



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

**ΔΙΕΠΑΦΕΣ ΜΕΣΩ WEB ΓΙΑ ΜΗ ΟΜΟΙΟΓΕΝΗ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΣΕ
ΠΡΑΓΜΑΤΙΚΟ ΧΡΟΝΟ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
του
Γιάκα Αθανασίου

Βόλος, Οκτώβριος 2013

Η σελίδα αυτή είναι σκόπιμα λευκή.



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

**ΔΙΕΠΑΦΕΣ ΜΕΣΩ WEB ΓΙΑ ΜΗ ΟΜΟΙΟΓΕΝΗ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΣΕ
ΠΡΑΓΜΑΤΙΚΟ ΧΡΟΝΟ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
του Γιάκα Αθανασίου

Επιβλέπων : Μανόλης Βάβαλης
Καθηγητής Τ.Η.Μ.Μ.Υ.

(Υπογραφή)

.....
Μανόλης Βάβαλης
Καθηγητής Τ.Η.Μ.Μ.Υ

(Υπογραφή)

.....
Κατσαρός Δημήτριος
Λέκτορας Τ.Η.Μ.Μ.Υ.

Βόλος, Οκτώβριος 2013

(Υπογραφή)

.....

Γιάκας Αθανάσιος

© 2013 – All rights reserved

Η σελίδα αυτή είναι σκόπιμα λευκή.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τους καθηγητές κ. Βάβαλη Εμμανουήλ και Κατσαρό Δημήτριο για την υποστήριξη και τις πολύτιμες συμβουλές που μου παρείχαν κατά την διάρκεια εκπόνησης της εργασίας μου αλλά και σε όλη την διάρκεια των σπουδών μου. Ιδιαίτέρως, τον κύριο Βάβαλη ο οποίος είχε την υπομονή να αναγνωρίσει τα ενδιαφέροντα μου και να με κατευθύνει προτείνοντάς μου την οπτικοποίηση δεδομένων ως αντικείμενο μελέτης, έρευνας και εφαρμογών.

Πίνακας Περιεχομένων

Περιεχόμενα

Εικόνες	10
1 Εισαγωγή.....	12
1.1 Οπτικοποίηση δεδομένων	12
1.2 Αντικείμενο της διπλωματικής	12
1.3 Οργάνωση κειμένου	14
2 Διεπαφές για οπτικοποίηση δεδομένων	14
2.1 Web Interfaces.....	14
2.1.1 Εμπειρία χρήστη (User Experience).....	15
2.2 Σύνολα δεδομένων	16
2.2.1 Τύπος δομής των δεδομένων	16
2.2.2 Δεδομένα πολλών διαστάσεων	17
2.3 Οπτική αναπαράσταση πληροφορίας.....	17
2.3.1 Αιτιολογική και διερευνητική οπτικοποίηση.....	18
2.3.2 Γραφήματα πληροφοριών (Info Graphics)	18
2.4 Γραφήματα.....	19
2.4.1 Γραμμικά δεδομένα.....	19
2.4.2 Χωρικά δεδομένα (Spatial Data).....	22
2.4.3 Δεδομένα δικτύου	23
2.4.4 Ιεραρχικά δεδομένα	28
2.4.5 Δεδομένα πολλών διαστάσεων	30
2.4.6 Επιλογές επιστημονικών σχεδιασμών στην οπτικοποίηση δεδομένων	31
2.4.7 Κλίμακες.....	32
2.4.8 Επικεφαλίδες, λεζάντες, σχόλια	32
2.5 Το ανθρώπινο οπτικό σύστημα	32
2.5.1 Preattentive μεταβλητές.....	33
2.5.2 Περιττή κωδικοποίηση (Redundant Encoding).....	34
2.5.3 Κατηγοριοποίηση οπτικών ιδιοτήτων	34
2.6 Δυναμικές οπτικοποιήσεις δεδομένων	36
2.6.1 Διαδραστικότητα	36

2.6.2	Συστήματα πραγματικού χρόνου	36
2.6.3	Ροή Γραφήματος (Graph Stream)	37
2.7	Εργαλεία οπτικοποίησης δεδομένων	37
2.7.1	SVG, HTML5 Canvas	37
2.7.2	Βιβλιοθήκες Javascript	38
2.7.3	Η γλώσσα προγραμματισμού R	40
2.7.4	Tableau	40
2.7.5	Gephi	41
2.7.6	Quadrigram	42
2.7.7	Visual.ly, infogr.am	43
2.7.8	Map Box	43
3	Οπτικοποίηση αποτελεσμάτων απογραφής πληθυσμού	44
3.1	Περιγραφή του προβλήματος	44
3.2	Στόχος	45
3.3	Προδιαγραφές	46
3.4	Σχεδιασμός	46
3.5	Λειτουργίες	48
3.6	Εργαλεία	48
3.6.1	Google Maps Javascript API V3	48
3.6.2	JSON	49
3.7	Επίδειξη	50
3.7.1	Επίπεδο Δημοτικών Ενοτήτων	50
3.7.2	Επίπεδο Δημοτικών Κοινοτήτων	50
3.7.3	Επίπεδο Οικισμών	51
4	Παρατηρητήριο ερευνητικών δραστηριοτήτων για το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Π.Θ.	52
4.1	Παρατηρητήρια Ερευνητικών Δραστηριοτήτων	52
4.1.1	Βάσεις δεδομένων επιστημονικών δημοσιεύσεων	53
4.1.2	Συστήματα CRIS	54
4.1.3	Βιβλιομετρικοί δείκτες μέτρησης της ερευνητικής δραστηριότητας	54
4.1.4	Επιστημονικά πεδία	56
4.1.5	Διαδικτυακά Παρατηρητήρια (Web Observatories)	57
4.1.6	Παρατηρητήρια ερευνητικών δραστηριοτήτων σε πανεπιστήμια της υφηλίου	58

4.2	Στόχος.....	59
4.3	Σχεδιασμός.....	60
4.3.1	Επεξεργασία δεδομένων και σχηματισμός βάσης δεδομένων.....	60
4.3.2	Model View Controller (MVC).....	62
4.3.3	Οπτικοποίηση της πληροφορίας.....	65
4.4	Λειτουργίες.....	65
4.5	Εργαλεία.....	66
4.5.1	HTML 5, CSS3 , JAVASCRIPT, JQuery.....	66
4.5.2	JSP, JAVA SERVLETS.....	67
4.5.3	Apache Tomcat.....	69
4.5.4	MySQL.....	69
4.5.5	NetBeans.....	69
4.5.6	Ωκεανός (Okeanos).....	70
4.6	Επίδειξη.....	72
5	Ανάλυση των οπτικοποιήσεων.....	76
5.1	Χάρτης συνεδριών.....	76
5.2	Σύννεφο ετικετών (Tag Cloud) για τις λέξεις κλειδιά.....	77
5.3	Δίκτυο συνεργασιών.....	77
5.4	Διάγραμμα ράβδων για αναπαράσταση των βιβλιομετρικών δεικτών.....	80
5.5	Heat Map για τον πληθυσμό του Δήμου Βόλου.....	81
6	Σύνοψη.....	82
7	Βιβλιογραφία.....	83
7.1	Βιβλία.....	83
7.2	Ιστοσελίδες.....	83
7.3	Αναφορές.....	84

Πίνακας Εικόνων

Εικόνες

Εικόνα 1-1 Από τα δεδομένα στην σοφία. Στην οπτικοποίηση της εικόνας βλέπουμε πως όσο βελτιώνεται το εννοιολογικό πλαίσιο και ο τρόπος κατανόησης, τα δεδομένα μας μετασχηματίζονται σε σοφία. Πηγή: http://www.easterbrook.ca/steve/2012/09/what-is-climate-informatics/	13
Εικόνα 2-1 Infographic που απεικονίζει την "έκρηξη των δεδομένων". Πηγή: http://www.ibmbigdatahub.com/blogs	19
Εικόνα 2-2 Γράφημα σωρευμένων ράβδων για την απεικόνιση του αριθμού των πατεντών στην πορεία του χρόνου ανά χώρα. Η αναλογία των πατεντών που αντιστοιχούν σε κάθε χώρα φαίνεται από την διαφοροποίηση του χρώματος σε κάθε ράβδο. Πηγή: <i>Quadrigram</i>	20
Εικόνα 2-3 Στο γράφημα με κουτιά μπορούμε να διακρίνουμε εύκολα το πεδίο τιμών αλλά και τον μέσο όρο που αντιστοιχεί σε κάθε τιμή της κατηγορικής μεταβλητής. Πηγή: <i>Highcharts.com</i> . Διάγραμμα σημείων (<i>Scatterplot</i>)	21
Εικόνα 2-4 Το διάγραμμα φυσαλίδας είναι ουσιαστικά ένα διάγραμμα σημείων με μία επιπλέον διάσταση που αναπαρίσταται από το εμβαδόν της κάθε φυσαλίδας. Πηγή: <i>Highcharts.com</i>	21
Εικόνα 2-5 <i>Wind Rose</i> για απεικόνιση των ανέμων στην πολιτεία του <i>Oregon</i> . Πηγή: <i>Highcharts.com</i>	22
Εικόνα 2-6 <i>Heat Map</i> απεικόνιση κατανομής εκπτώσεων στην Ισπανία. Τα θερμά χρώματα συμβολίζουν υψηλές πυκνότητες. Πηγή: <i>Quadrigram</i>	23
Εικόνα 2-7 Οι κόμβοι που βρίσκονται στο κέντρο των τοπικών κοινωνιών έχουν υψηλότερο <i>closeness centrality</i> . Πηγή: <i>activatenetworks.net</i>	25
Εικόνα 2-8 Οι κόμβοι που συμπεριφέρονται ως γέφυρες μεταξύ των συστάδων του δικτύου έχουν υψηλότερη διαμεσότητα. Πηγή: <i>activatenetworks.net</i>	26
Εικόνα 2-9 Παρατηρούμε ότι η σελίδα C έχει υψηλότερο <i>ragerank</i> από την σελίδα E, παρόλο που οι υπερσύνδεσμοι προς την E είναι λιγότεροι από αυτούς προς την C. Αυτό συμβαίνει γιατί ο μοναδικός σύνδεσμος προς την C έρχεται από τον κόμβο B ο οποίος έχει μεγάλο <i>ragerank</i> . Πηγή: <i>Wikipedia</i>	27
Εικόνα 2-10 Σε έναν γράφο οι οντότητες συμβολίζονται ως κόμβοι και οι μεταξύ τους σχέσεις ως σύνδεσμοι. Στην εικόνα μας, οι κόμβοι χρωματίζονται με βάση την ομάδα στην οποία ανήκουν. Πηγή: <i>Gephi.org</i>	27
Εικόνα 2-11 <i>Circle Limit IV</i> . Ο M.C. Escher γύρω στο 1956 ασχολήθηκε με τις Υπερβολικές Ψηφιδωτεθήςεις, που είναι κανονικά πλακίδια στο Υπερβολικό επίπεδο. Για να παρουσιάσει την εργασία του δημιούργησε το 1959 τέσσερις ξυλογραφίες με την ονομασία <i>Circle Limit</i> . Πηγή: <i>Wikipedia</i>	28
Εικόνα 2-12 Οπτικοποίηση του BBC με την κατανομή των 100 site με την μεγαλύτερη επισκεψιμότητα ανά θεματική κατηγορία (με βάση το χρώμα). Σε ένα διαδραστικό περιβάλλον ο χρήστης έχει την επιλογή να τοποθετήσει τον κέρσορα του ποντικιού πάνω από ένα τετράγωνο για να δει το όνομα του ιστότοπου και τον αριθμό των επισκέψεων. Πηγή: <i>BBC</i>	29
Εικόνα 2-13 Στο <i>Mosaic plot</i> της εικόνας τα κουτιά διαχωρίζονται ανάλογα με τις τιμές των τεσσάρων μεταβλητών: <i>hair, sex, eye color, pearson residuals</i> . Πηγή: http://statmath.wu.ac.at/projects/vcd/	30
Εικόνα 2-14 Κάθε παράλληλη συντεταγμένη αναπαριστά ένα μέτρο επίδοσης κάποιου αυτοκινήτου, ενώ κάθε γραμμή ένα συγκεκριμένο αυτοκίνητο. Σε ένα διαδραστικό περιβάλλον μπορούμε να φιλτράρουμε το πλήθος γραμμών προς εμφάνιση, με βάση το ποια αυτοκίνητα μας ενδιαφέρουν, για να κάνουμε ευκολότερα τις συγκρίσεις. Πηγή: http://mbostock.github.io/protovis/ex/cars-full.html	31
Εικόνα 2-15 Στο παράδειγμα της εικόνας, ενώ και τα τέσσερα αντικείμενα έχουν ακριβώς το ίδιο χρώμα (κατά απόλυτη τιμή), ο εγκέφαλος τα αντιλαμβάνεται ως διαφορετικά, λόγω της αλλαγής στην φωτεινότητα του φόντου. Στην συγκεκριμένη περίπτωση ο εγκέφαλος μας έχει αντιληφθεί την οπτική διαφορά του χρώματος του αντικειμένου με το χρώμα που το περιβάλλει. Μεταξύ των άλλων, ένα συμπέρασμα που μπορούμε να βγάλουμε μελετώντας το παραπάνω παράδειγμα είναι ότι πρέπει να αποφεύγουμε την χρήση χρωματικών διαβαθμίσεων	

