μόνο μπάρας δηλώνει την ύπαρξη ότι ο συγκεκριμένος συγγραφέας έχει αντίκτυπο σε μία μόνο επιστήμη. Αντίθετα, η δεξιά έκδοση του ιστογράμματος δείχνει την ύπαρξη πολλών επηρεασμένων επιστημών. Το μέγεθος της κάθε μπάρας προδίδει και το μέγεθος της επιρροής. Η μέθοδος αυτή μπορεί βέβαια να μεταφέρει επιτυχώς όλες τις πληροφορίες σε μία ενιαία εικόνα για την επιρροή ενός ή και περισσότερων ερευνητών, αλλά η αλήθεια είναι ότι κάποιος χρειάζεται να κοιτάξει προσεκτικά τις μπάρες και τους τίτλους αυτών για να προσδιορίσει τα πεδία που κάθε επιμέρους μπάρα σχετίζεται. Συνεπώς, η αναπαράσταση αυτή δεν είναι ιδιαίτερα έξυπνη, παρόλο που η χρήση χρωμάτων για τον καθορισμό επιστημονικών πεδίων μπορεί να την ενισχύσει.

Σχήμα 6.1: Ιστόγραμμα



Η δική μας προσέγγιση πάνω στο ιστόγραμμα, που αποτελεί και την πρώτη γραφική αναπαράσταση των δεδομένων μας, είναι ο διαχωρισμός του νοητού άξονα των x σε περιοχές και συνεπώς ο σχεδιασμός κάθε μπάρας στην αντίστοιχη περιοχή ανάλογα με την επιστήμη που αντιπροσωπεύει. Αναλυτικότερα, ο διαχωρισμός γίνεται σε πέντε περιοχές όσες και οι επιστήμες σύμφωνα με την ιεραρχία που δημιουργήσαμε και παρουσιάσαμε στο κεφάλαιο 4. Για την αναπαράσταση χρησιμοποιήθηκε μπλέ χρώμα για τις φυσικές & θετικές επιστήμες, κυανό για τις εφαρμοσμένες, πράσινο για τις κοινωνικές, κίτρινο για τις επιστήμες σύμπαντος και κόκκινο για τις ανθρωπιστικές. Η σειρά με την οποία εμφανίζονται είναι αντίστοιχη με τη σειρά που αναφέρθηκαν και τα χρώματά τους παραπάνω. Στο παρακάτω σχήμα 6.2 φαίνεται ότι υπάρχει επιρροή μόνο στις θετικές & φυσικές, στις κοινωνικές και τις ανθρωπιστικές επιστήμες. Οι μπάρες με κυανό και κίτρινο χρώμα απουσιάζουν γιατί δεν υπάρχει επιρροή στις εφαρμοσμένες επιστήμες και στις επιστήμες σύμπαντος αντίστοιχα.

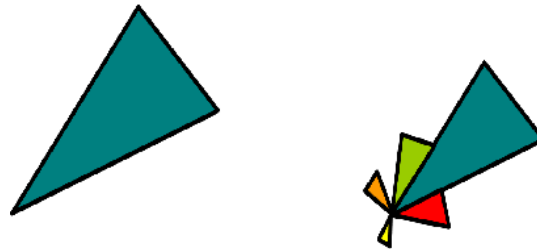
Σχήμα 6.2: Ιστόγραμμα



Ωστόσο, μία πιο έξυπνη προσέγγιση είναι να απεικονίσουμε τις τιμές όχι ως μήκος αλλά μάλλον σαν περιοχές. Η κατανομή της περιοχής μπορεί στη συνέχεια να χρησιμοποιηθεί για να απεικονίσει τις επιστήμες με τις οποίες σχετίζεται και κάθε κατεύθυνση ή χρώμα να συνδέεται και με έναν διαφορετικό κλάδο σχήμα 6.3. Με τον τρόπο αυτό επιτυγχάνουμε κάθε χρήστης που έχει

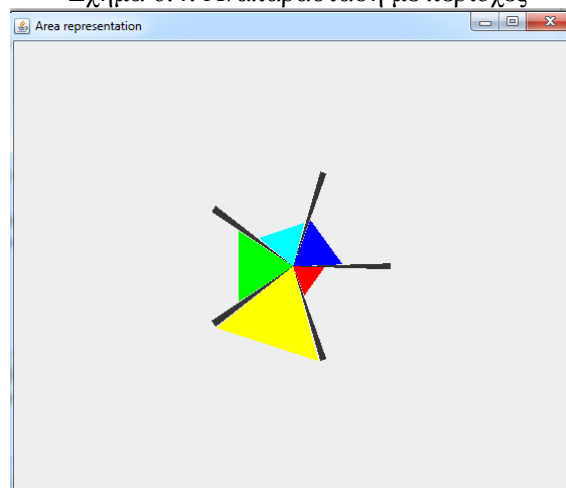
εξοικειωθεί με αυτή την απεικόνιση να είναι σε θέση να πάρει μία πολύ σαφή ιδέα για το νόημα του σχήματος με μία γρήγορη και ενιαία ματιά.

Σχήμα 6.3: Αναπαράσταση με περιοχές



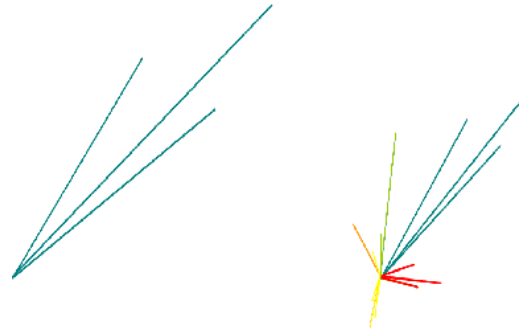
Ίδια λογική με αυτή του ιστογράμματος χρησιμοποιήσαμε και στην αναπαράσταση περιοχών. Οι επιστήμες και επιλογή των χρωμάτων είναι ίδια με πριν αλλά τώρα ο διαχωρισμός των περιοχών γίνεται ανάλογα με τις μοίρες. Σαφέστερα, από 0 έως 72 μοίρες είναι η περιοχή σχεδίασης για τις θετικές & φυσικές επιστήμες, από 72 έως 144 μοίρες είναι η περιοχή των εφαρμοσμένων επιστημών, από 144 έως 216 μοίρες η περιοχή των κοινωνικών επιστημών, από 216 με 288 μοίρες οι επιστήμες σύμπαντος και τέλος από 288 έως 360 μοίρες οι ανθρωπιστικές επιστήμες. Για κάθε μία από τις παραπάνω περιοχές διατίθεται ένα μικρό κομμάτι (μοιρών) για επιστήμες που είναι πιθανό να προστεθούν στο μέλλον (black areas). Το μέγεθος των τριγώνων που δημιουργούνται μέσα στις περιοχές είναι ανάλογο του βαθμού επιρροής που ασκεί ο συγγραφέας στην αντίστοιχη επιστήμη. Όσο μεγαλύτερο το τρίγωνο τόσο μεγαλύτερη η επιρροή του συγγραφέα. Στο σχήμα 6.4 φαίνονται όλες οι περιοχές και οι αναπαραστάσεις των επιστημών όπως αναφέρθηκαν προηγουμένως. Με μία γρήγορη ματιά μπορούμε να καταλάβουμε ότι η μεγαλύτερη επιρροή είναι στις επιστήμες σύμπαντος.

Σχήμα 6.4: Αναπαράσταση με περιοχές



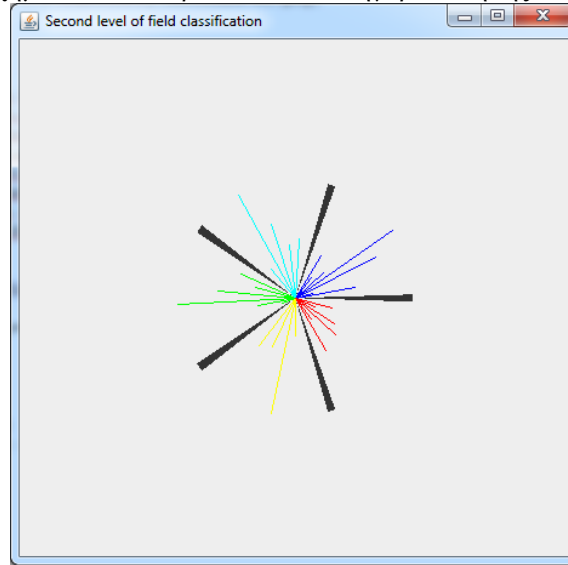
Ομοίως, περισσότερες πληροφορίες μπορούν να απεικονιστούν εάν η ανάλυση δεν είναι περιορισμένη, αλλά επεκτείνεται σε πιο λεπτομερείς ταξινομήσεις κλάσεων με αποτέλεσμα να καθιστά αυτή τη μέθοδο οπτικοποίησης προτιμότερη και πιο διαισθητική από μία απλή αριθμητική παράσταση σχήμα 6.5.

Σχήμα 6.5: Δεύτερο επίπεδο κατηγοριοποίησης πεδίων



Στην δική μας αναπαράσταση διατηρούμε τόσο το διαχωρισμό των περιοχών όσο και τα χρώματα που έχουμε επιλέξει για την κάθε επιστήμη και στα προηγούμενα μέσα οπτικοποίησης που χρησιμοποιήσαμε. Ωστόσο, εδώ ορίζουμε συγκεκριμένες μοίρες για την σχεδίαση των ευθύγραμμων τμημάτων, που αναπαριστούν τα επιμέρους επιστημονικά πεδία της κάθε επιστήμης, μέσα στις περιοχές που ήδη έχουν οριστεί από το σχήμα 6.4 της αναπαράστασης περιοχών (area representation). Έτσι με τον τρόπο αυτό προχωράμε ένα επίπεδο βαθύτερα στην ιεραρχία των επιστημονικών πεδίων. Τα χρώματα θα παραμένουν τα ίδια για ευθύγραμμα τμήματα που ανήκουν στην ίδια επιστήμη όπως φαίνεται στο σχήμα 6.6. Οι μαύρες περιοχές είναι για τη μελλοντική προσθήκη νέων επιστημονικών πεδίων όπως και στην περίπτωση του area representation. Ο αριθμός των ευθύγραμμων τμημάτων σε κάθε τομέα είναι όσος και ο αντίστοιχος αριθμός των επιστημονικών πεδίων του.

Σχήμα 6.6: Δεύτερο επίπεδο κατηγοριοποίησης πεδίων



6.1 Πλεονεκτήματα

Έχοντας λοιπόν παρουσιάσει τα μέσα οπτικοποίησης μπορούμε να συνοψίσουμε τα πλεονεκτήματα αυτού του τρόπου αναπαράστασης των δεδομένων.