(διανυσμάτων – color gradient) ως φόντο σε κάποιο γράφημα. Πηγή: <i>Table and graph design for enlightening communication</i> by Stephen Few. <i>Perceptual Edge</i> . Σελίδα 86	33
Εικόνα 2-16 Οι 8 preattentive μεταβλητές του Tidwell. Πηγή: <i>An introduction to visualizing data</i> by Joel Laumans (σελίδα 14).	34
Εικόνα 2-17 Ο παραπάνω πίνακας των συνηθισμένων οπτικών ιδιοτήτων μπορεί να μας φανεί αρκετά χρήσιμος κατά την διαδικασία κωδικοποίησης για τους τύπους των δεδομένων μας. Πηγή: <i>Designing data visualizations</i> by Noab Iliinski & Julie Steele (σελίδα 28).	35
2-18 Με το Tableau Public μπορεί να δημιουργήσει κανείς, αυτόματα, οπτικοποιήσεις βελτιστοποιημένες για οθόνες αφής και χρήση μέσω iPad ή Android Tablets. Πηγή: www.tableausoftware.com	41
2-19 Το περιβάλλον εργασίας στο Overview του Gephi.	42
2-20 Το περιβάλλον εργασίας του Quadrigram.	43
2-21 Το περιβάλλον εργασίας του TileMill.	44
3-1 Μέρος του συνόλου των δεδομένων μας, τα οποία αντλήθηκαν από την ιστοσελίδα του Δήμου Βόλου σε μορφή Excel.	46
3-2 Στην αριστερή εικόνα βλέπουμε ότι η φυσαλίδα του Δήμου Βόλου καλύπτει πολλές από τις γειτονικές περιοχές. Μεγαλώνουμε το μέγεθος της κλίμακας και στα δεξιά βλέπουμε ότι πλέον δεν υπάρχει το πρόβλημα αυτό.	47
3-3 Πληθυσμός ανά δημοτική ενότητα	50
3-4 Ο πληθυσμός των κοινοτήτων αναπαρίσταται ως φυσαλίδα.	51
3-5 Ο πληθυσμός των οικισμών αναπαρίσταται ως φυσαλίδα	51
4-1 Υπολογισμός του h index. Πηγή: Wikipedia	56
4-2 Οπτικοποίηση Δικτύου Συνεργατών στο παρατηρητήριο του HKU.	58
4-3 Προφίλ ερευνητή στο παρατηρητήριο του Πανεπιστημίου του Κονκόρντια.	59
4-4 Σχηματική απεικόνιση της βάσης δεδομένων μας	61
4-5 Σχηματική απεικόνιση της MVC αρχιτεκτονικής. Πηγή: http://www.c-sharpcorner.com/UploadFile/3d39b4/AspNet-mvc-introduction/Images/MVC-Introduction2.jpg	63
Εικόνα 6 Ο κκεανός ως υπηρεσία τριών σύννεφων, ένα σύννεφο για πρόσβαση σε μηχανές μεγάλης υπολογιστικής ισχύος, ένα σύννεφο-υπηρεσία αποθηκευτικού χώρου, και ένα σύννεφο πρόσβασης σε εικονικά δίκτυα.	70
Εικόνα 7 Η επιφάνεια εργασίας της εικονικής μας μηχανής.	71
4-8 Αναζήτηση ερευνητή	72
4-9 Σελίδα προφίλ ερευνητή.	73
4-10 Παράδειγμα επερώτησης μέσω συμπλήρωσης της φόρμας αναζήτησης δημοσίευσης.	74
4-11 Ο πίνακας με τα αποτελέσματα της αναζήτησης που πραγματοποιήσαμε πριν ως παράδειγμα.	75
Εικόνα 5-1 Δισδιάστατη οπτικοποίηση για την απεικόνιση των συνεδρίων του Αναπληρωτή Καθηγητή του Τ.Η.Μ.Μ.Υ. κ. Μπέλλα. Παρατηρούμε ότι τα περισσότερα συνέδρια στο οποία συμμετείχε έχουν διεξαχθεί στις Η.Π.Α.	76
Εικόνα 5-2 Παρατηρώντας το σύννεφο ετικετών του κ. Κατσαρού βλέπουμε ότι οι περιοχές που δραστηριοποιείται περισσότερο είναι, μεταξύ άλλων, τα ασύρματα δίκτυα και το Web.	77
5-3 Το δίκτυο των συνεργασιών των μελών ΔΕΠ. Οι κόμβοι εδώ έχουν διαταχθεί με τον αλγόριθμο Reingold.	78
Εικόνα 5-4 Το δίκτυο συνεργασιών του κυρίου Μποζάνη	79
Εικόνα 5-5 Κατάταξη με βάση το H-Index.	81
5-6 Χάρτης θερμότητας για τον πληθυσμό του Δήμου Βόλου.	81

Κεφάλαιο 1

1 Εισαγωγή

1.1 Οπτικοποίηση δεδομένων

Τα τελευταία χρόνια παρατηρούμε μια μεγάλη πρόοδο αναφορικά με την δυνατότητα του ανθρώπου να συγκεντρώνει και να αποθηκεύει δεδομένα αλλά και να εξάγει βασικούς δείκτες απόδοσης αποκρυπτογραφώντας τις διάφορες μετρήσεις. Είναι σύνηθες φαινόμενο, παραδείγματος χάριν στον κόσμο των πωλήσεων, να βασιζόμαστε σε ποσοτικά δεδομένα για να μετρήσουμε το πόσο υγιής είναι μία οργάνωση, να βγάζουμε συμπεράσματα, να εντοπίζουμε ευκαιρίες και να προσπαθούμε να προβλέψουμε το μέλλον. Ωστόσο, σε αντίθεση με την σοφία, η πληροφορία δεν μπορεί να μιλήσει από μόνη της. Μπορεί κάποιος να επενδύσει στο να χτίσει την αποδοτικότερη αποθήκη δεδομένων, που να λειτουργεί πάνω σε ένα πανίσχυρο υλικό και να είναι προσβάσιμη από ένα πανέξυπνο λογισμικό, αλλά αν οι άνθρωποι που εργάζονται με τα δεδομένα δεν ξέρουν πως να αντλήσουν νόημα από αυτά ή πώς να το παρουσιάσουν στους φορείς λήψης των αποφάσεων, όλη αυτή η επένδυση θα είναι κυριολεκτικά μια σπατάλη.

Η **οπτικοποίηση των δεδομένων** (data visualisation ή σύντομα data vis) είναι η οπτική αναπαράσταση των δεδομένων, με την έννοια ότι η πληροφορία είναι αφηρημένη σε μια μορφή που συμπεριλαμβάνει χαρακτηριστικά ή μεταβλητές για τις μονάδες που την σχηματίζουν.

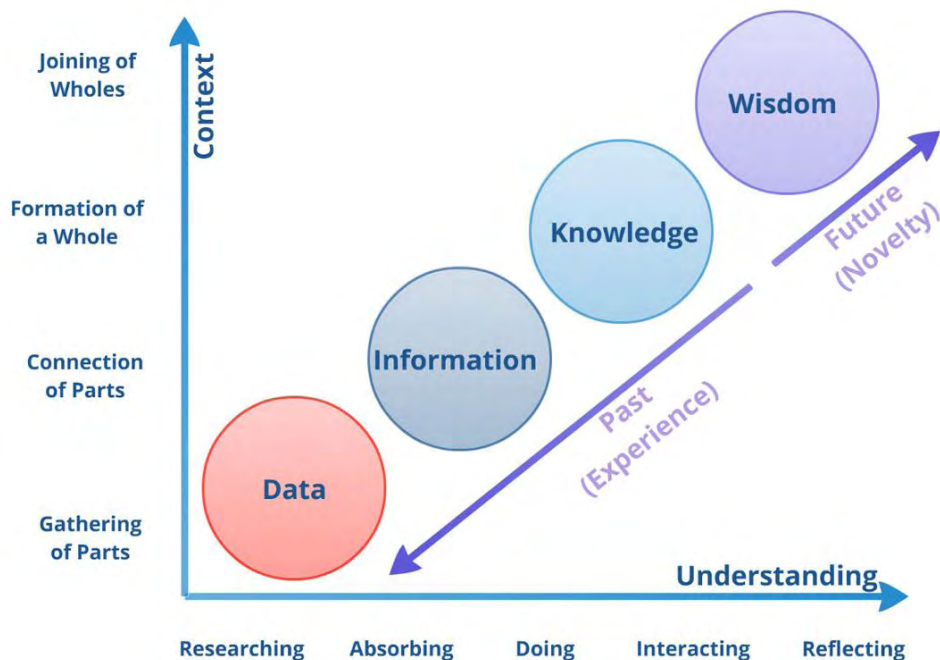
1.2 Αντικείμενο της διπλωματικής

Η παρούσα διπλωματική εξετάζει την εφαρμογή των κατάλληλων διεπαφών για την παραγωγή γνώσης η οποία προέρχεται από την οπτικοποίηση συνόλων δεδομένων.

Στόχος μας είναι η σχεδίαση λογισμικών τα οποία αντλούν από τα δεδομένα (DATA) πληροφορία (INFORMATION) και την παρουσιάζουν στον χρήστη με τρόπο που αυτός μπορεί να την αντιληφθεί εύκολα και απευθείας με την όρασή του. Η αλληλεπίδραση αυτή του ανθρώπου με την μηχανή (λογισμικό) θα πρέπει να έχει ως αποτέλεσμα την παραγωγή γνώσης (KNOWLEDGE). Εν συνεχεία ο χρήστης του λογισμικού θα πρέπει να έχει την δυνατότητα να κατευθύνει το πεδίο γνώσης εστιάζοντας σε αυτά που τον ενδιαφέρουν περισσότερο και συνθέτοντας τα όσα έχει κατανοήσει σε βάθος. Κάτι τέτοιο θα μπορούσε να έχει σαν αποτέλεσμα την παραγωγή εφαρμοσμένης γνώσης και ιδανικά σοφίας (WISDOM). Η παραπάνω συζήτηση μπορεί να παρουσιαστεί συνοπτικά στην εικόνα 1-1.