Πρώτα απ' όλα, δεν θα χρειάζεται κάποιος να χρονοτριβεί διαβάζοντας λεζάντες και επεξηγήσεις για να κατανοήσει τα διαγράμματα και να πάρει όλη την πληροφορία που χρειάζεται. Η μόνη προϋπόθεση είναι να έχει εξοικειωθεί με αυτού του είδους την οπτικοποίηση. Με αυτόν τον τρόπο με μία απλή ματιά, γνωρίζοντας ήδη βέβαια σε ποιές περιοχές και χρώματα αντιστοιχίζεται η κάθε επιστήμη, μπορεί εύκολα κανείς να συμπεράνει από το χρώμα και την θέση του τριγώνου στο επίπεδο, σε ποιά επιστήμη έχει αντίκτυπο ο εξεταζόμενος συγγραφέας.

Επιπλέον, από το μέγεθος-εμβαδό του κάθε τριγώνου προκύπτει άμεσα με μία απλή ματιά, η επιστήμη στην οποία ασκείται περισσότερη επιρροή και η επιστήμη στην οποία ασκείται μικρότερη. Ακόμα και στην περίπτωση που είναι πολλές επιστήμες και πάλι χωρίς δυσκολία μπορούμε να ταξινομήσουμε τις επιστήμες εύκολα από την μικρότερη στην μεγαλύτερη επιρροή ή και αντίστροφα. Το εμβαδόν του τριγώνου είναι αυτό που προσδίδει με σχετική ακρίβεια στα ποσοστά κάνοντας φανερά ακόμα και τις μικρές διαφορές. Η μέθοδος αυτή μας απαλλάσσει από τη χρονοβόρα διαδικασία της προσεκτικής μελέτης των αυστηρών ποσοστών που αναγράφονται στην κάθε περιοχή.

Ακόμη, στην περίπτωση του σχήματος 6.6, με την αναπαράσταση των επιστημονικών πεδίων με ευθύγραμμα τμήματα, μπορεί κάποιος να εξάγει εύκολα συμπεράσματα για τον αριθμό των επιστημονικών πεδίων ανά επιστήμη που επηρεάζει. Ο αριθμός αυτός προφανώς είναι ίδιος με τον αριθμό των ευθύγραμμων τμημάτων. Η θέση του κάθε ευθύγραμμου τμήματος προδίδει και το όνομα του επιστημονικού πεδίου.

Τέλος, με αυτόν τον τρόπο αναπαράστασης των δεδομένων καθίσταται πιο εύκολη και η σύγκριση μεταξύ δύο συγγραφέων-ερευνητών. Παραθέτοντας δίπλα δίπλα τα οπτικοποιημένα αποτελέσματα για δύο εξεταζόμενους συγγραφείς μπορούμε να συμπαιράνουμε ποιός από τους δύο έχει μεγαλύτερο αντίκτυπο και πού. Ακόμα και στην περίπτωση που ο ένας έχει μόνο σε μία επιστήμη επιρροή και ο άλλος επηρεάζει πολλές επιστήμες ταυτόχρονα η σύγκριση είναι εφικτή χωρίς να οδηγεί σε λάθος συμπεράσματα. Αυτό συμβαίνει γιατί έχουμε ήδη υπολογίσει με διαφορετικούς συντελεστές τα έργα με πολλαπλό αντίκτυπο και έχουμε λάβει υπόψιν και τον όγκο των εξεταζόμενων αναφορών για τον κάθε συγγραφέα-ερευνητή.

Κεφάλαιο 7

Υλοποίηση

Για την υλοποίηση αυτού του έργου χρησιμοποιήθηκε η γλώσσα προγραμματισμού java. Με διάφορους υπολογισμούς και κρατώντας τα απαραίτητα δεδομένα σε πίνακες υπολογίσαμε τους βαθμούς επιρροής με τους κατάλληλους συντελεστές σε κάθε περίπτωση. Δημιουργήσαμε φόρμες για τη διευκόλυνση του χρήστη, και κυρίως για τα κομμάτια της εισαγωγής κάποιου νέου μέσου δημοσίευσης, και για την εύρεση του κάθε συγγραφέα που συμμετάσχει σε αυτά και υπάρχει στο αρχείο xml που μας δόθηκε.

7.1 Αρχιτεκτονική

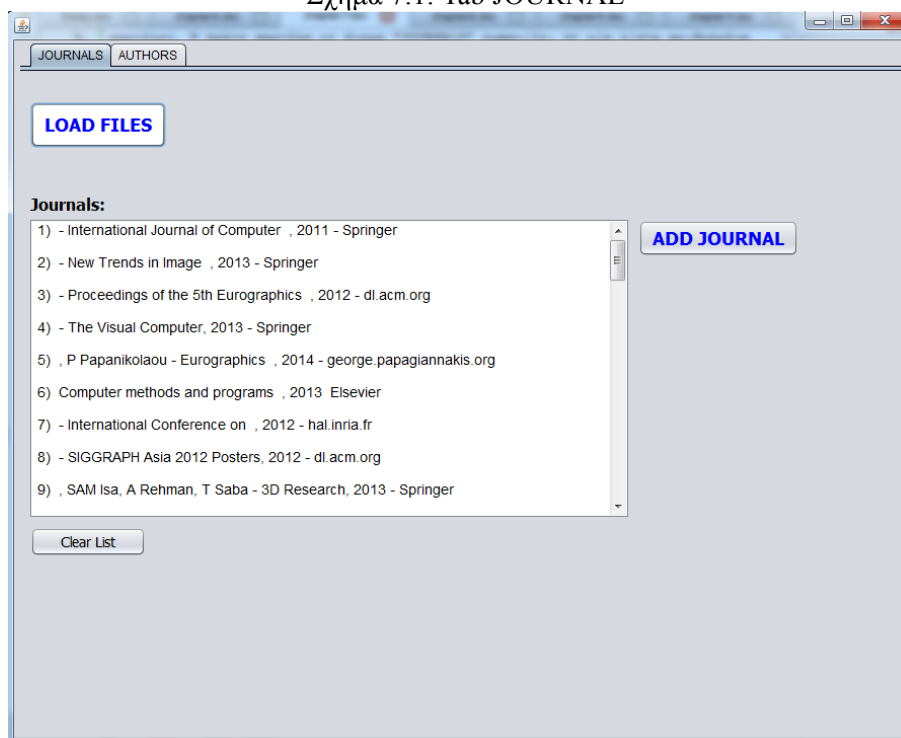
Η βασική ιδέα της υλοποίησης είναι να διαβάζουμε το αρχείο papers.xml που μας δόθηκε και να κρατάμε μοναδικά τα ονόματα των συγγραφέων που εντοπίζονται. Εν συνεχεία, θα κρατάμε για κάθε συγγραφέα τη λίστα με τα μέσα δημοσίευσης στα οποία και εμφανίστηκε. Έτσι θα παίρνουμε ένα ενδιάμεσο αποτέλεσμα για να μπορούμε έπειτα να συνδέσουμε τα μέσα δημοσίευσης με τα αντίστοιχα επιστημονικά πεδία. Η ανάγνωση των δύο αυτών αρχείων xml που εμπεριέχουν τις απαραίτητες πληροφορίες γίνεται με XML Parsing using DOM. Τα επιμέρους βήματα που ακολούθησαμε για να γίνει αυτό παρουσιάζονται συνοπτικά παρακάτω:

1. Βήμα 1: Δημιουργία του Parser
2. Βήμα 2: Parsing στο αρχείο xml που θέλουμε
3. Βήμα 3: Ανάγνωση κόμβων και διάβασμα του root κόμβου παίρνοντας όλα τα στοιχεία (elements).
4. Βήμα 4: Διάσχιση παιδιών (child).

Τρέχοντας την εφαρμογή ο χρήστης θα ανοίγει αυτόματα ένα παράθυρο με δύο καρτέλες. Η πρώτη καρτέλα εικόνα 7.1 με όνομα “JOURNALS” εμφανίζει σε μία λίστα αριθμημένα όλα τα μέσα δημοσίευσης που διαβάστηκαν και βρέθηκαν από το αρχείο που μας δόθηκε και για κάθε αντικείμενο journal που εντοπίζεται προστίθεται στη λίστα το όνομά του. Δεξιά του παραθύρου υπάρχει ένα κουμπί με όνομα “ADD JOURNAL” με το πάτημα που οποίου ανοίγει ένα νέο παράθυρο με τη φόρμα συμπλήρωσης κάποιου νέου μέσου δημοσίευσης. Έτσι δίνεται η δυνατότητα στον χρήστη να εισάγει κάποιο νέο μέσο δημοσίευσης το οποίο δεν υπάρχει στη λίστα και να το αντιστοιχίσει με τα κατάλληλα επιστημονικά πεδία με τα οποία σχετίζεται και ασκεί επιρροή. Θεωρούμε ότι ο χρήστης έχει τις απαραίτητες γνώσεις και εμπειρίες για να κάνει αυτή την προσθήκη

σωστά. Εάν η εισαγωγή είναι επιτυχημένη ο χρήστης θα ενημερωθεί με κατάλληλο μήνυμα σε νέο αναδυόμενο παράθυρο (pop up window) εικόνα 7.2 και η εγγραφή θα προστεθεί στο αρχείο JournalScienceMatch.xml. Για να γίνει αυτό ο χρήστης θα πρέπει να έχει συμπληρώσει όλα τα πεδία της φόρμας που έχουν σημειωθεί ως απαραίτητα (*required fields) και να πατήσει το κουμπί με όνομα “SUBMIT” στο τέλος της φόρμας συμπλήρωσης. Τα πεδία εμφανίζονται με της εξής σειρά όπως φαίνονται και στην εικόνα. Πρώτο είναι το πεδίο με το όνομα του μέσου δημοσίευσης (απαραίτητο), δεύτερο το πεδίο για την αντιστοίχιση της επιστήμης (φυσικές & θετικές, εφαρμοσμένες, κοινωνικές, σύμπαντος και ανθρωπιστικές, επίσης απαραίτητο), τρίτο είναι το πεδίο συμπλήρωσης του επιστημονικού κλάδου της κάθε επιστήμης (απαραίτητο), τέταρτο το πεδίο για τυχόν ύπαρξη υποκλάδου του επιστημονικού κλάδου που αναφέρθηκε νωρίτερα, και τέλος το πεδίο για τυχόν (υπο)υποκλάδο του υποκλάδου της επιστήμης. Οι επιστήμες και τα επιστημονικά πεδία εμφανίζονται σε λίστες και η προηγούμενη επιλογή λειτουργεί ως φίλτρο για την επόμενη. Με τον τρόπο αυτό επιτυγχάνεται ένας μικροέλεγχος για την εισαγωγή του μέσου δημοσίευσης. Γυρίζοντας πίσω στην καρτέλα “JOURNALS” παρατηρούμε ότι υπάρχουν άλλα δύο κουμπιά με όνοματα “LOAD FILES” και “clear list”. Όπως προδίδει και το όνομά τους το δεύτερο επιτρέπει στον χρήστη να καθαρίσει τη λίστα με τα ονόματα των μέσων δημοσίευσης ενώ από την άλλη το “LOAD FILES” ξαναφορτώνει τα αρχεία και συνεπώς εμφανίζει ξανά τη λίστα.

Σχήμα 7.1: Tab JOURNAL



Σχήμα 7.2: Add Journal

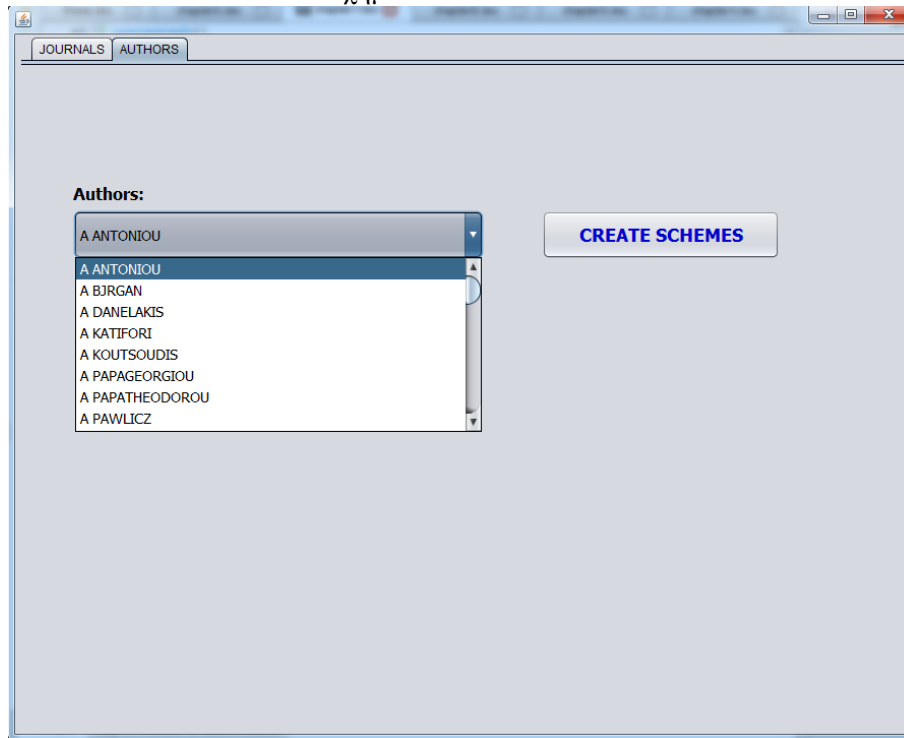
The screenshot shows a web form titled "ADD JOURNAL". It contains the following fields:

- Journal:** An empty text input field, marked as a required field.
- Science:** A dropdown menu with options: social science, applied science, universe science, natural science, and human science. Marked as a required field.
- Field:** A dropdown menu with options: life science, mathematics, statistics, and computer science. Marked as a required field.
- SubField:** A dropdown menu with options: information systems, computing methodologies, computer applications, and computing milieux.
- SubSubField:** A dropdown menu with options: symbolic and algebraic manipulation, artificial intelligence, computer graphics, image processing and computer vision, pattern recognition, and simulation and modeling.

A blue "SUBMIT" button is located at the bottom left of the form.

Στη δεύτερη καρτέλα του παραθύρου εικόνα 7.3 με όνομα "AUTHORS" αριστερά εμφανίζεται ένα combobox που περιέχει τη λίστα με όλα τα ονόματα των συγγραφέων που βρέθηκαν διαβάζοντας το αρχείο papers.xml. Τα ονόματα εμφανίζονται με αλφαβητική σειρά και ο χρήστης μπορεί να περιηγηθεί στη λίστα πατώντας το βελάκι και χρησιμοποιώντας το scroll bar. Στη συνέχεια μπορεί να επιλέξει τον συγγραφέα-ερευνητή για τον οποίο θέλει να μελετήσει τα στατιστικά του και να πατήσει το κουμπί "CREATE SCHEMES" για να εμφανιστούν τα γραφήματα.

Σχήμα 7.3: Tab Authors



7.2 Λίστες

Το πρόγραμμα παίρνει τρεις λίστες ως είσοδο σε αρχεία xml. Η πρώτη από αυτές (papers.xml), σχήμα 7.4 είναι η λίστα με τις αναφορές που έχουν δημοσιευτεί και σχετικές πληροφορίες σχετικά με αυτές όπως ο τίτλος της αναφοράς, τα ονόματα των συγγραφέων, ο τίτλος του μέσου δημοσίευσης και το έτος δημοσίευσης. Οι πληροφορίες αυτές περικλείονται μέσα στις κατάλληλες ετικέτες (tags). Κάθε μία δημοσιευμένη αναφορά ορίζεται μέσα στην ετικέτα <paper>. Η ετικέτα <title> δίνει τον τίτλο της αναφοράς, η <author> το όνομα του συγγραφέα (ένας συγγραφέας τη φορά), η <journal> το τίτλο του μέσου δημοσίευσης και η <year> το έτος δημοσίευσης της αναφοράς. Δεν είναι απαραίτητο να υπάρχουν συμπληρωμένες όλες η ετικέτες για κάποια αναφορά. Στην περίπτωση αυτή πρέπει να εξετάσουμε αν η ετικέτα που λείπει προσδίδει σημαντική πληροφορία ή όχι. Για παράδειγμα, αν λείπει το <year>, το οποίο δεν συνεισφέρει στον υπολογισμό των στατιστικών μας, δεν τίθεται κάποιο πρόβλημα. Αντίθετα, εάν λείπουν οι ετικέτες με τους συγγραφείς τότε δεν έχει νόημα να υπολογίσουμε βαθμούς επιρροής γιατί δεν θα ξέρουμε το άτομο που επηρεάζει την κάθε επιστήμη. Επιπρόσθετα, είναι απαραίτητο να υπάρχει το πεδίο <journal> που εμπεριέχει το μέσο δημοσίευσης στο οποίο και στηριζόμαστε, για να αποφανθούμε για το με ποιά επιστήμη συσχετίζεται. Στην περίπτωση λοιπόν που απουσιάζει κάποια από τις παραπάνω πληροφορίες που κρίθηκαν σημαντικές, τότε η αναφορά δεν λαμβάνεται υπόψιν και κατά συνέπεια δεν

προσμετράται στον υπολογισμό του βαθμού επιρροής. Το ίδιο συμβαίνει και με τις περιπτώσεις όπου τα ονόματα των συγγραφέων ή τα ονόματα των μέσων δημοσίευσης δεν είχαν τη σωστή κωδικοποίηση και ως εκ τούτου δεν ήμασταν σε θέση να τα κατανοήσουμε για να εξάγουμε χρήσιμες πληροφορίες.

Σχήμα 7.4: papers.xml

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <research>
3  <full>
4  <paper>
5  <title>Graphics and visualization: Principles &amp; algorithms</title>
6  <author>T Theoharis</author>
7  <author>G Pasaiouannou</author>
8  <author>N Elatis</author>
9  <author>NM Patrikalakis</author>
10 <journal>CRC Press</journal>
11 <year>2008</year>
12 </paper>
13 <paper>
14 <paper>
15 <title>PANORAMA: A 3D shape descriptor based on panoramic views for unsupervised 3D object retrieval</title>
16 <author>P Papadakis</author>
17 <author>I Pratikakis</author>
18 <author>T Theoharis</author>
19 <author> </author>
20 <journal> - International Journal of , 2010 - Springer</journal>
21 </paper>
22 <paper>
23 <title>Automatic 3D Facial Region Retrieval from Multi-pose Facial Datasets.</title>
24 <author>P Perakis</author>
25 <author>T Theoharis</author>
26 <author>G Passalis</author>
27 <author>IA Kakadiaris</author>
28 <author> </author>
29 <journal> - 3DOR, 2009 - researchgate.net</journal>
30 </paper>
31 <paper>
32 <title>Rosy+: 3d object pose normalization based on pga and reflective object symmetry with application in 3d object retrieval</title>
33 <author>K Sfikas</author>
34 <author>T Theoharis</author>
35 <author>I Pratikakis</author>
36 <author> </author>
37 <journal> - International Journal of Computer , 2011 - Springer</journal>
38 </paper>

```

Η δεύτερη λίστα που χρησιμοποιήσαμε ως είσοδο βρίσκεται στο αρχείο JournalScienceMatch.xml, σχήμα 7.5 το οποίο περιέχει όλες τις αντιστοιχίες των μέσων δημοσίευσης, που βρέθηκαν στο αρχείο papers.xml, με τις κατάλληλες επιστήμες και επιστημονικά πεδία. Αναλυτικότερα, κάθε ετικέτα <journal> περικλύει τον τίτλο (<title>) του μέσου δημοσίευσης και το <science> που δηλώνει την ευρύτερη επιστήμη που επηρεάζει, όπως αυτές έχουν καθοριστεί από την ιεραρχία μας. Έπειτα, εμφανίζονται σα παιδιά του <science> οι αντίστοιχες κατηγορίες για τα επιστημονικά πεδία (<fields>, <subfields>, <subsubfields>) της κάθε επιστήμης. Τα ονόματα του κάθε επιπέδου σημειώνονται με την προσθήκη γνωρισμάτων (attributes) στην κάθε ετικέτα (tag). Στη συνέχεια διαβάζουμε το αρχείο με βάση το γνώρισμα “category” και αντίστοιχα παίρνουμε τα παιδιά, έπειτα τα παιδιά των παιδιών (εάν υπάρχουν) και ούτω καθεξής (XML Parsing). Έτσι δημιουργείται ένα δέντρο το οποίο ορίζει το μονοπάτι για τα χαμηλότερα και βαθύτερα επίπεδα της ιεραρχίας. Εάν υπάρχουν και τα τρία πεδία που αναφέρθηκαν παραπάνω τότε το μονοπάτι είναι πλήρης και έχουμε την καλύτερη δυνατή πληροφορία. Σε άλλη περίπτωση, μιάς και τα σχήματα δημιουργούνται μόνο με την αξιοποίηση των ετικετών <science> και <field> δημιουργείται δέντρο, αλλά δεν μπορούμε να παράξουμε όλο το μονοπάτι μέχρι τα τελικά φύλλα, αλλά αυτό δεν έχει κάποιο αντίκτυπο στο έργο μας. Απαραίτητη προϋπόθεση λοιπόν είναι να υπάρχουν τουλάχιστον αυτά τα δύο πεδία για να οδηγηθούμε σε σωστά αποτελέσματα και τελικά να σχεδιαστούν τα γραφήματά μας.

Σχήμα 7.5: JournalScienceMatch.xml

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <research>
3  <journal>
4  <title> - International Journal of Computer , 2011 - Springer</title>
5  <science category="natural science">
6  <field category="computer science">
7  </field>
8  </science>
9  </journal>
10
11 <journal>
12 <title> - New Trends in Image , 2013 - Springer</title>
13 <science category="natural science">
14 <field category="computer science">
15 <subfield category="computing methodologies">
16 <subsubfield category="image processing">
17 </subsubfield>
18 </subfield>
19 </field>
20 </science>
21 </journal>
22
23 <journal>
24 <title> - Proceedings of the 5th Eurographics , 2012 - dl.acm.org</title>
25 <science category="natural science">
26 <field category="computer science">
27 <subfield category="computing methodologies">
28 </subfield>
29 </field>
30 </science>
31 </journal>
32
33 <journal>
34 <title> - The Visual Computer, 2013 - Springer</title>
35 <science category="natural science">
36 <field category="computer science">
37 <subfield category="computing methodologies">
38 <subsubfield category="computer graphics">
39 </subsubfield>
40 </subfield>
41 </field>
42 </science>