Επιπρόσθετα της προσπάθειάς μας για πλήρη κατανόηση των βασικών εμπλεκόμενων εννοιών και των υφισταμένων και αναδυόμενων τεχνολογιών σκοπός μας είναι η πρακτική αξιολόγηση τους μέσω δύο μελετών πεδίου. Στην πρώτη από αυτές προσπαθούμε να αποτυπώσουμε τα αποτελέσματα των απογραφών των πληθυσμών στον Δήμο Βόλου με οπτικοποιημένη μορφή πάνω σε έναν χάρτη. Στην δεύτερη, σκοπός μας είναι να οπτικοποιήσουμε τα δεδομένα της σε "πραγματικό" χρόνο ενημερωμένης βάσης που έχει δημιουργηθεί για τις ανάγκες ενός παρατηρητηρίου ερευνητικών

δραστηριοτήτων των μελών Διδακτικού Ερευνητικού Προσωπικού (Δ.Ε.Π.) του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών (Τ.Η.Μ.Μ.Υ.) του Πανεπιστημίου Θεσσαλίας (Π.Θ.).



Εικόνα 1-1 Από τα δεδομένα στην σοφία. Στην οπτικοποίηση της εικόνας βλέπουμε πως όσο βελτιώνεται το εννοιολογικό πλαίσιο και ο τρόπος κατανόησης, τα δεδομένα μας μετασχηματίζονται σε σοφία. Πηγή: <http://www.easterbrook.ca/steve/2012/09/what-is-climate-informatics/>

Επιλέξαμε ως μελέτη περίπτωσης την απεικόνιση των αποτελεσμάτων της απογραφής του Δήμου Βόλου, τα οποία είναι δημοσιευμένα στην ιστοσελίδα του Δήμου, επειδή παρά το γεγονός ότι κάποιος επισκέπτης της ιστοσελίδας μπορεί να δει πόσους κατοίκους έχει το κάθε δημοτικό διαμέρισμα και ο κάθε οικισμός, ωστόσο είναι αρκετά δύσκολο κάποιος απλά μελετώντας τους αριθμούς να αποκτήσει μια συνολική εικόνα για την γεωγραφική κατανομή του πληθυσμού ή να συγκρίνει δύο περιοχές. Επιπρόσθετα η δημοτική αρχή οφείλει να έχει μια πλήρη εικόνα (εικόνα τόσο κυριολεκτικά όσο και από εννοιολογικής άποψης) τόσο σε συγκεντρωτικό επίπεδο όσο και σε λεπτομέρειες οι οποίες θα επιτρέψουν τοπικές ζωτικές σημασίας σχεδιασμούς και δράσεις.

Στην δεύτερη μελέτη πεδίου, προσπαθήσαμε να οπτικοποιήσουμε την ερευνητική δραστηριότητα των μελών Δ.Ε.Π. του Τ.Η.Μ.Μ.Υ. του Π.Θ. Μέσω του αντίστοιχου παρατηρητηρίου (web observatory) έχουμε πρόσβαση σε πληροφορία για τις επιστημονικές δημοσιεύσεις των μελών ΔΕΠ, η οποία συγκεντρώνεται από διάφορες μηχανές αναζήτησης σε πραγματικό χρόνο. Σκοπός μας είναι χρησιμοποιώντας την παραπάνω πληροφορία να αναπαραστήσουμε οπτικά το επιστημονικό προφίλ του τμήματος, την παραγωγικότητα, την εξέλιξη στον χρόνο, το εύρος των θεματικών περιοχών, την εξωστρέφεια, το είδος των συνεργασιών κλπ..

Για να επιτύχουμε τους σκοπούς μας θα μελετήσουμε σε βάθος τα χαρακτηριστικά και τις ικανότητες του ανθρωπίνου οπτικού συστήματος, μέσω των οποίων ο άνθρωπος μπορεί να αντιληφθεί και να κατανοήσει. Επίσης θα λάβουμε σοβαρά υπ' όψιν τους κανόνες τους οποίους το διέπουν.

1.3 Οργάνωση κειμένου

Το υπόλοιπο της παρούσας εργασίας είναι οργανωμένο με τον εξής τρόπο:

Στο **2^ο κεφάλαιο** αναφερόμαστε γενικά στα σύνολα των δεδομένων τα οποία μπορεί να αντλούνται από διάφορες πηγές. Επιπρόσθετα εξηγούμε πως οι διεπαφές οπτικοποίησης δεδομένων λειτουργούν ως δίαυλος επικοινωνίας υψηλού εύρους ζώνης μεταξύ των ανθρώπινων αισθητήρων όρασης και του εκάστοτε λογισμικό επεξεργασίας των δεδομένων. Στην συνέχεια παρουσιάζουμε τους κανόνες λειτουργίας του ανθρώπινου οπτικού συστήματος και πως, και με ποια εργαλεία, αυτοί μπορούν να χρησιμοποιηθούν για την οπτικοποίηση δεδομένων τα οποία μπορεί να είναι διαδραστικά και να αλλάζουν συνεχώς σε πραγματικό χρόνο.

Στο **3^ο κεφάλαιο** εξετάζουμε το πρόβλημα της οπτικής απεικόνισης των αποτελεσμάτων των απογραφών πληθυσμού στον Δήμο Βόλου. Στόχος μας είναι ο χρήστης της εφαρμογής μας να μπορεί να αντιλαμβάνεται το πως κατανέμεται ο πληθυσμός στον δήμο επιλέγοντας ανάμεσα σε δύο επίπεδα (δημοτικές ενότητες, οικισμοί), εστιάζοντας στον χώρο που τον ενδιαφέρει.

Στο **4^ο κεφάλαιο** χρησιμοποιούμε τις τεχνικές οπτικοποίησης για να σχεδιάσουμε την διεπαφή ενός παρατηρητηρίου (web observatory) των επιστημονικών δραστηριοτήτων των μελών Δ.Ε.Π. του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας.

Στο **5^ο κεφάλαιο** κάνουμε μια μελέτη ξεχωριστά για κάθε οπτικοποίηση που αναπτύξαμε και προσπαθούμε να καταγράψουμε την γνώση που μπορούμε να αντλήσουμε από αυτήν.

Στο **6^ο κεφάλαιο** κάνουμε μία σύνοψη της εργασίας μας τονίζοντας τις διαπιστώσεις στις οποίες οδηγηθήκαμε κατά την μελέτη του αντικείμενου της διπλωματικής. Παραθέτουμε ακόμη κάποιους σχολιασμούς και παρατηρήσεις και προβληματιζόμαστε για το κατά πόσο λύσαμε το πρόβλημα. Τέλος εκθέτουμε τις σκέψεις μας για το τι πρέπει να γίνει στο μέλλον για να είμαστε σε θέση να σχεδιάζουμε άμεσα διαδραστικές οπτικοποιήσεις.

Κεφάλαιο 2

2 Διεπαφές για οπτικοποίηση δεδομένων

2.1 Web Interfaces

Η σχεδίαση διεπαφών αποτελεί υποκατηγορία του επιστημονικού πεδίου που ονομάζεται επικοινωνία ανθρώπου – μηχανής (HCI Human-computer Interaction). Η επικοινωνία ανθρώπου μηχανής είναι η μελέτη και ο σχεδιασμός του τρόπου με τον οποίον άνθρωποι και μηχανές δουλεύουν μαζί έτσι ώστε οι

ανάγκες του ανθρώπου να ικανοποιούνται με τον πιο αποδοτικό τρόπο, με βάση το τι θέλουν και τι περιμένουν οι άνθρωποι, τους φυσικούς περιορισμούς, τις ικανότητες που κατέχει ο άνθρωπος και το τι βρίσκει ελκυστικό και απολαυστικό.

Η διεπαφή χρήστη αποτελεί κομμάτι ενός λογισμικού με το οποίο κάποιος μπορεί να αντιληφθεί με κάποια από τις αισθήσεις του (συνήθως με την ακοή ή την όραση) όπως επίσης να κατανοήσει κάτι ή να το κατευθύνει.

Μία εξαιρετικά σχεδιασμένη διεπαφή, είναι αυτή που επιτρέπει τον χρήστη να εστιάσει στην πληροφορία καθαυτή και όχι στους μηχανισμούς που παρουσιάζουν την πληροφορία. Με άλλα λόγια μια πολύ καλά σχεδιασμένη διεπαφή είναι αυτή που είναι σχεδόν "αόρατη" στον χρήστη, καθώς αυτός δεν μπορεί να την αντιληφθεί. Ο χρήστης άλλωστε επισκέπτεται έναν ιστότοπο ή χρησιμοποιεί μια διαδικτυακή υπηρεσία για να πετύχει έναν σκοπό π.χ. να κάνει μία αγορά ή να μοιραστεί κάτι με τους φίλους του και όχι να παίξει με την διεπαφή δοκιμάζοντας όλες τις δυνατότητες της.

Αν μια διεπαφή μπερδεύει τον χρήστη και δεν ικανοποιεί τις ανάγκες του, ο χρήστης δυσκολεύεται να πραγματοποιήσει τις εργασίες και οδηγείται σε επιπρόσθετα λάθη, κάτι που μπορεί να έχει ως αποτέλεσμα την οριστική εγκατάλειψη του συστήματος από την χρήστη και την καλλιέργεια αισθημάτων απογοήτευσης και άγχους.

Μια διεπαφή στο web (web interface) έχει ως στόχο την διαδραστικότητα ανάμεσα σε έναν χρήστη και σε ένα λογισμικό το οποίο τρέχει σε έναν Web server. Η διεπαφή χρήστη είναι ο φυλλομετρητής και η σελίδα ή η εφαρμογή η οποία έχει κατέβει και έχει αποδοθεί σε αυτόν.

Αξίζει εδώ να αναφέρουμε ότι ένα γραφικό σύστημα αποτελείται από αντικείμενα και πράξεις. Σε μια διεπαφή τα αντικείμενα είναι οι ανεξάρτητες μονάδες που μπορεί να δει ένας άνθρωπος στην οθόνη.

2.1.1 Εμπειρία χρήστη (User Experience)

Ο όρος εμπειρία χρήστη (User Experience, αρχικά UX) συμπεριλαμβάνει τα συναισθήματα κάποιου ανθρώπου που αναπτύχθηκαν κατά την χρήση ενός προϊόντος, ενός συστήματος ή μιας υπηρεσίας. Τονίζει τις πειραματικές, τις συναισθηματικές, τις με νόημα και αξία εκείνες όψεις της επικοινωνίας ανθρώπου και μηχανής και της κυριότητας προϊόντων. Επιπρόσθετα, συμπεριλαμβάνει τις αντιλήψεις του ανθρώπου για πρακτικά θέματα όπως χρησιμότητα, ευκολία χρήσης και αποδοτικότητα συστήματος. Η εμπειρία χρήστη είναι υποκειμενική μιας και έχει να κάνει με προσωπική αντίληψη και δυναμική αφού συνεχώς αλλάζει στο πέρασμα του χρόνου, εξαιτίας των συνεχών πάσης φύσεως αλλαγών του περιεχομένου αλλά και των νέων τεχνολογικών ανακαλύψεων.

Πολλοί παράγοντες μπορούν να επηρεάσουν την συνολική εμπειρία χρήστη. Για να συγκεκριμενοποιήσουμε κάπως αυτήν την πληθώρα, οι παράγοντες αυτοί έχουν ομαδοποιηθεί σε τρεις κύριες κατηγορίες:

- κατάσταση χρήστη και προηγούμενη εμπειρία,
- ιδιότητες συστήματος,
- εννοιολογικό πλαίσιο χρήσης (κατάσταση).