```

Η τρίτη λίστα, αρχείο ScienceTree.xml σχήμα 7.6, αναπαριστά το δένδρο των επιστημών και χρησιμοποιήθηκε για τη συμπλήρωση των αντίστοιχων πεδίων με xml parsing για τη φόρμα συμπλήρωσης προσθήκης νέου μέσου δημοσίευσης. Έτσι για την επιλογή natural science κρατάμε όλα τα επιστημονικά πεδία με τα οποία και γεμίζουμε τη λίστα στο πεδίο field. Αντίστοιχα, για το επιλεγμένο field του natural science κρατάμε όλα τα υποπεδία του και τα τοποθετούμε στη λίστα του subfield. Η προηγούμενη επιλογή χρησιμοποιείται ως φίλτρο για την επόμενη.

Σχήμα 7.6: ScienceTree.xml

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <tree>
3  <science category="natural science">
4    <field category="physics">
5      <subfield category="applied physics"> </subfield>
6      <subfield category="cosmology"> </subfield>
7      <subfield category="astronomy">
8        <subsubfield category="astrophysics"> </subsubfield>
9      </subfield>
10     <subfield category="biophysics"> </subfield>
11     <subfield category="geophysics"> </subfield>
12   </field>
13   <field category="chemistry">
14     <subfield category="biochemistry"> </subfield>
15     <subfield category="geochemistry"> </subfield>
16     <subfield category="physicochemical"> </subfield>
17   </field>
18   <field category="earth and planetary science">
19     <subfield category="geology">
20       <subsubfield category="volcanology"> </subsubfield>
21       <subsubfield category="seismology"> </subsubfield>
22       <subsubfield category="mineralogy"> </subsubfield>
23       <subsubfield category="paleontology"> </subsubfield>
24     </subfield>
25     <subfield category="meteorology"> </subfield>
26     <subfield category="climatology"> </subfield>
27     <subfield category="oceanology"> </subfield>
28   </field>
29   <field category="life science">
30     <subfield category="botanic"> </subfield>
31     <subfield category="zoology"> </subfield>
32     <subfield category="ecology"> </subfield>
33     <subfield category="genetics"> </subfield>
34     <subfield category="biotechnology"> </subfield>
35     <subfield category="molecular biology"> </subfield>
36     <subfield category="physiology"> </subfield>
37   </field>
38   <field category="mathematics">
39     <subfield category="general/foundations (mathematical logic)"> </subfield>
40     <subfield category="discrete mathematics/algebra"> </subfield>
41     <subfield category="geometry and topology"> </subfield>

```

7.3 Αποτελέσματα

Όπως εξηγήσαμε και στο κεφάλαιο 6, τα αποτελέσματά μας για την επιρροή του ασκεί ο κάθε συγγραφέας θα εμφανίζονται σε τρία γραφήματα. Ένα ιστόγραμμα, ένα γράφημα με αναπαράσταση περιοχών και ένα γράφημα με αναπαράσταση δεύτερου επιπέδου πεδίων με χρήση ευθύγραμμων τμημάτων.

Ιστόγραμμα (Histogram)

Για την δημιουργία του ιστογράμματος αρχικά καταμετρήσαμε τις συχνότητες της κάθε κατηγορίας που θα εμφανίζεται. Έπειτα κρατήσαμε σε ένα πίνακα τα ονόματα των κατηγοριών για να μπορούμε να κάνουμε τις σωστές αντιστοιχίες με χρώματα θέσεις και συχνότητες. Σε όποια κατηγορία η συχνότητα, δηλαδή ο βαθμός επιρροής βρεθεί μηδέν τότε η αντίστοιχη μπάρα δεν σχεδιάζεται δηλώνοντας έτσι ότι δεν έχει επιρροή στη επιστήμη που θα αντιπροσώπευε η συγκεκριμένη μπάρα. Για τις μπάρες που αντιστοιχίζονται με κάποιο βαθμό επιρροής υπολογίζουμε το πλάτος της μπάρας που θέλουμε να δημιουργήσουμε και το μέγεθος του panel. Στη συνέχεια

βρίσκουμε το μέγιστο βαθμό επιρροής και τον αντιστοιχίζουμε με το μέγιστο δυνατό ύψος που θέλουμε να δώσουμε στην μπάρα. Καθορίζουμε τη θέση που θέλουμε να εμφανίζεται στον άξονα των x η πρώτη μπάρα και τέλος γεμίζουμε τις μπάρες με κατάλληλα χρώματα και στο κατάλληλο μέρος ανάλογα με το όνομα της επιστήμης που αντιπροσωπεύουν.

Αναπαράσταση με περιοχές (Area representation)

Για τη δημιουργία της αναπαράστασης περιοχών (area representation) αρχικά γεμίζουμε ένα πίνακα με τα χρώματα τα οποία θέλουμε να αντιστοιχίσουμε την κάθε επιστήμη ανάλογα με τη σειρά που θέλουμε να εμφανίζονται. Οι σειρά είναι η ίδια που ακολουθήσαμε και στο ιστόγραμμα. Σε ένα νοητό άξονα x και ξεκινώντας από 0 μοίρες θα εμφανίζονται πρώτα οι φυσικές & θετικές επιστήμες με μπλε χρώμα, οι εφαρμοσμένες με κυανό, οι κοινωνικές με πράσινο, οι επιστήμες σύμπαντος με κίτρινο και οι ανθρωπιστικές με κόκκινο. Στη συνέχεια θα παίρνουμε σε ένα πίνακα το βαθμό επιρροής που έχουμε υπολογίσει για την κάθε επιστήμη και θα βρίσκουμε τη μέγιστη τιμή. Ξεκινάμε τη σχεδίαση των τριγώνων ξεκινώντας από τις 3 μοίρες γιατί έχουμε αφήσει ένα περιθώριο 2 μοιρών σε κάθε περιοχή για μελλοντικές επιστήμες. Αφού έχουμε καταλήξει σε πέντε επιστήμες χρειαζόμαστε τη σχεδίαση πέντε τριγώνων τα οποία θα εμφανίζονται ανά 72 μοίρες. Οι μαύρες περιοχές και τα τρίγωνα σχεδιάζονται και γεμίζουν με χρώματα με τον ίδιο τρόπο. Κρατάμε το ένα σημείο σταθερό και υπολογίζουμε τα άλλα δύο (startx και starty) με μαθηματικές συναρτήσεις που χρησιμοποιούν ημίτονα και συνημίτονα σημειώνοντας εδώ ότι οι άξονες x και y ξεκινούν από την πάνω αριστερή γωνία του παραθύρου. Τα τρίγωνα γεμίζουν με τα κατάλληλα χρώματα που δηλώνουν την επιστήμη και όπως και στην περίπτωση του ιστογράμματος για επιστήμες που δεν επηρεάζονται δεν δημιουργούνται τρίγωνα.

Δεύτερο επίπεδο κατηγοριοποίησης πεδίων (Second level of field classification)

Παρόμοιο τρόπο υλοποίησης ακολουθήσαμε και στην αναπαράσταση περιοχών δεύτερου επιπέδου (second level field classification). Διαφορά εδώ είναι ότι για κάθε περιοχή (area), όπως αυτές ορίστηκαν στο προηγούμενο γράφημα, χρησιμοποιείται νέος πίνακας (σύνολο πέντε διαφορετικοί πίνακες) που παίρνει το βαθμό επιρροής για κάθε επιμέρους επιστημονικό πεδίο κάθε επιστήμης. Έπειτα βρίσκουμε τον αριθμό των δυνατών γραμμών σε κάθε περιοχή ώστε να τοποθετήσουμε την κάθε μία με τρόπο τέτοιο ώστε να ισαπέχουν μεταξύ τους. Η υπόλοιπη διαδικασία είναι ίδια με πριν και τα χρώματα των γραμμών διατηρούνται ανά επιστημονική περιοχή.

Κεφάλαιο 8

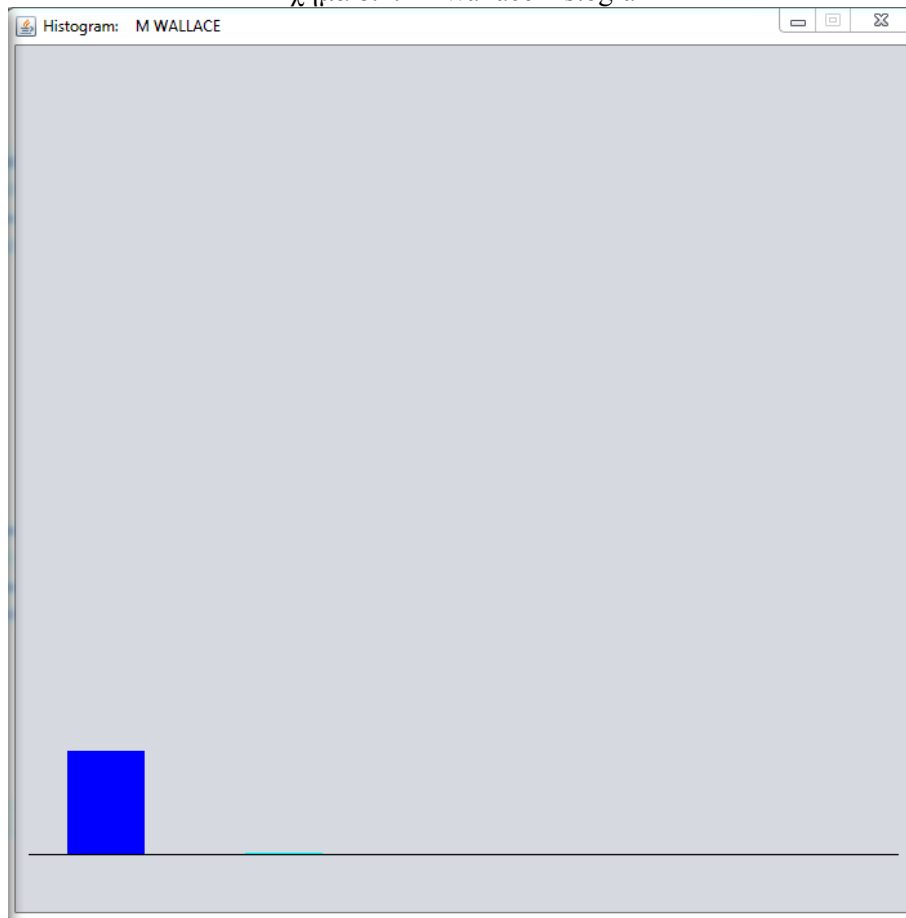
Παραδείγματα

Για να γίνει περισσότερο κατανοητός ο σκοπός και η λειτουργία της εφαρμογής θα παρουσιάσουμε μερικά παραδείγματα.

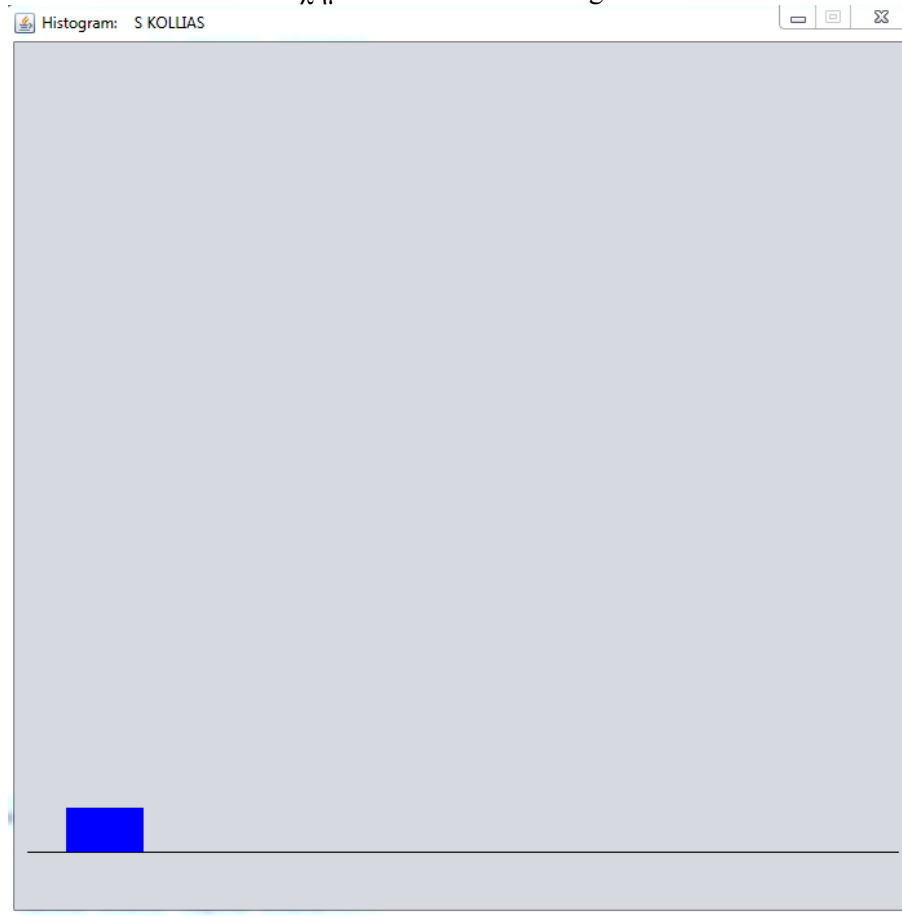
Νωρίτερα, αναφέραμε πως ένα σημαντικό πλεονέκτημα της εφαρμογής αυτής είναι η δυνατότητα εύκολης σύγκρισης βαθμών επιρροής μεταξύ δύο ερευνητών με μία απλή ματιά. Για παράδειγμα, ας δούμε τα αποτελέσματα των M Wallace και S Kollias.

Στην περίπτωση του ιστογράμματος σχήματα 8.1 και 8.2 μπορούμε να διακρίνουμε εύκολα ότι και οι δύο ερευνητές έχουν καθαρή επιρροή στις φυσικές & θετικές επιστήμες. Αυτό φαίνεται αφού και στις δύο περιπτώσεις η μπάρα είναι μπλέ και εμφανίζεται στην αρχή του άξονα των x. Η θέση και το χρώμα αυτό όπως έχουμε αναφέρει αντιστοιχούν στις φυσικές & θετικές επιστήμες. Το ύψος της μπάρας όμως για τον κ. M Wallace διακρίνεται να είναι μεγαλύτερο από του κ. S Kollia, πράγμα που σημαίνει ότι ο πρώτος έχει περισσότερες αναφορές από τον δεύτερο ή πιο σωστά ότι για τον πρώτο ερευνητή οι αναφορές που λήφθηκαν υπόψιν από αυτές που μας δόθηκαν ήταν περισσότερες από ότι για τον δεύτερο. Συγκεκριμένα, αξιοποιήθηκαν 84 αναφορές για τον κ. M Wallace και 36 για τον κ. S Kollia. Έτσι ο κ. Wallace όπως φαίνεται έχει μεγαλύτερη επιρροή (περισσότερες αναφορές) στις φυσικές & θετικές επιστήμες από ότι ο κ. S Kollias. Εάν ρίχναμε μία πιο προσεκτική ματιά στο σχήμα 8.1 μία πολύ λεπτή γραμμή στο σημείο που θα εμφανιζόταν η κυανή μπάρα που θα αναπαραστούσε τις εφαρμοσμένες επιστήμες (δεύτερη σε σειρά μπάρα). Το ύψος της μπάρας όμως είναι τόσο μικρό που αυτή δύσκολα διακρίνεται. Αυτό συμβαίνει γιατί οι αναφορές με επιρροή στις εφαρμοσμένες επιστήμες ήταν μόνο τρεις (όπως φαίνεται στο output της εφαρμογής σχήμα 8.3). Αντίστοιχα, για τον κ. S Kollia βρίσκεται μόνο μία αναφορά στις εφαρμοσμένες επιστήμες και για το λόγο αυτό η επιρροή του εκεί είναι ανύπαρκτη.

Σχήμα 8.1: M Wallace histogram



Σχήμα 8.2: S Kollias histogram

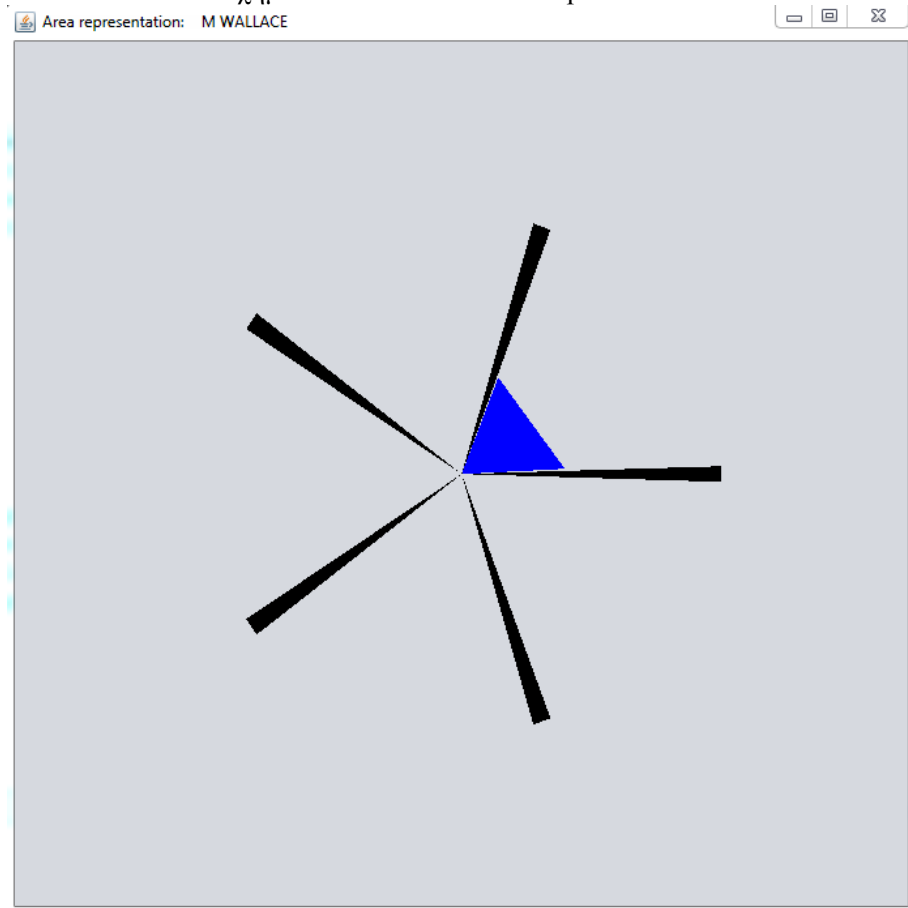


Σχήμα 8.3: Output

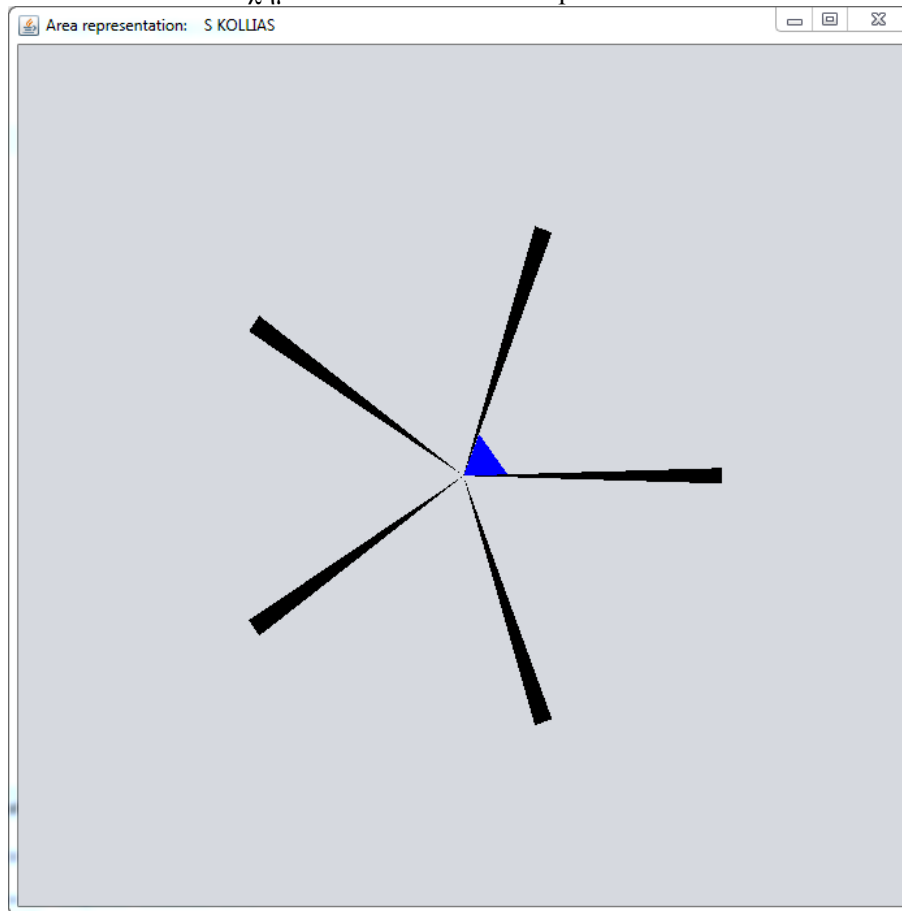
```
Output - VisualizationPtyxiaki (run) ✖
run:
author:M WALLACE
FIELDS
[natural science, computer science]=81.39999999999999
[applied science, health science]=3.0
author:S KOLLIAS
FIELDS
[natural science, computer science]=35.0
[applied science, health science]=1.0
BUILD SUCCESSFUL (total time: 44 seconds)
```

Τα αποτελέσματα που αναφέραμε παραπάνω, αναπαριστώνται και στην περίπτωση του area representation. Έτσι βλέποντας μόνο τα σχήματα 8.4 και 8.5 μπορεί κανείς να συμπεράνει και πάλι ότι ο κ. M Wallace έχει περισσότερες αναφορές και κατά συνέπεια και μεγαλύτερη επιρροή στον τομέα των φυσικών & θετικών επιστημών. Ο τομέας αυτός όπως έχουμε ορίσει αναπαριστάται μεταξύ των 3 και 72 μοιρών και το τρίγωνο είναι μπλε χρώματος. Εδώ όπως και παραπάνω, το κυανό τρίγωνο για τις εφαρμοσμένες επιστήμες (περιοχή 75 με 147 μοιρών) είναι εξαιρετικά δύσκολο να διακριθεί μιας και είναι σχεδόν ανύπαρκτο.

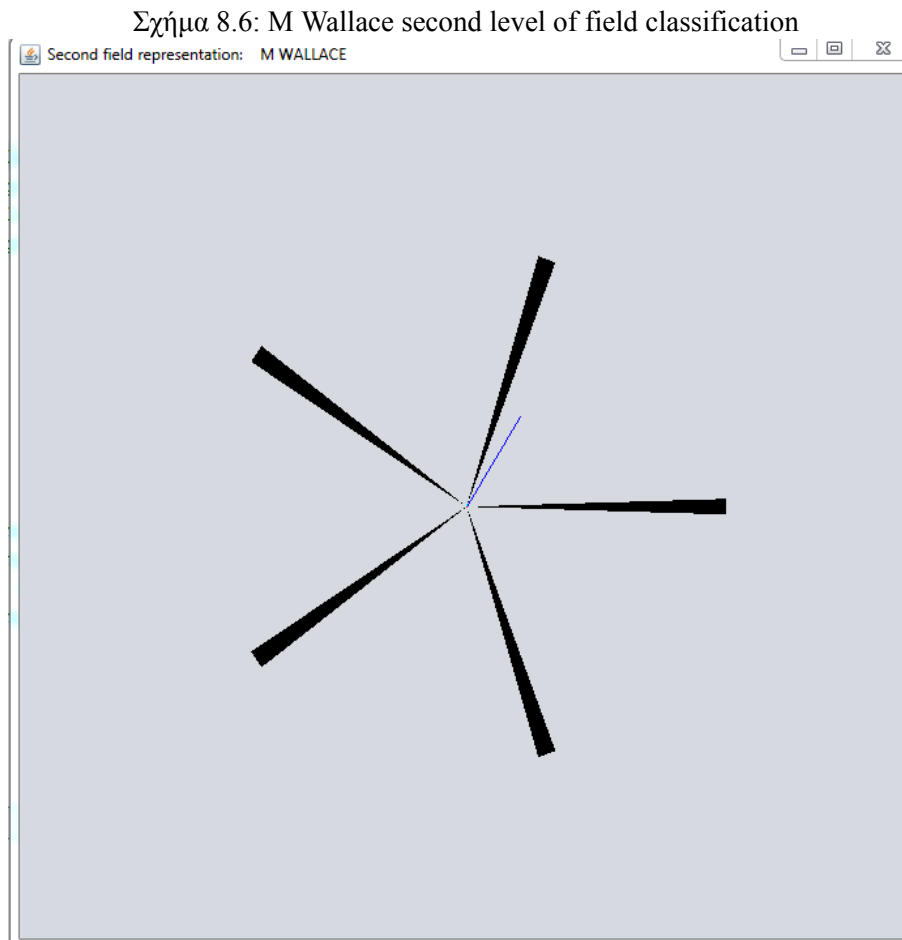
Σχήμα 8.4: M Wallace area representation



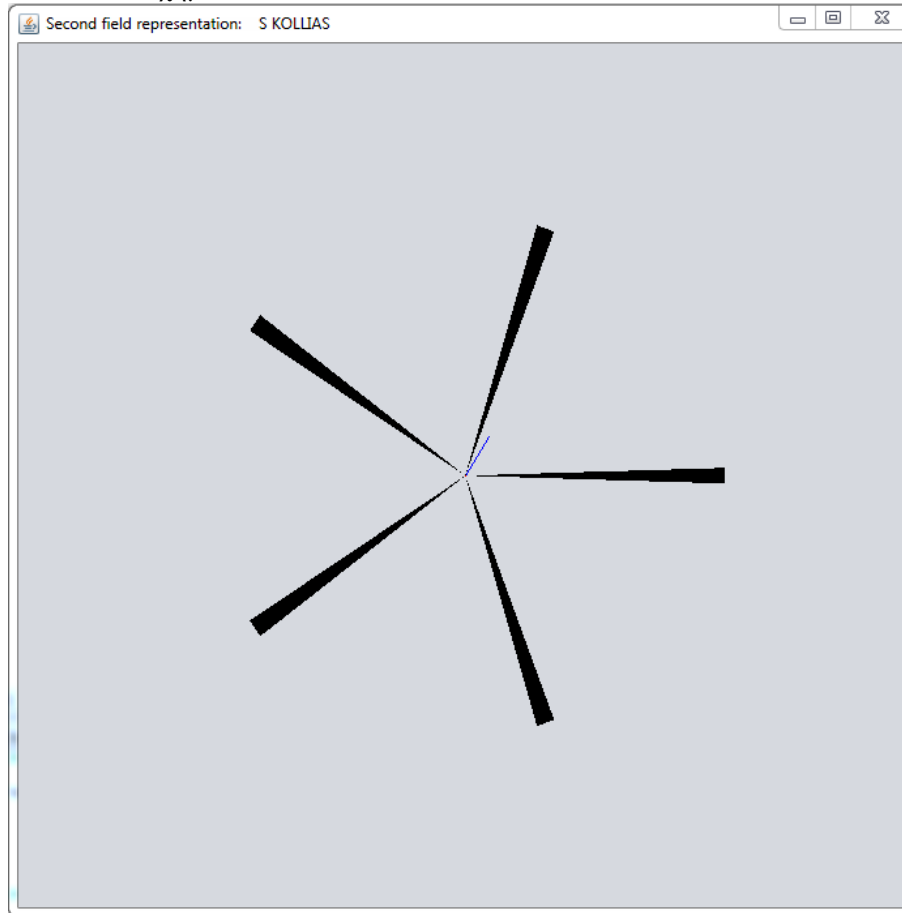
Σχήμα 8.5: S Kollias area representation



Τέλος, στα σχήματα 8.6 και 8.7, του second level of field classification μπορούμε να διακρίνουμε και πάλι ότι ο κ. M Wallace έχει μεγαλύτερη επιρροή στον τελευταίο επιστημονικό κλάδο των φυσικών & θετικών επιστημών, που όπως έχουμε ορίσει είναι η επιστήμη των υπολογιστών (computer science) από τον κ. S Kollia. Το μήκος του ευθύγραμμου τμήματος δείχνει το μέγεθος της επιρροής και η θέση της στη συγκεκριμένη περιοχή προδίδει τον επιστημονικό κλάδο (το πιο κοντινό σε απόσταση ευθύγραμμο τμήμα στις 72 μοίρες είναι ο κλάδος της επιστήμης των υπολογιστών). Το μήκος των κυανών ευθύγραμμων τμημάτων είναι αμελητέο και στις δύο περιπτώσεις.