Η μελέτη της συμπεριφοράς του τυπικού χρήστη, του περιβάλλοντος και της μεταξύ τους αλληλεπίδρασης μας βοηθά στην σχεδίαση του συστήματος.

2.2 Σύνολα δεδομένων

Μιλήσαμε πριν για οπτικοποίηση δεδομένων, χωρίς όμως να δώσουμε τον ακριβή ορισμό του τι είναι τα δεδομένα. Ως δεδομένα, χαρακτηρίζεται ένα μη αξιολογημένο σύνολο διακριτών στοιχείων, μιας παρατήρησης-αναφοράς, μια συλλογή που αποτυπώνει "τιμές" επί αντικειμένων, προσώπων, γεγονότων κλπ. Στην ουσία τα δεδομένα είναι καταγραφές από παρατηρήσεις μας για τον έξω κόσμο, ή προϊόντα κάποιου συστήματος προσομοίωσης και μελέτης. Ενδεχομένως κάποιοι από τους τύπους δεδομένων που παρατίθενται εμπεριέχουν ήδη εμφανείς πληροφορίες, όχι όμως την πληροφορία στο επίπεδο που συνθέτει μια αξιολόγησή τους.

Μια συλλογή δεδομένων ονομάζεται σύνολο δεδομένων (data set ή dataset). Συνήθως ένα σύνολο δεδομένων αντιστοιχεί σε έναν πίνακα μιας βάσης δεδομένων ή ενός στατιστικού πίνακα δεδομένων, όπου κάθε στήλη αναπαριστά μια συγκεκριμένη μεταβλητή και κάθε γραμμή ένα συγκεκριμένο μέλος του συνόλου δεδομένων.

Πολλές φορές αναγκαζόμαστε να συγκεντρώσουμε δεδομένα από διάφορες πηγές. Όταν τα δεδομένα μας δεν έχουν την ίδια μορφή τότε έχουμε να κάνουμε με μη ομοιογενή σύνολα δεδομένων. Για παράδειγμα, όπως θα δούμε και στην συνέχεια, για να δημιουργήσουμε ένα παρατηρητήριο ερευνητικών δραστηριοτήτων για τα μέλη ΔΕΠ του Πανεπιστημίου Θεσσαλίας θα πρέπει να συγκεντρώσουμε δεδομένα από διάφορες μηχανές αναζήτησης επιστημονικών δημοσιεύσεων, ώστε τα αποτελέσματα που θα προκύψουν να είναι αξιόπιστα. Στην περίπτωση αυτή, το σύνολο δεδομένων που παίρνουμε μετά από κάποια ερώτηση μας σε μια Διεπαφή Προγραμματισμού Εφαρμογών (API) της μίας μηχανής θα είναι διαφορετικά δομημένο σε σχέση με αυτό κάποιας άλλης μηχανής. Σε αυτή την περίπτωση έχουμε να κάνουμε με ανομοιογενή σύνολα δεδομένων τα οποία πρέπει να τα επεξεργαστούμε κατάλληλα ώστε να μπορέσουμε να τα ενοποιήσουμε.

2.2.1 Τύπος δομής των δεδομένων

Γραμμική δομή: Τα δεδομένα μπορούν να οργανωθούν σε πίνακες. Για παράδειγμα, πηγαίος κώδικας προγράμματος, αλφαβητική λίστα, χρονολογικά ταξινομημένα αντικείμενα, καταταγμένα αποτελέσματα αναζήτησης.

Χρονική δομή: Δομή που αποτελείται από δεδομένα που μεταβάλλονται στον χρόνο.

Χωρική δομή (Spatial): Τα δεδομένα σχετίζονται με κάτι που αφορά την γεωγραφική τους θέση.

Ιεραρχική δομή: Τα δεδομένα είναι ιεραρχικά όπως μια γενεαλογία και έχουν την μορφή δέντρου όπως για παράδειγμα, ένα σύστημα αρχείων ή ένα σύστημα κατηγοριοποίησης σε μια βιβλιοθήκη.

Δίκτυα: Τα δεδομένα αποτελούνται από οντότητες που σχετίζονται μεταξύ τους. Αποτελούν γενικές δομές γραφημάτων, όπως τα γραφήματα υπερμέσων με κόμβους και συνδέσμους, σημασιολογικά δίκτυα, πλέγματα (webs) και άλλα.

Πολυδιάστατη: Μια πολυδιάστατη δομή εμπεριέχει χαρακτηριστικά μεταδεδομένων όπως τύπος, μέγεθος, εκδότης, ημερομηνία τροποποίησης κτλ. Αντικείμενα με N χαρακτηριστικά γίνονται σημεία στον N -διάστατο χώρο.

Χώρος γνωρισμάτων (Feature spaces): Από την Ανάκτηση Πληροφορίας (Information Retrieval - IR), ένα διάνυσμα γνωρισμάτων αναπαριστά κάθε αντικείμενο της συλλογής μας. Ο Χώρος Γνωρισμάτων προβάλλεται πάνω σε δύο ή τρεις οπτικές διαστάσεις. Τα αντικείμενα συνήθως τοποθετούνται στον χώρο δυναμικά με κάποια τεχνική κατευθυνόμενη από δυνάμεις (βλ. ενότητα 2-4-5).

2.2.2 Δεδομένα πολλών διαστάσεων

Οι οπτικοποιήσεις των δεδομένων χαρακτηρίζονται από τον αριθμό των διαστάσεων που αυτές αναπαριστούν. Μία διάσταση σε μια οπτικοποίηση είναι ένας διακριτός τύπος πληροφορίας που κωδικοποιείται οπτικά στο γράφημα. Για παράδειγμα, ένα γράφημα μπορεί να αναπαριστά τις τιμές των μετοχών μιας εταιρίας για όλες τις μέρες μιας εβδομάδας. Το γράφημα αυτό έχει δύο διαστάσεις: την τιμή και τον χρόνο. Αν τώρα στο ίδιο γράφημα προσθέσουμε και άλλες εταιρίες τότε το γράφημα μας θα αποκτήσει μία τρίτη διάσταση. Ο αριθμός των διαστάσεων ενός γραφήματος μπορεί να περιγραφεί ως το επίπεδο της πολυπλοκότητας της οπτικοποίησης. Η πολυπλοκότητα μιας οπτικοποίησης, λοιπόν, αυξάνει όσο μεγαλώνει ο αριθμός των διαστάσεων και όχι όσο αυξάνει ο αριθμός των δεδομένων της κάθε διάστασης.

Κάθε διάσταση ενός συνόλου δεδομένων αντιστοιχεί σε κάποια συγκεκριμένη οπτική ιδιότητα. Στην ενότητα 2.5 θα εξετάσουμε λεπτομερώς τις οπτικές ιδιότητες.

2.3 Οπτική αναπαράσταση πληροφορίας

Τα τελευταία χρόνια οι μηχανές αυξάνουν με γρήγορους ρυθμούς την χωρητικότητά τους. Για τον λόγο αυτό, μπορούμε να δημιουργήσουμε (μέσω μετρήσεων, δειγματοληψίας, προσομοίωσης) και να αποθηκεύσουμε τεράστιες ποσότητες δεδομένων. Τα δεδομένα μας μπορούμε επίσης να τα αντλήσουμε από το διαδίκτυο στο οποίο παρατηρούμε καταιγισμό δεδομένων. Για να διαπιστώσουμε την αφθονία των δεδομένων στο διαδίκτυο αρκεί να σκεφτούμε ότι ένα μεγάλο μέρος της ζωής μας το περνάμε online. Σχεδόν οποιαδήποτε δραστηριότητά μας συμπυκνώνεται σε bits και στέλνεται σε όλον τον κόσμο με την ταχύτητα του φωτός. Για παράδειγμα στα κοινωνικά δίκτυα παράγεται από τις δραστηριότητες των χρηστών τους ένας τεράστιος όγκος δεδομένων ανά δευτερόλεπτο. Ας σκεφτούμε επιπλέον την περίπτωση των κινητών συσκευών οι οποίες μπορούν ανά πάσα στιγμή να συγκεντρώνουν καταγραφές που αφορούν την καθημερινότητα του χρήστη, όπως η τοποθεσία στην οποία βρίσκεται, η μουσική που ακούει, οι συνομιλίες του, η ταχύτητα με την οποία οδηγεί κ.ο.κ. Στην εικόνα 2-1 βλέπουμε κάποιες από τις τεράστιες ποσότητες δεδομένων που παράγουν οι άνθρωποι καθημερινά. Αυτή η "έκρηξη των δεδομένων" όμως, έχει πολλές φορές ως συνέπεια να μην μπορούμε να συσχετίσουμε τα δεδομένα μας με τον αρχικό μας στόχο για τον οποίο ξεκινήσαμε να τα συγκεντρώνουμε. Όλη αυτή η κατάσταση γεννάει το ερώτημα: "Πως μπορούμε να κατανοήσουμε τόσα πολλά δεδομένα ;".

Η οπτικοποίηση των δεδομένων επωφελείται από τις τεράστιες δυνατότητες του ανθρώπινου οπτικού συστήματος το οποίο λειτουργεί ως δίαυλος επικοινωνίας υψηλού εύρους ζώνης μεταξύ των αισθητήρων όρασης και του ανθρώπινου εγκεφάλου. Επιπρόσθετα χρησιμοποιεί το πλεονέκτημα του "λογισμικού" του εγκεφάλου μας να αναγνωρίζει πρότυπα και να μεταδίδει σχέσεις και νοήματα. Τέλος

αξίζει να τονίσουμε ότι ενδεχομένως να χρησιμοποιηθούν και άλλα ιδιαίτερα επιθυμητά πλεονεκτήματα του ανθρώπινου νου που εν μέρει τουλάχιστον οφείλονται στην όραση όπως η πλαστικότητα της μνήμης και η δυνατότητα υποσυνείδητα [απώλειας μνήμης](#) (Christopher G. Healey, James. T. Enns, 2011). Μια οπτικοποίηση μπορεί να εμπνεύσει νέες ερωτήσεις και περαιτέρω ανακαλύψεις. Μπορεί επίσης να βοηθήσει στο να εντοπίσουμε υποπροβλήματα, τάσεις, ακραίες τιμές, συγκεκριμένα σημεία δεδομένων που βρίσκονται μέσα σε ένα μεγάλο σύνολο δεδομένων.

Η οπτικοποίηση δεδομένων δεν αφορά απλά τον μετασχηματισμό ενός πίνακα δεδομένων σε οπτικοποιημένη μορφή. Με την αποδοτική οπτικοποίηση των δεδομένων επιτυγχάνουμε την αναπαράσταση ακόμα και των πιο σύνθετων μεγάλων και πολύπλοκων συνόλων δεδομένων σε κατανοητή μορφή.

Ας σκεφτούμε ότι ένα σύνολο δεδομένων με M εγγραφές έχει 2^M υποσύνολα δεδομένων. Κάθε ένα από αυτά τα σύνολα ενδέχεται να είναι αυτό που εμείς αναζητούμε. Με μία καλή απεικόνιση δεδομένων, οι δικές μας ικανότητες αναγνώρισης προτύπων, μπορούν όχι μόνο να ταξινομήσουν τα δεδομένα μέσα από αυτήν την συνδυαστική έκρηξη, αλλά μπορούν ακόμη να εξάγουν εννοήσεις (Insights) μέσα από τα οπτικά πρότυπα.

2.3.1 Αιτιολογική και διερευνητική οπτικοποίηση

Σύμφωνα με την λίστα του Tufte (Tufte, Edward R., 2001) όλες οι οπτικοποιήσεις δεδομένων πρέπει να παρουσιάζουν τα δεδομένα, να είναι ακριβείς χωρίς να αλλοιώνουν τα δεδομένα, να κάνουν μεγάλα σύνολα δεδομένων συνεκτικά, να εξυπηρετούν έναν σαφή στόχο, να αποκαλύπτουν τα δεδομένα σε διαφορετικά επίπεδα (γενικά ή λεπτομερή) και να ενθαρρύνουν τον χρήστη.

Υπάρχουν δύο ειδών οπτικοποιήσεις δεδομένων η αιτιολογική και η διερευνητική:

Η **αιτιολογική οπτικοποίηση** των δεδομένων (explanatory data viz), η οποία χρησιμοποιείται για να μεταδώσει δεδομένα ή μια γνώμη από τον σχεδιαστή στον αναγνώστη. Συνήθως υπάρχει μία συγκεκριμένη ιστορία ή πληροφορία την οποία ο σχεδιαστής έχει σκοπό να μεταδώσει σε κάποιον άλλο. Για τον σκοπό αυτό ο σχεδιαστής θα πρέπει να αποφασίσει από πριν ποια πληροφορία είναι χρήσιμη και ποια είναι άσχετη. Όλη αυτή η διαδικασία συλλογής εστιασμένων δεδομένων στηρίζει την ιστορία που θέλει ο σχεδιαστής να περιγράψει. Η αιτιολογική οπτικοποίηση δεδομένων είναι κομμάτι της φάσης αναπαράστασης των δεδομένων

Η **διερευνητική οπτικοποίηση** (exploratory data viz) των δεδομένων χρησιμοποιείται από τον ίδιο τον σχεδιαστή για να ανακαλύψει μοτίβα, τάσεις ή υποπροβλήματα σε ένα σύνολο δεδομένων. Συνήθως δεν υπάρχει από πριν κάποια ήδη γνωστή ιστορία και ο σχεδιαστής δεν είναι σίγουρος για τα συμπεράσματα που θα προκύψουν από τα δεδομένα. Για το λόγο αυτό τα οπτικοποιεί προκειμένου να ανακαλύψει κάποια ενδιαφέρουσα καμπύλη, γραμμή, τάση ή κάποιες μη ομαλές ακραίες τιμές. Με τον τρόπο αυτό ο σχεδιαστής ανακαλύπτει την ιστορία που τα ίδια τα δεδομένα έχουν να πουν. Η διερευνητική οπτικοποίηση δεδομένων είναι κομμάτι της φάσης ανάλυσης των δεδομένων.

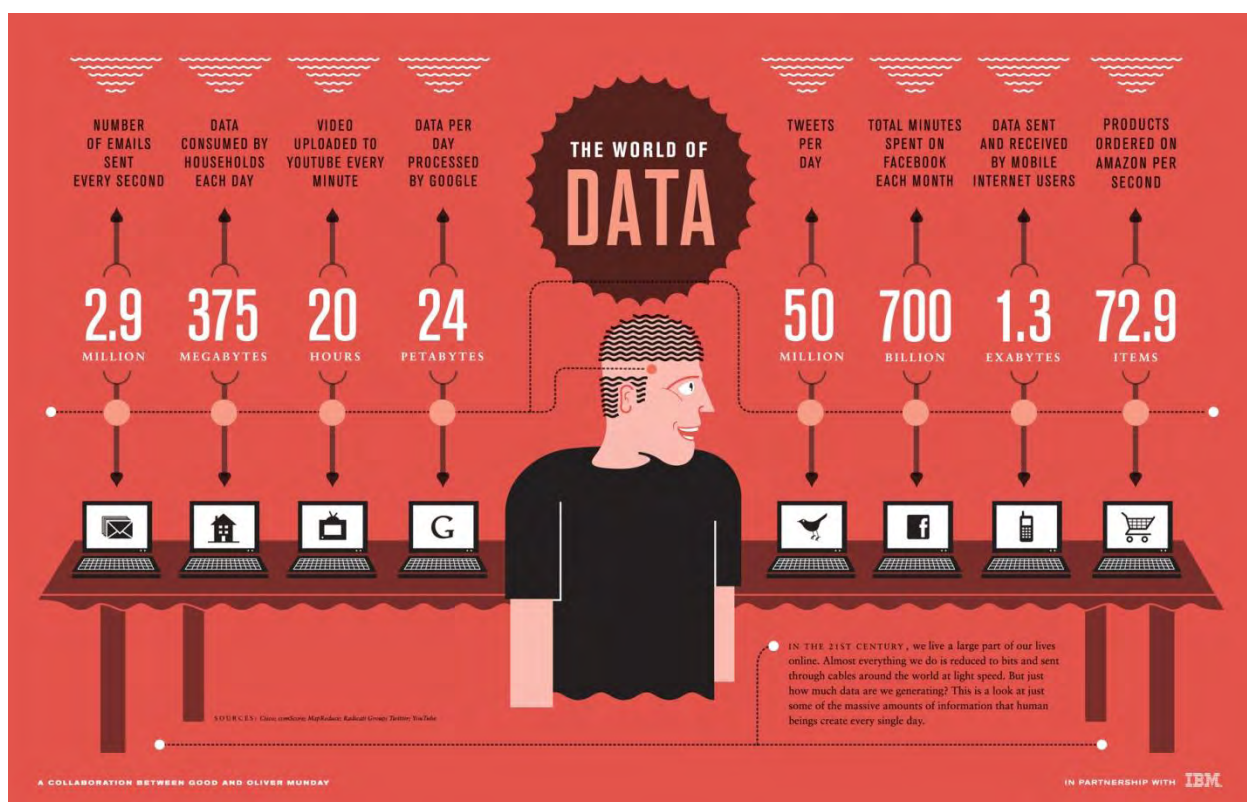
2.3.2 Γραφήματα πληροφοριών (Info Graphics)

Ο όρος γράφημα πληροφοριών (Infographic) αναφέρεται σε κάθε οπτική αναπαράσταση δεδομένων η οποία σχεδιάζεται χειροκίνητα για κάποια συγκεκριμένα δεδομένα. Ωφείλει να είναι εξαιρετικά

καλαίσθητη με πλούσιο περιεχόμενο που προσελκύει το μάτι και κρατά το ενδιαφέρον και είναι σχετικά φτωχή ως προς την ποσότητα των δεδομένων.

Στην ουσία τα γραφήματα πληροφοριών είναι εικονογραφήματα όπου η αναπαράσταση των δεδομένων σχεδιάζεται χειρωνακτικά ή με κάποιο λογισμικό όπως το Adobe Illustrator.

Πολλές φορές γίνεται σύγχυση των όρων οπτικοποίηση δεδομένων και γραφήματος πληροφοριών. Από τον παραπάνω ορισμό συγκρίνοντας τα βλέπουμε ότι η οπτικοποίηση της πληροφορίας αν και αρχικά σχεδιάζεται από τον άνθρωπο, στην συνέχεια σχεδιάζεται αλγοριθμικά με κάποιο λογισμικό οπτικοποίησης πληροφορίας ή σχεδίασης γραφημάτων. Για τον λόγο αυτό μια οπτικοποίηση δεδομένων μπορεί να επαναχρησιμοποιηθεί ως έχει για περισσότερα, νέα ή διαφορετικά δεδομένα. Από την άλλη, οι οπτικοποιήσεις δεδομένων αν και δείχνουν μεγάλο όγκο δεδομένων, δεν είναι τόσο καλαίσθητες όσο τα γραφήματα πληροφοριών.



Εικόνα 2-1 Infographic που απεικονίζει την "έκρηξη των δεδομένων". Πηγή: <http://www.ibmbigdatahub.com/blogs>

2.4 Γραφήματα

2.4.1 Γραμμικά δεδομένα

Στα μαθηματικά, ένα γράφημα είναι μια αφηρημένη αναπαράσταση ενός συνόλου στοιχείων, όπου μερικά ζευγάρια στοιχείων συνδέονται μεταξύ τους με δεσμούς. Τα διασυνδεδεμένα στοιχεία αναπαριστώνται με μαθηματικές έννοιες και ονομάζονται κορυφές ενώ οι δεσμοί που συνδέουν τα ζευγάρια των κορυφών ονομάζονται ακμές. Συνήθως, ένας γράφος απεικονίζεται σε διαγραμματική

μορφή ως ένα σύνολο κουκκίδων για τις κορυφές, ενωμένα μεταξύ τους με γραμμές ή καμπύλες για τις ακμές.

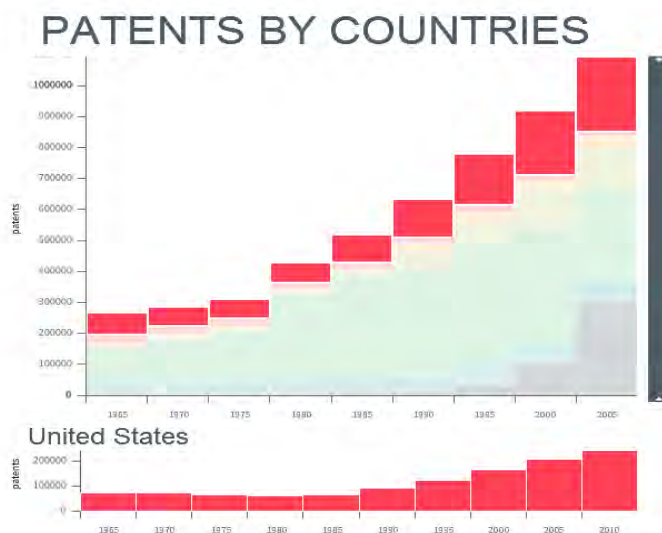
Τα γραφήματα είναι πολύ σημαντικά για την οπτική αναπαράσταση των δεδομένων, καθώς μοντελοποιούν πολλά σημαντικά προβλήματα.

Τα γραμμικά δεδομένα μπορούν να οπτικοποιηθούν με χρήση των διαγραμμάτων που υπάρχουν και στο Excel: διάγραμμα με σημεία, διάγραμμα με γραμμές (line chart), διάγραμμα πίτας (pie chart), διάγραμμα με ράβδους (bar chart). Πρέπει να προσέξουμε ότι, όσον αφορά τα διαγράμματα πίτας, ενώ αυτά είναι αρκετά δημοφιλή, πολλοί είναι αυτοί που προτείνουν να αποφεύγεται η χρήση τους, αφού η αντίληψη του ανθρώπου για τις γωνίες σε δισδιάστατες περιοχές ως ποσοτικό μέτρο, είναι περιορισμένη. Στο Excel έχουμε επίσης την επιλογή να κάνουμε τα γραφήματα μας να φαίνονται τρισδιάστατα (π.χ. 3D bars, 3D pie). Κάτι τέτοιο όμως θα πρέπει να αποφεύγεται (εκτός αν πρέπει αναγκαστικά με τον τρόπο αυτό να προσθέσουμε μία ακόμα διάσταση) γιατί οι συγκρίσεις στον τρισδιάστατο χώρο είναι πολύ πιο δύσκολες.

Στην συνέχεια θα παρουσιάσουμε κάποια άλλα είδη γραφημάτων που απεικονίζουν γραμμικά δεδομένα και χρησιμοποιούνται συχνά όταν έχουμε περισσότερες από δύο διαστάσεις.

Γράφημα σωρευμένων ράβδων (Stacked Bar Chart)

Αναμφίβολα, είναι πιο δύσκολες οι συγκρίσεις σε ένα γράφημα σωρευμένων ράβδων από ότι σε ένα γράφημα με ράβδους τοποθετημένες η μία δίπλα στην άλλη. Ωστόσο όταν έχουμε μια παραπάνω μεταβλητή και θέλουμε να οπτικοποιήσουμε τόσο ολόκληρο το μέτρο όσο και τα επί μέρους στοιχεία του, η χρήση του γραφήματος σωρευμένων ράβδων είναι η κατάλληλη επιλογή.