Σχήμα 8.7: S Kollias second level of field classification



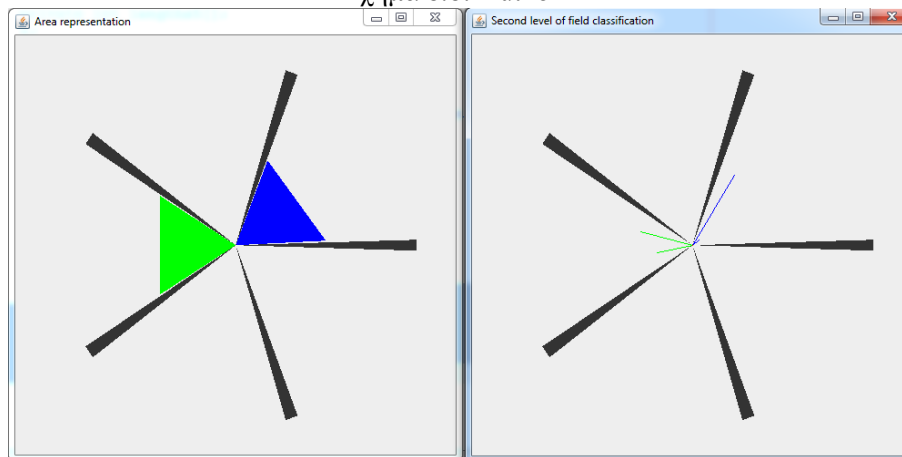
Δυστυχώς τα δεδομένα εισόδου που έχουμε είναι ατελή μέχρι αυτό το σημείο και για το λόγο αυτό δεν μας δίνεται η δυνατότητα να παρουσιάσουμε περισσότερα ρεαλιστικά παραδείγματα. Ωστόσο, μπορούμε να προιδεάσουμε για αυτά.

Σε αυτό το θεωρητικό παράδειγμα λοιπόν θα κατανοήσουμε καλύτερα τη σημασία της αναπαράστασης του second level of field classification. Ας θεωρήσουμε λοιπόν για έναν ερευνητή Α ότι έχει 100 δημοσιευμένες αναφορές με επιρροή στον τομέα των φυσικών & θετικών επιστημών και άλλες 100 με επιρροή στις κοινωνικές επιστήμες. Στις περιπτώσεις του ιστογράμματος και του area representation, η επιρροή θα φαίνεται ίση και στους δύο τομείς. Τι θα γίνει όμως αν τα δεδομένα μας λένε ότι από τις 100 αναφορές στις φυσικές & θετικές επιστήμες οι 90 ήταν στο επιστημονικό πεδίο της επιστήμης των υπολογιστών και οι 10 ήταν στα μαθηματικά, ενώ αντίστοιχα για τις κοινωνικές επιστήμες οι 60 αναφορές ήταν στο επιστημονικό κλάδο της πολιτικής και οι 40 στα οικονομικά;

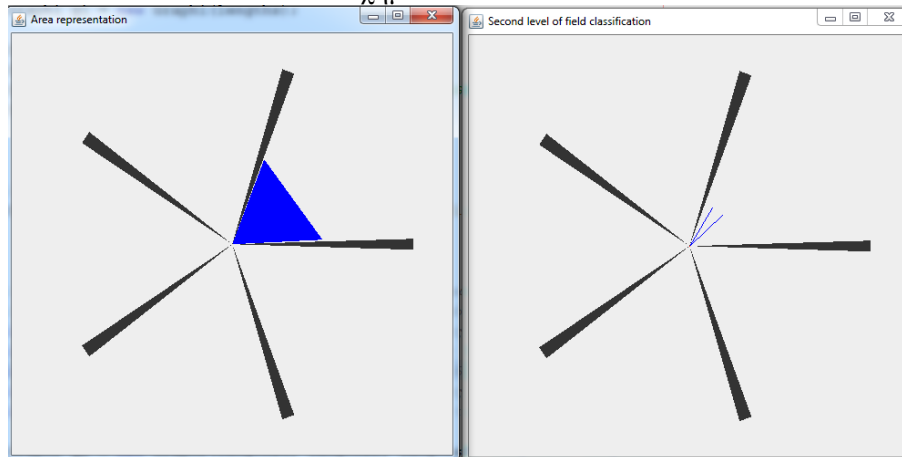
Η αναπαράσταση των αποτελεσμάτων με second level of field classification έρχεται να ξεκαθαρίσει το τοπίο για αυτή την περίπτωση. Οι 90 αναφορές στον κλάδο του computer science έχουν πρωταγωνιστικό ρόλο έτσι για τον ερευνητή Α μπορεί να συμπεράνει κανείς πως η μεγαλύτερη επιρροή που ασκεί είναι στην επιστήμη των υπολογιστών των φυσικών & θετικών επιστημών.

Εάν τώρα υποθέσουμε ότι θέλουμε να συγκρίνουμε τον ερευνητή A που συζητήσαμε προηγουμένως με έναν ερευνητή B. Ο δεύτερος έχει κι αυτός 100 δημοσιευμένες αναφορές στον τομέα των φυσικών & θετικών επιστημών και συγκεκριμένα 50 στον κλάδο του computer science και 50 στον κλάδο των μαθηματικών. Παρατηρώντας μόνο το ιστόγραμμα και το σχήμα του area representation θα συμπεραίναμε ότι και οι δύο έχουν την ίδια επιρροή στις φυσικές & θετικές επιστήμες. Αντίθετα, με μία πιο προσεκτική ματιά στο σχήμα second level of field classification θα βλέπαμε ότι ο ερευνητής A έχει μεγάλη επιρροή στην επιστημη των υπολογιστών και πολύ μικρή στον τομέα των μαθηματικών, ενώ ο ερευνητής B ίση επιρροή και στους δύο κλάδους. Συνεπώς, μπορεί να έχουν και οι δύο ίση συνολική επιρροή στις φυσικές & θετικές επιστήμες αλλά ειδικότερα ο ερευνητής A έχει μεγαλύτερη επιρροή στην επιστήμη των υπολογιστών και ο ερευνητής B μεγαλύτερη στα μαθηματικά.

Σχήμα 8.8: Author A



Σχήμα 8.9: Author B



Κεφάλαιο 9

Συμπεράσματα

Ανακεφαλαιώνοντας, στο έργο αυτό προτείνουμε και υλοποιούμε ένα τρόπο για την εξόρυξη χρήσιμων πληροφοριών από τα επιστημονικά άρθρα. Ειδικότερα, εξετάζουμε το περιοδικό ή το συνέδριο στο οποίο δημοσιεύτηκε μία αναφορά, και με βάση αυτό το αντιστοιχίζουμε με τον κατάλληλο επιστημονικό κλάδο με τον οποίο και συνδέεται. Έχουμε ήδη προτείνει μία ιεραρχία των επιστημών, με ιδιαίτερη έμφαση στο τομέα της πληροφορικής. Στη συνέχεια, με τη βοήθεια γράφου εξάγαμε βαθμό επιρροής για να καθορίσουμε τη σύνδεση των διεπιστημών. Έπειτα, με ένα νέο και περισσότερο διαισθητικό τρόπο οπτικοποιήσαμε τα αποτελέσματά μας με τη χρήση της γλώσσας προγραμματισμού java. Οι αναπαραστάσεις που επιλέξαμε βασίζονται κυρίως σε περιοχές (area representation) και όχι τόσο σε μπάρες (bar representation). Με τον τρόπο αυτό επιτυγχάνουμε περισσότερη πληροφορία δοσμένη σε ένα απλό σχήμα. Μόνη προϋπόθεση είναι ο ερευνητής να έχει αποκτήσει ήδη μία οικειότητα με το συγκεκριμένο τρόπο οπτικοποίησης. Τελειώνοντας, δίνουμε τη δυνατότητα, με τη χρήση ενός απλού προγράμματος, να διακρίνουμε τους επιστημονικούς τομείς με τους οποίους συνδέεται η κάθε αναφορά και κατά συνέπεια και ο κάθε ερευνητής/συγγραφέας αυτής.

Το έργο αυτό, θα μπορούσε να χρησιμοποιηθεί μεμονομένα ως ελεύθερο πρόγραμμα, το οποίο δύναται να χρησιμοποιεί ο κάθε ερευνητής για να ελέγξει την απήχηση της δουλειάς κάποιου άλλου ερευνητή, να αποφανθεί εάν τελικά σχετίζεται με τον κλάδο που τον ενδιαφέρει και τι αντίκτυπο είχε σε αυτόν. Ωστόσο όμως δεν αποκλείεται να χρησιμοποιηθεί και ως κομμάτι κάποιας εφαρμογής στην οποία και θα μπορούσε να συνεισφέρει επιπλέον χρήσιμη πληροφορία. Ένα σημαντικό βήμα ίσως να ήταν η ένταξη των οπτικοποιημένων αποτελεσμάτων της εφαρμογής στα web based εργαλεία όπως google scholar, research gate κ.α. τα οποία έχουν αναφερθεί νωρίτερα. Η εισαγωγή τέτοιων γραφημάτων θα είχε ενδιαφέρον και στην προσωπική σελίδα του κάθε ερευνητή και θα έδινε αυτομάτως σε κάθε επισκέπτη μία γενική εικόνα για το συνολικό έργο του, καταναμένο στις προκαθορισμένες επιστήμες.