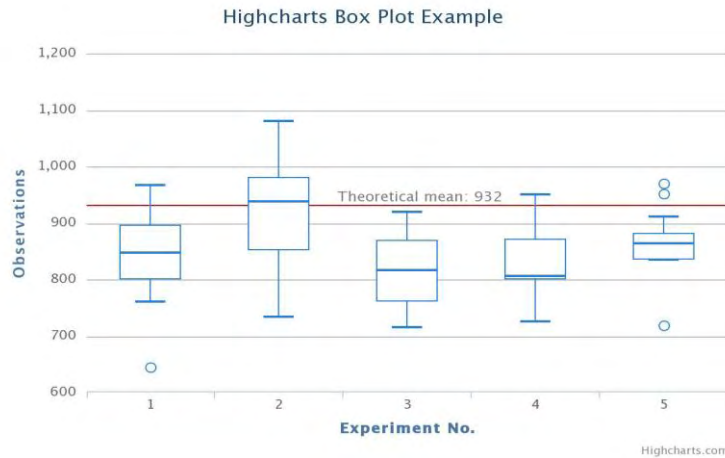


Εικόνα 2-2 Γράφημα σωρευμένων ράβδων για την απεικόνιση του αριθμού των πατεντών στην πορεία του χρόνου ανά χώρα. Η αναλογία των πατεντών που αντιστοιχούν σε κάθε χώρα φαίνεται από την διαφοροποίηση του χρώματος σε κάθε ράβδο.

Πηγή: [Quadriagram](#)

Γράφημα με κουτιά (Box Plot)

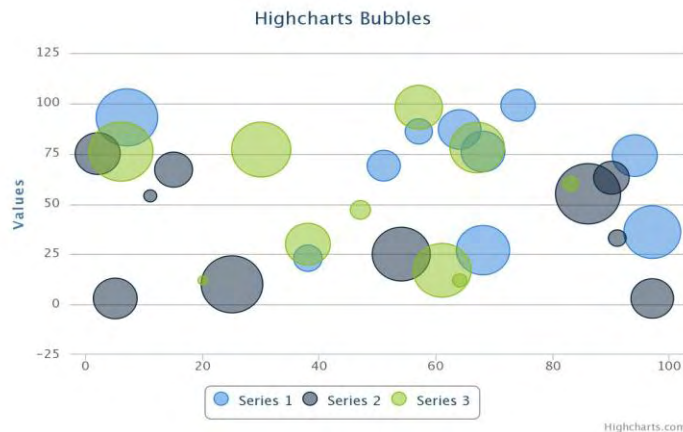
Το κουτί είναι παρόμοιο με την ράβδο με την μόνη διαφορά ότι και οι δύο άκρες του κουτιού χρησιμοποιούνται για να ορίσουν μια τιμή στον γράφο. Κάθε κουτί αναπαριστά μια κατανομή ενός συνόλου τιμών: το κάτω μέρος αναπαριστά την ελάχιστη τιμή, το πάνω την μέγιστη και το μήκος του, το πεδίο τιμών. Μπορούμε επιπλέον να χρησιμοποιήσουμε μία γραμμή για να δείξουμε το κέντρο της κατανομής που συνήθως είναι ο μέσος όρος. Χρησιμοποιούνται σε γραφήματα πολλαπλών κατανομών.



Εικόνα 2-3 Στο γράφημα με κουτιά μπορούμε να διακρίνουμε εύκολα το πεδίο τιμών αλλά και τον μέσο όρο που αντιστοιχεί σε κάθε τιμή της κατηγορικής μεταβλητής. Πηγή: [Highcharts.com](https://highcharts.com).

Διάγραμμα σημείων (Scatterplot)

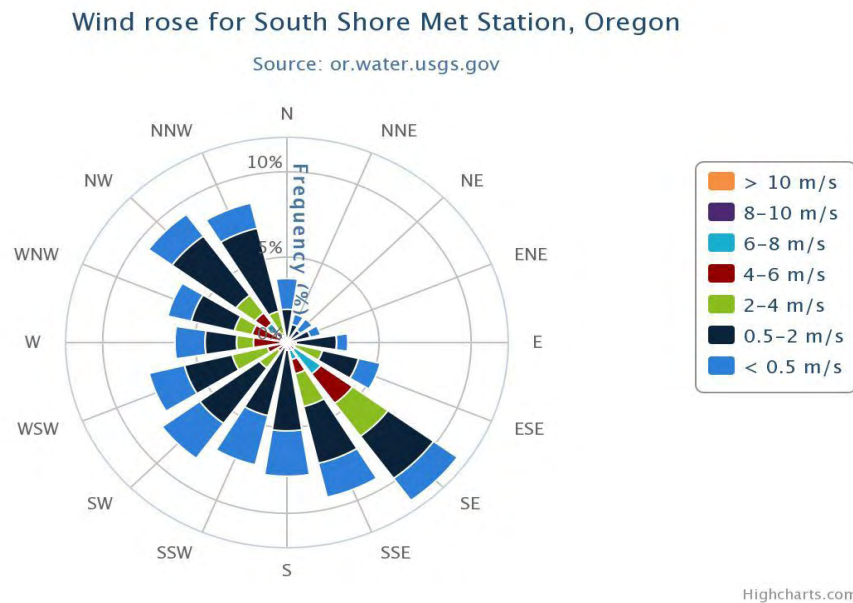
Ιδανικά για αναζήτηση συσχετίσεων μεταξύ δύο ποσοτικών μεταβλητών ή για οπτικοποίηση δεδομένων που μεταβάλλονται κατά μήκος δύο διαστάσεων. Είναι δυνατό να αναπαραστήσουν τρεις ακόμα και τέσσερις διαστάσεις κωδικοποιώντας σημεία δεδομένων ως φυσαλίδες (**Διάγραμμα Φυσαλίδας – Bubble Chart**), πίτες ή σωρευμένες ράβδους και χρωματίζοντας τα ανάλογα αν υπάρχει κάποια κατηγορική μεταβλητή. Συνήθως, η εξαρτημένη μεταβλητή σχεδιάζεται στον κάθετο άξονα.



Εικόνα 2-4 Το διάγραμμα φυσαλίδας είναι ουσιαστικά ένα διάγραμμα σημείων με μία επιπλέον διάσταση που αναπαρίσταται από το εμβαδόν της κάθε φυσαλίδας. Πηγή: [Highcharts.com](https://highcharts.com).

Wind Rose Plot

Χρησιμοποιείται κατά κόρον από τους μετεωρολόγους για να αποδώσει μια λακωνική όψη του πως η ταχύτητα και η κατεύθυνση του αέρα κατανέμονται σε μία συγκεκριμένη περιοχή. Χρησιμοποιούν ένα πολικό σύστημα συντεταγμένων σε μορφή πλέγματος και με τον τρόπο αυτό η συχνότητα των ανέμων μετά από μία μεγάλη χρονική περίοδο σχεδιάζεται πάνω στην διεύθυνση του ανέμου. Μπορούμε να κατηγοριοποιήσουμε τους ανέμους ανάλογα με το είδος τους (π.χ. την ταχύτητα τους) χρωματίζοντας ανάλογα προς την κάθε κατεύθυνση. Τελικά έχουμε τρεις διαστάσεις: κατεύθυνση, εύρος ταχύτητας, συχνότητα.



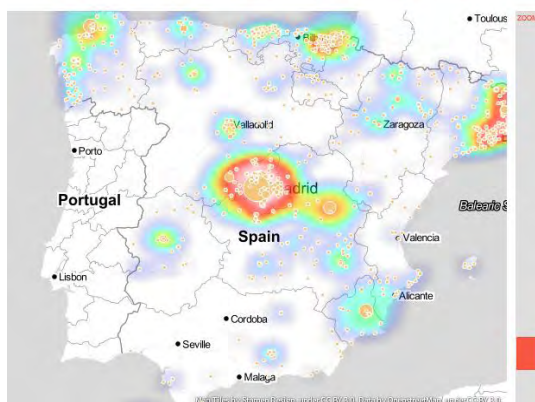
Εικόνα 2-5 Wind Rose για απεικόνιση των ανέμων στην πολιτεία του Oregon. Πηγή: Highcharts.com.

2.4.2 Χωρικά δεδομένα (Spatial Data)

Χάρτες Θερμότητας (Heat Maps)

Οι χάρτες θερμότητας είναι τρισδιάστατα γραφήματα περιοχής τα οποία χρησιμοποιούν το χρώμα και την φωτεινότητα για να υποδείξουν τιμές μεγάλων συνόλων δεδομένων. Οι δύο χωρικές διαστάσεις αναπαρίστανται από το γεωγραφικό μήκος και πλάτος, ενώ η τρίτη διάσταση χρησιμοποιείται για να δώσει την ένταση ενός σημείου δεδομένων σε σχετική σύγκριση με την μέγιστη τιμή του συνόλου δεδομένων. Η ένταση κωδικοποιείται με το χρώμα, συνήθως χρησιμοποιούμε θερμά χρώματα και τιμές μεγάλης έντασης και ψυχρά χρώματα για μικρής έντασης τιμές. Με τον τρόπο αυτό αναδεικνύουμε περιοχές με μεγάλες αλλαγές, τιμές εκτός εύρους ή άλλα ενδιαφέροντα χαρακτηριστικά. Καλή τακτική είναι η χρήση χρωμάτων με αυξημένη αντίθεση, ώστε να ξεχωρίσουμε τις ενδιαφέρουσες από τις κανονικές τιμές.

GEOGRAPHIC SALES DISTRIBUTION IN SPAIN



Εικόνα 2-6 Heat Map απεικόνισης κατανομής εκπτώσεων στην Ισπανία. Τα θερμά χρώματα συμβολίζουν υψηλές πυκνότητες. Πηγή: Quadrigram.

2.4.3 Δεδομένα δικτύου

Ένα δίκτυο μπορεί να αναπαρασταθεί ως γράφος μαθηματικά με τον πίνακα γειτνίασης. Ο πίνακας γειτνίασης ενός γράφου G με N κόμβους είναι ένας $N \times N$ πίνακας A , όπου τα στοιχεία του $a_{i,j}$ είναι ίσα με 1 αν ο κόμβος i είναι γειτονικός με τον κόμβο j , διαφορετικά ισούται με 0. Το σύνολο των ιδιοτιμών αυτού του πίνακα ονομάζεται φάσμα γραφήματος και χρησιμοποιείται στον προσδιορισμό του χώρου στον οποίον θα ενσωματωθεί ο γράφος, ως ένα σύνολο κορυφών και ακμών.

Για να σχεδιάσουμε έναν γράφο εισάγουμε τα δεδομένα μας, ορίζουμε τις οντότητες που θα αναπαρασταθούν ως κόμβοι και τις μεταξύ τους σχέσεις ως ακμές. Κατόπιν, εφαρμόζουμε κάποιον αλγόριθμο που θα αναδιατάξει την θέση των κόμβων ώστε οι ακμές να είναι περίπου ίδιου μήκους και να υπάρχουν όσο το δυνατόν λιγότερες διασταυρώσεις ακμών.

Τεχνικές κατευθυνόμενες με δυνάμεις (Force Directed Techniques)

Ο σχεδιασμός γράφων με αλγόριθμους κατευθυνόμενους με δυνάμεις είναι κλάσεις αλγορίθμων για σχεδίαση γράφων με καλαίσθητο και ευχάριστο τρόπο. Οι εν λόγω τεχνικές δανείζονται μια αναλογία που υπάρχει στην φύση, όπου οι κορυφές είναι σώματα με μάζα που έλκονται ή απωθούνται μεταξύ τους εξαιτίας μιας ελαστικότητας ή ενός ηλεκτρικού φορτίου. Οι δυνάμεις εφαρμόζονται σε ένα σύνολο κόμβων ή ακμών ανάλογα με την σχετική τους θέση. Στην συνέχεια, χρησιμοποιούμε αυτές τις δυνάμεις είτε για να προσομοιώσουμε την κίνηση των ακμών και των κορυφών είτε για να ελαχιστοποιήσουμε την ενέργεια τους.