Ένα από τα προβλήματα που αντιμετωπίσαμε αρχικά ήταν να αποφασίσουμε αν πράγματι το μέσο δημοσίευσης παρέχει την απαιτούμενη πληροφορία σε αρκετά αξιόπιστο επίπεδο. Διαλέγοντας λοιπόν γνωστά συνέδρια και περιοδικά παρατηρήσαμε ότι το ποσοστό των αναφορών που αποκλείει από την επιστήμη συσχετισμού του μέσου δημοσίευσής του, ή η σχέση του με κάποια επιστήμη είναι αμφίβολη είτε δυσδιάκριτη, είναι σχετικά μικρό. Το δεύτερο δίλημμα στο οποίο βεθήκαμε ήταν η επιλογή της ιεραρχίας των επιστημών. Στους περισσότερους τομείς αποφασίσαμε να χρησιμοποιήσουμε τις δημοφιλέστερες κατηγοριοποιήσεις. Ειδικότερα, στον τομέα της

πληροφορικής πήραμε αυτούσιο το μοντέλο της ACM (2012) το οποίο έχει αναπτυχθεί σε μία πολύ-ιεραρχική οντολογία και δίνει μία καλή και σαφή εικόνα για τους υποκλάδους της ενώ παράλληλα φαίνεται να είναι και ανεκτικό στις τεχνολογικές εξελίξεις που αναμένονται στο μέλλον. Αφού καταλήξαμε σε κάποια συγκεκριμένη ιεραρχία, βρεθήκαμε αντιμέτωποι με το πρόβλημα των διεπιστημών. Δηλαδή, για τις επιστήμες που παρουσιάζουν στοιχεία από διαφορετικούς επιστημονικούς κλάδους, να εξάγουμε τον αντίστοιχο βαθμό συσχέτισης με την καθεμιά. Τη λύση σε αυτό το πρόβλημα έδωσε η κατασκευή κατευθυνόμενου άκυκλικού γράφου με βάρη που παρουσιάζει ξεκάθαρα την κάθε δυνατή συσχέτιση. Ιδιαίτερη λεπτό φάνηκε να είναι το ζήτημα του "ξεκαθαρίσματος" της χρήσιμης πληροφορίας από τις λίστες που μας δόθηκαν. Έπρεπε να ελέγξουμε ενδελεχώς τις περιπτώσεις που δεν εμφανιζόταν σωστή κωδικοποίηση, ή που έλειπαν απαραίτητες πληροφορίες όπως ονόματα συγγραφέων και τίτλος δημοσίευσης, και ακόμα δυσκολότερα να κρίνουμε εάν ο τίτλος του μέσου δημοσίευσης έδινε όλη της απαραίτητη πληροφορία ή ήταν ανεπαρκής. Στο κομμάτι της υλοποίησης μας απασχόλησαν κυρίως δύο πράγματα. Ο τρόπος με τον οποίο θα κρατάμε σωστά τα δεδομένα που μας δίνονται μαζί με τις απαραίτητες συσχετίσεις τους. Δηλαδή, τον τρόπο με τον οποίο θα συνδέαμε τους συγγραφείς με τις επιστήμες, τις επιστήμες με τα αντίστοιχα πεδία τους και κατά συνέπεια και τους συγγραφείς με τα πεδία και ούτω καθεξής για τα υποπεδία κτλ. Τη λύση σε αυτό το πρόβλημα δώσαμε με τη δημιουργία καταλληλών αντικειμένων και τη χρήση του xml parser για την ανάλυση των δεδομένων στα xml αρχεία. Τέλος, στο κομμάτι της οπτικοποίησης η σχεδίαση των τριγώνων σε συγκεκριμένες περιοχές ανάλογα με τις μοίρες επέφερε κάποιες δυσκολίες οι οποίες όμως αντιμετωπίστηκαν με κατανόηση βασικής γεωμετρίας. Η λογική που ακολουθήθηκε στο κομμάτι του area representation χρησιμοποιήθηκε με μικρές τροποποιήσεις και στο second level field classification.

Η εργασία αυτή θα είχε ιδιαίτερο ενδιαφέρον αν επεκτεινόταν για περισσότερα συνέδρια, περιοδικά, δημοσιεύσεις και συγγραφείς. Έτσι, αν περιείχε μία μεγάλη γκάμα θα μπορούσε να χρησιμοποιηθεί και να ενταχθεί από τα web based εργαλεία που ήδη υπάρχουν και προσφέρουν σημαντική πληροφορία για την ερευνητική περιοχή. Παράλληλα λοιπόν με τις λίστες παραπομπών, που παρέχονται από αυτά, θα μπορούσε να παρέχεται και μία λίστα με τις επιστήμες που επηρεάζονται είτε με βάση το επιστημονικό άρθρο είτε με βάση τον κάθε ερευνητή/συγγραφέα. Επίσης χρήσιμο θα ήταν με την κατάλληλη συλλογή των απαραίτητων πληροφοριών για κάθε αναφορά τα σχήματα να προχωρούν πιο βαθιά στο δένδρο, σε συγκεκριμένες περιοχές, έτσι ώστε να είναι φανερές και οι υποκατηγορίες για μία καλύτερη εικόνα του αντικτύπου του κάθε ερευνητή. Τέλος, ένα μεγάλο βήμα θα ήταν να εκπαιδύσουμε το σύστημα ώστε να μη παίρνει έτοιμες από κάποιον εμπειρογνώμονα τις αντιστοιχίες των επιστημών με μέσα δημοσίευσης. Αντίθετα, ελέγχοντας τον τίτλο του δεύτερου να αντιστοιχίζει φράσεις κλειδιά (π.χ. image processing, computer graphics) με επιστημονικά πεδία διατρέχοντας (π.χ. με DFS) κόμβους γράφου (δένδρου), όπου ο κάθε κόμβος θα αναπαριστά επιστημονικό πεδίο ή επιστήμη ανάλογα με το επίπεδο που βρίσκεται (ύπαρξη φύλλων κοντά του).

Παράρτημα

Ιεραρχία επιστημών

1. Physical/Natural Sciences

(α) Physics

- i. Applied Physics
- ii. Cosmology
- iii. Astronomy
 - A. Astrophysics
- iv. Biophysics
- v. Geophysics

(β) Chemistry

- i. Biochemistry
- ii. Geochemistry
- iii. Physicochemistry

(γ) Earth and Planetary Science

- i. Geology
 - A. Volcanology
 - B. Seismology
 - Γ. Mineralogy
 - Δ. Paleontology
- ii. Meteorology
- iii. Climatology
- iv. Oceanography

(δ) Life Sciences/Biology

- i. Botanic
- ii. Zoology
- iii. Ecology
- iv. Genetics
- v. Biotechnology
- vi. Molecular Biology
- vii. Physiology

(ε) Mathematics

- i. General/foundations (mathematical logic)
- ii. Discrete mathematics/algebra
- iii. Analysis
- iv. Geometry and topology
- v. Applied mathematics

(στ) Statistics

(ζ) Computer Science

- i. General Literature
- ii. Hardware
 - A. Control structures
 - B. Microprogramming
 - Γ. Arithmetic and logic structures
 - Δ. memory structures