Ο βελτιστοποιημένος γράφος είναι αυτός στον οποίο οι "φυσικές" αυτές δυνάμεις βρίσκονται σε ισορροπία. Σε αυτή τη κατάσταση οι δυνάμεις που εφαρμόζονται σε κάθε επανάληψη δεν αλλάζουν την σχετική θέση των κόμβων. Η ισότητα που αποτυπώνει αυτήν την αναλογία είναι η εξής:

$$Q(X|A, B) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \phi(d_{ij}(X)) - \sum_{i=1}^n \sum_{j=1}^n b_{ij} \psi(d_{ij}(X)),$$

όπου ο $n \times p$ πίνακας X περιέχει τις συντεταγμένες n κορυφών στον R^p και το $d_{ij}(X)$ υποδηλώνει τις αποστάσεις μεταξύ των σημείων με συντεταγμένες x_i και x_j . Τα βάρη a_{ij} αναφέρονται σε εκείνα του πίνακα γειτνίασης A του γράφου G (αν $a_{ij} = \{0,1\}$ μη βεβαρημένος γράφος), ενώ τα βάρη b_{ij} μπορεί να προέρχονται είτε από τον πίνακα γειτνίασης είτε από κάποιον εξωτερικό περιορισμό. Τελικά οι συναρτήσεις $\phi(\cdot)$ και $\psi(\cdot)$ είναι μετασχηματισμοί των οποίων ο ρόλος είναι να επιβάλουν μερικές καλίσθητες θεωρήσεις στην διάταξη μας. Για παράδειγμα η κυρτή συνάρτηση ϕ θα ενισχύσει τις μεγάλες αποστάσεις αποδίδοντας ακόμα μεγαλύτερες και ως εκ τούτου θα ανιχνεύσει μοναδικά χαρακτηριστικά στα δεδομένα, ενώ ένας κοίλος μετασχηματισμός θα μετριάσει το φαινόμενο των απομονωμένων κορυφών.

Μέτρα κεντρικότητας κόμβου

Στην θεωρία των γράφων και στην ανάλυση δικτύων για να μετρήσουμε πόσο κεντρικός, δηλαδή πόσο σχετικά σημαντικός, είναι ένας κόμβος σε ένα δίκτυο μπορούμε να χρησιμοποιήσουμε κάποιο μέτρο κεντρικότητας για τους κόμβους του δικτύου μας. Για παράδειγμα θα χρησιμοποιήσουμε κάποιο μέτρο κεντρικότητας για να δούμε πόσο μπορεί κάποιο μέλος ενός κοινωνικού δικτύου να επηρεάσει τα άλλα μέλη ή κατά πόσο είναι early adopter, πόσο σημαντικό είναι ένα δωμάτιο μέσα σε ένα σπίτι, πόσο καλά χρησιμοποιείται ένας δρόμος κάποιου συγκοινωνιακού δικτύου. Στην συνέχεια θα εξετάσουμε κάποια από τα μέτρα κεντρικότητας τόσο από την μαθηματική όσο και από την κοινωνιολογική τους σκοπιά.

Απλή συνδεσιμότητα (Degree Centrality)

Η απλή συνδεσιμότητα ορίζεται ως ο αριθμός των συνδέσεων που ξεκινάνε από (out) ή καταλήγουν (in) στον κόμβο μας. Αλλιώς ονομάζεται βαθμός. Όταν το γράφημα μας είναι κατευθυνόμενο μπορούμε να εξετάσουμε χωριστά τον εσωτερικό βαθμό (In Degree) και τον εξωτερικό βαθμό κάποιου κόμβου. (Out Degree).

Μπορούμε να ερμηνεύσουμε την απλή συνδεσιμότητα ως το πόσο άμεσο ρίσκο έχει κάποιος κόμβος να "πιάσει" αυτό που κυκλοφορεί στο δίκτυο μας (π.χ. Ένας ιός ή μια πληροφορία). Στην περίπτωση των κατευθυνόμενων γράφων όταν οι συνδέσεις σχετίζονται με κάτι θετικό όπως για παράδειγμα μια σύνδεση φιλίας ή συνεργασίας, ο αριθμός των εισερχόμενων συνδέσεων μπορεί να ερμηνευθεί ως μέτρο δημοτικότητας (*Wikipedia*) και δείχνει το πρεστίτζ που έχει κάποιος κόμβος δηλαδή το κατά πόσο η οντότητα που αναπαρίσταται από τον κόμβο προκαλεί θαυμασμό, επιρροή ή εμπιστοσύνη. Από την άλλη, ο αριθμός των εξερχόμενων συνδέσεων μπορεί να ερμηνευθεί ως μέτρο της κοινωνικότητας του κόμβου.

Κεντρικότητα εγγύτητας (Closeness centrality)

Για να βρούμε την απόσταση δύο κόμβων σε έναν γράφο βρίσκουμε το μήκος του συντομότερου μονοπατιού. Το πόσο απομακρυσμένος είναι ένας κόμβος το βρίσκουμε αθροίζοντας τις αποστάσεις

του από όλους τους υπόλοιπους κόμβους. Το αντίστροφο της παραπάνω διαδικασίας μας δίνει την κεντρικότητα εγγύτητας. Έτσι όσο πιο κεντρικός θα είναι ένας κόμβος, τόσο χαμηλότερη θα είναι η συνολική του απόσταση από τους υπόλοιπους κόμβους. Μπορεί επίσης να θεωρηθεί ότι η κεντρικότητα εγγύτητας είναι ένας τρόπος μέτρησης του πόσο θα διαρκέσει η διάδοση της πληροφορίας από τον προς εξέταση κόμβο σε όλους τους άλλους, αν αυτή γίνει ακολουθιακά.

Μία ιδιότητα της κεντρικότητας εγγύτητας είναι ότι τείνει να αποδώσει υψηλό σκορ σε κόμβους που βρίσκονται κοντά στο κέντρο των τοπικών κοινωνιών (local clusters). Κοινωνιολογικά, μέλη μιας τοπικής κοινωνικής ομάδας με υψηλή κεντρικότητα εγγύτητας είναι αυτά τα μέλη που ασκούν μεγάλη επιρροή στα υπόλοιπα άτομα της κοινωνικής αυτής ομάδας. Συχνά, τα άτομα αυτά είναι δημόσιες φιγούρες σε ολόκληρο το δίκτυο, δέχονται τον σεβασμό της κοινωνικής ομάδας. Σε κάθε περίπτωση, "καταλαμβάνουν" τα συντομότερα μονοπάτια διάδοσης πληροφορίας.



Εικόνα 2-7 Οι κόμβοι που βρίσκονται στο κέντρο των τοπικών κοινωνιών έχουν υψηλότερο closeness centrality. Πηγή: activatenetworks.net.

Εκκεντρότητα Διαμεσότητας (Betweenness Centrality)

Η διαμεσότητα προτάθηκε από τον κοινωνιολόγο Linton Freeman (Freeman, Linton, 2004). Η διαμεσότητα ενός κόμβου ισούται με τον αριθμό των συντομότερων μονοπατιών από όλες τις κορυφές σε όλες τις υπόλοιπες που περνάνε από τον προς εξέταση κόμβο. Η διαμεσότητα είναι ακόμα πιο χρήσιμο μέτρο από την απλή συνδεσιμότητα (connectivity) όσον αφορά την μέτρηση τόσο του φορτίου όσο και της σπουδαιότητας του κόμβου. Η διαμεσότητα σε σχέση με την απλή συνδεσιμότητα είναι πιο γενικό και όχι τοπικό μέτρο.

Για να υπολογίσουμε την εκκεντρότητα διαμεσότητας πρέπει να ξεκινήσουμε βρίσκοντας όλα τα συντομότερα μονοπάτια μεταξύ όλων των δυνατών ζευγαριών κόμβων στο δίκτυο μας. Κατόπιν μετράμε πόσα από αυτά περνάνε από κάθε κόμβο. Οι αριθμοί αυτοί είναι η εκκεντρότητα διαμεσότητας του αντίστοιχου κόμβου.

Αν υποθέσουμε ότι κάθε κόμβος αναπαριστά ένα άτομο, το αποτέλεσμα του παραπάνω υπολογισμού είναι η εύρεση των ατόμων που είναι απαραίτητοι αγωγοί για την πληροφορία που διασχίζει διάφορα μέρη του δικτύου. Αυτοί συνήθως διαφέρουν από αυτούς με υψηλή κεντρικότητα εγγύτητας. Άτομα με υψηλή εκκεντρότητα διαμεσότητας δεν έχουν το κοντινότερο μέσο μονοπάτι σε οποιονδήποτε άλλον

κόμβο, αλλά έχουν τον μεγαλύτερο αριθμό συντομότερων μονοπατιών που πρέπει να περάσουν από αυτούς.

Σε ένα κοινωνικό δίκτυο κόμβοι με υψηλή εκκεντρότητα διαμεσότητας εμφανίζονται συχνά στις τομές των πιο πυκνά συνδεδεμένων κοινωνιών. Όπως βλέπουμε στην εικόνα 2.8 όλοι είναι τοποθετημένοι για να πάρουν ρόλο διαμεσολαβητή κατά μήκος των συστάδων, με την έννοια ότι οι διαμεσολαβητές συνδέουν άτομα τα οποία διαφορετικά θα ήταν ασύνδετα μεταξύ τους, αλλά πλέον επωφελούνται από την ανταλλαγή της πληροφορίας.

Στην πράξη άτομα με υψηλή εκκεντρότητα διαμεσότητας είναι συχνά κρίσιμα για συνεργασίες μεταξύ διαφορετικών τμημάτων και για να διατηρούν την διασπορά ενός προϊόντος σε όλο το δίκτυο. Εξαιτίας της θέσης τους είναι διαμεσολαβητές πληροφορίας και συνεργασίας. Η μόνη διαφορά τους με τους πραγματικούς μεσίτες είναι ότι οι δεύτεροι έχουν συνήθως ένα δημόσιο προφίλ σαν μέρος της εργασίας τους, ενώ οι πρώτοι συχνά παραβλέπονται. Αυτό συμβαίνει γιατί δεν είναι κεντρικοί σε καμία κοινωνική ομάδα, αντιθέτως κινούνται στην περιφέρεια πολλών κοινωνικών ομάδων σε καθεμία από τις οποίες προκαλούν εμπιστοσύνη και θαυμασμό.

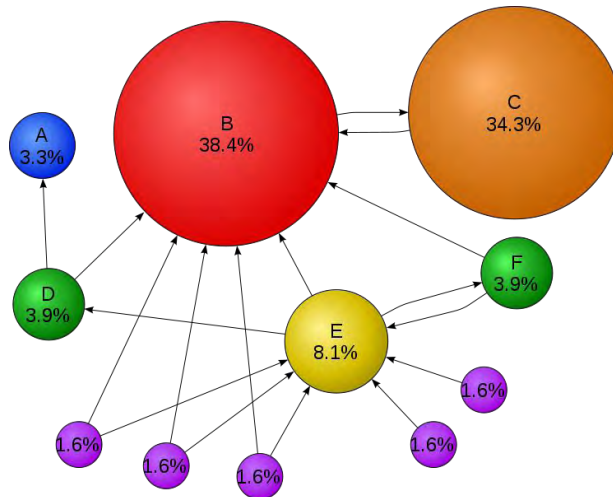


Εικόνα 2-8 Οι κόμβοι που συμπεριφέρονται ως γέφυρες μεταξύ των συστάδων του δικτύου έχουν υψηλότερη διαμεσότητα.
Πηγή: activatenetworks.net.