 - E. Input/Output and data communications
 - ΣΤ. Register-transfer-level implementation
 - Z. logic design
 - H. Integrated circuits
 - Θ. performance tuning
 - I. reliability engineering
- iii. Computer Systems Organization
 - A. Processor architectures
 - B. Computer-communication networks
 - Γ. Special purpose and application-based systems
 - Δ. Performance of system implementation
- iv. Software
 - A. Programming technics
 - B. Software engineering
 - Γ. Programming languages
 - Δ. Operating systems
- v. Data
 - A. Data structures
 - B. Data storage representation
 - Γ. Data encryption
 - Δ. Coding
 - E. Information theory
 - ΣΤ. Files
- vi. Theory of Computation
 - A. Computation by abstract devices
 - B. Analysis of algorithms

-
- Γ. Problem complexity
 - Δ. Logic and meanings of program
 - E. Mathematical logic
 - ΣΤ. Formal languages
 - vii. Mathematics of Computing
 - A. Numerical analysis
 - B. Discrete mathematics
 - Γ. Probability and statistics
 - Δ. Mathematical software
 - viii. Information Systems
 - A. Models and principles
 - B. Database management
 - Γ. Information storage and retrieval
 - Δ. Information systems applications
 - E. Information interfaces and presentation
 - ix. Computing Methodologies
 - A. Symbolic and algebraic manipulation
 - B. Artificial intelligence
 - Γ. Computer graphics
 - Δ. Image processing and computer vision
 - E. Pattern recognition
 - ΣΤ. Simulation and modeling
 - Z. Document and text processing
 - x. Computer Applications
 - A. Physics science and engineering
 - B. Life and medical science
 - Γ. Social and behavioral sciences
 - Δ. Arts and humanities
 - E. Computer-aided engineering
 - ΣΤ. Computers in other systems
 - xi. Computing Milieux
 - A. Computer industry
 - B. History of computing
 - Γ. Computers and education
 - Δ. Computers and society
 - E. Legal aspects of computing
 - ΣΤ. Management of computing and information systems
 - Z. Computing profession
 - H. Personal computing

2. Applied Sciences

- (α) Mechanical engineering
 - i. Architecture
 - ii. Mechanics
 - iii. Mechanology
 - iv. Topology
 - v. Chemical and mechanical engineering
- (β) Health sciences
 - i. Medicine
 - ii. Pharmacy
 - iii. Veterinary
- (γ) Geotechnical sciences
 - i. Veterinary
 - ii. Geology
 - iii. Geoponics
 - iv. Enviromental engineering
 - v. Forest engineering
 - vi. Ichthyology
- (δ) Science of management

3. Social Sciences

- (α) Psychology
- (β) Politics
- (γ) Law and order
- (δ) Sociology
- (ε) Economics
- (στ) Geography

4. Universe Sciences

- (α) Cosmology
- (β) Metaphysics
- (γ) Theology
- (δ) Astronomy

5. Human Sciences

- (α) Philology
- (β) History
- (γ) Archeology
- (δ) Anthropology
- (ε) Philosophy
- (στ) Anatomy

Βιβλιογραφία

- [] *Histogram*. URL: www.cs.armstrong.edu/liang/intro06/evennumberedexercise/Histogram.java.
- [Blo14] Google Scholar Blog. *Google Scholar Citations Open To All*. retrieved 2. Αύγ. 2014.
- [Dan05] Sips M. Daniel A. Keim. «Visual Data-Mining Techniques». Στο: *The Visualization Handbook* 43 (2005), σσ. 831–843.
- [Kei02] Daniel A. Keim. «Information Visualization and Visual Data Mining». Στο: *IEEE transactions on visualization and computer graphics* 7.1 (Μαρ. 2002), σσ. 100–107.
- [Ora] Oracle. *Reading XML Data into a DOM*. URL: <https://docs.oracle.com/javase/tutorial/jaxp/dom/readingXML.html>.
- [ove] Stack overflow. *Triangles and lines*. URL: <http://stackoverflow.com/questions/3536428/draw-a-line-at-a-specific-angle-in-java>.
- [W3s] W3schools. *XML Tree*. URL: http://www.w3schools.com/xml/xml_tree.asp.
- [Wal14] Manolis Wallace. «Extracting and visualizing research impact semantics». Στο: *SMAP 9th International Workshop on Semantic and Social Media Adaptation and Personalization* (Νοέ. 2014), σσ. 104–109.
- [Wika] Wikipedia. *Applied sciences*. URL: http://en.wikipedia.org/wiki/Outline_of_applied_science#Fields_of_applied_science.
- [Wikb] Wikipedia. *List of academic databases and search engines*. URL: http://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines.
- [Wikc] Wikipedia. *Natural sciences*. URL: http://en.wikipedia.org/wiki/Natural_science.
- [Wikd] Wikipedia. *Science classifications*. URL: <http://el.wikipedia.org/wiki/%C3%8EC2%95%C3%8FC2%80%C3%8EC2%B9%C3%8FC2%83%C3%8FC2%84%C3%8EC2%AE%C3%8EC2%BC%C3%8EC2%B7>.
- [Wike] Wikipedia. *Sciences/Categories*. URL: <http://en.wikipedia.org/wiki/Portal:Science/Categories>.
- [Wikf] Wikipedia. *Social sciences*. URL: http://en.wikipedia.org/wiki/List_of_social_sciences.
- [Wik04] Wikipedia. *List of scientific journals*. retrieved 3. Αύγ. 2004. URL: <http://en.wikipedia.org/wiki/List%20of%20scientific%20journals>.
- [Wik12] Wikipedia. *ACM Computing Classification System*. 2012. URL: <http://www.acm.org/about/class/class/2012>.