Κεντρικότητα Page Rank

Ο αλγόριθμος page rank, (*Wikipedia*) ο οποίος χρησιμοποιήθηκε από την μηχανή αναζήτησης της Google, αποδίδει ένα βάρος σε κάθε στοιχείο ενός υπερσυνδεδεμένου (hyperlinked) συνόλου εγγράφων, όπως για παράδειγμα το World Wide Web, με σκοπό την μέτρηση της σχετικής σημαντικότητας μέσα στο σύνολο. Ο αλγόριθμος μπορεί να εφαρμοσθεί σε κάθε γράφο, για μέτρηση της φασματικής κεντρικότητας αλλά χρησιμοποιείται ευρέως από τις μηχανές αναζήτησης για την κατάταξη των ιστοσελίδων στα αποτελέσματα αναζήτησης για κάθε δεδομένη επερώτηση.

Κάθε υπερσύνδεσμος σε μια σελίδα μετριέται ως ψήφος υποστήριξης. Το page rank μιας σελίδας ορίζεται αναδρομικά και εξαρτάται από τον αριθμό των σελίδων που αναφέρονται μέσω υπερσυνδέσμου σε αυτή αλλά και από την τιμή του page rank των σελίδων αυτών.

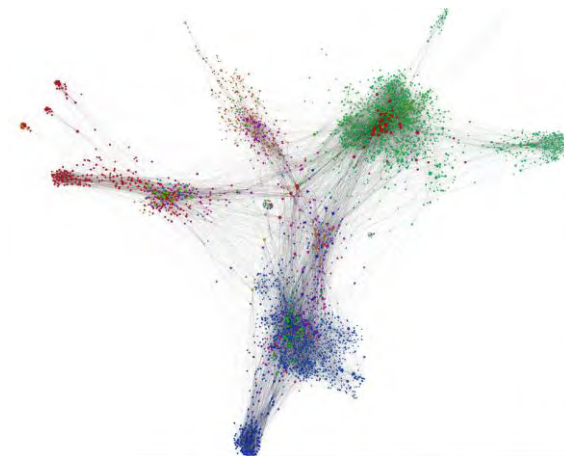


Εικόνα 2-9 Παρατηρούμε ότι η σελίδα C έχει υψηλότερο pagerank από την σελίδα E, παρόλο που οι υπερσύνδεσμοι προς την E είναι λιγότεροι από αυτούς προς την C. Αυτό συμβαίνει γιατί ο μοναδικός σύνδεσμος προς την C έρχεται από τον κόμβο B ο οποίος έχει μεγάλο pagerank. Πηγή: Wikipedia.

Διάρθρωση (Modularity)

Η διάρθρωση μετρά την δομή ενός γράφου. Πιο συγκεκριμένα, υπολογίζει την δύναμη της κάθε υποδιαίρεσης ενός γράφου σε κοινωνίες ή συστάδες (clusters). Δίκτυα με υψηλό modularity έχουν πυκνές συνδέσεις μεταξύ των κόμβων που ανήκουν σε μια συστάδα και αραιές συνδέσεις μεταξύ των κόμβων που ανήκουν σε διαφορετικές συστάδες.

Ο αλγόριθμος υπολογισμού του modularity χρησιμοποιείται για να ανακαλύψουμε την κοινωνική δομή των δικτύων. Μειονέκτημα του αλγορίθμου είναι ότι δεν μπορεί να εντοπίσει μικρές ομάδες, λόγω περιορισμού στην ανάλυση του. Βιολογικά δίκτυα όπως το ανθρώπινο μυαλό παρουσιάζουν πολύ υψηλό modularity.



Εικόνα 2-10 Σε έναν γράφο οι οντότητες συμβολίζονται ως κόμβοι και οι μεταξύ τους σχέσεις ως σύνδεσμοι. Στην εικόνα μας, οι κόμβοι χρωματίζονται με βάση την ομάδα στην οποία ανήκουν. Πηγή: Gephi.org.

2.4.4 Ιεραρχικά δεδομένα

Ένα **δέντρο** στη Θεωρία γράφων, είναι ένας μη κατευθυνόμενος Γράφος, στον οποίο οποιοσδήποτε δύο κορυφές συνδέονται με ένα και μόνο απλό μονοπάτι. Με άλλα λόγια κάθε συνεκτικός γράφος χωρίς κύκλους είναι ένα δέντρο.

Ιεραρχικά Δένδρα

Κατευθυνόμενα δένδρα με ένα σύνολο κόμβων-φύλλων (κόμβοι μοναδιαίου βαθμού) που αναπαριστούν ένα σύνολο αντικειμένων και ένα σύνολο κόμβων-γονέων που αναπαριστούν σχέσεις επί των αντικειμένων. Σε ένα ιεραρχικό δένδρο κάθε κόμβος, εκτός της ρίζας του δέντρου, έχει ακριβώς έναν γονέα.

Υπερβολικά Δένδρα (Hyper Trees)

Εμπνευσμένα από την υπερβολική γεωμετρία, απεικονίζουν ιεραρχικά δεδομένα, ενώ ο αριθμός των κόμβων σε κάθε επίπεδο μπορεί να αυξάνει εκθετικά. Χρησιμοποιούν τον Υπερβολικό Χώρο, ο οποίος ουσιαστικά έχει περισσότερο μέρος από τον Ευκλείδειο. Για παράδειγμα, μια γραμμική αύξηση μιας ακτίνας ενός κύκλου στον Ευκλείδειο χώρο αυξάνει την περιφέρεια γραμμικά, ενώ μια ανάλογη γραμμική αύξηση στον Υπερβολικό χώρο θα την αύξανε εκθετικά.



Εικόνα 2-11 Circle Limit IV. Ο M.C. Escher γύρω στο 1956 ασχολήθηκε με τις Υπερβολικές Ψηφιοδοθετήσεις, που είναι κανονικά πλακίδια στο Υπερβολικό επίπεδο. Για να παρουσιάσει την εργασία του δημιούργησε το 1959 τέσσερις ξυλογραφίες με την ονομασία Circle Limit. Πηγή: Wikipedia.

Σε ένα υπερβολικό δένδρο οι κόμβοι στους οποίους εστιάζουμε τοποθετούνται στο κέντρο και καταλαμβάνουν περισσότερο χώρο, ενώ οι κόμβοι εκτός εστίασης βρίσκονται συμπιεσμένοι κοντά στα όρια του χώρου. Αν εστιάσουμε σε κάποιον άλλον κόμβο, τον φέρνουμε με την σειρά του αυτόν στο

κέντρο και γύρω τα παιδιά του, ενώ οι κόμβοι που δεν μας ενδιαφέρουν συμπιέζονται καταλαμβάνοντας πολύ λίγο χώρο.

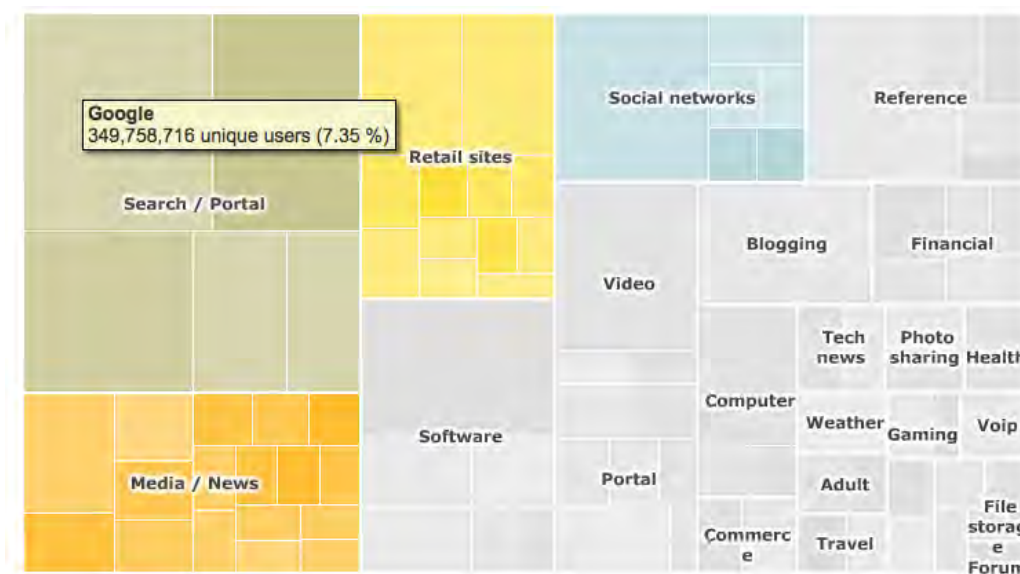
Τα υπερβολικά δένδρα κάνουν εύκολη την πλοήγηση σε μεγάλα δένδρα ακόμα και με εκατοντάδες οι χιλιάδες κόμβους. Με την περιστροφή και το "τράβηγμα" της απεικόνισης στον Καρτεσιανό χώρο μπορούμε εύκολα να αναζητήσουμε και να επιλέξουμε τους κόμβους που μας ενδιαφέρουν.

Δενδρικοί χάρτες (Treemaps)

Οι δενδρικοί χάρτες (treemaps) επινοήθηκαν από τον Ben Shneiderman (*Ben Shneiderman-University of Maryland, 2009*) σαν μια λύση στο πρόβλημα της αναπαράστασης ανάλογων τιμών και ιεραρχικών σχέσεων την ίδια στιγμή. Είναι ιδανικά στο αναπαριστούν μια πληθώρα ιεραρχικά φωλιασμένων τιμών δεδομένων.

Στους δενδρικούς χάρτες κάθε τερματικός κόμβος αναπαρίσταται ως ένα τετράγωνο. Η κύρια ιδέα είναι να διαμερίσουμε τον διαθέσιμο χώρο αναδρομικά με τον ίδιο τρόπο που ένα δένδρο διαμερίζει τα δεδομένα. Αυτό σημαίνει ότι οι δενδρικοί χάρτες είναι αναπαραστάσεις μοντέλου οδηγούμενες από τα δεδομένα.

Ο τετράγωνος χώρος του δενδρικού χάρτη ανταποκρίνεται σε ολόκληρο το σύνολο δεδομένων. Στο πρώτο βήμα ο χώρος διαμερίζεται οριζόντια σύμφωνα με τις αναλογίες των περιπτώσεων που περνάνε σε κάθε κόμβο παιδί. Στο επόμενο βήμα κάθε διαμέριση χωρίζεται κάθετα ανταποκρινόμενη στις αναλογίες των παιδιών της. Η διαδικασία αυτή επαναλαμβάνεται αναδρομικά με εναλλαγή κάθε φορά της κατεύθυνσης διαμέρισης από οριζόντια σε κάθετη και το αντίστροφο, έως ότου φτάσουμε στους τερματικούς κόμβους.



Εικόνα 2-12 Οπτικοποίηση του BBC με την κατανομή των 100 site με την μεγαλύτερη επισκεψιμότητα ανά θεματική κατηγορία (με βάση το χρώμα). Σε ένα διαδραστικό περιβάλλον ο χρήστης έχει την επιλογή να τοποθετήσει τον κέρσορα του ποντικιού πάνω από ένα τετράγωνο για να δει το όνομα του ιστότοπου και τον αριθμό των επισκέψεων. Πηγή: BBC.

Στο τελικό αποτέλεσμα κάθε τετράγωνο ανταποκρίνεται σε έναν τερματικό κόμβο. Η περιοχή που καλύπτει κάθε τετράγωνο είναι ανάλογη με τον αριθμό των περιπτώσεων που αντιστοιχούν σε αυτόν τον κόμβο.

Κατά την σχεδίαση μπορούμε να προσαρμόσουμε τα κενά μεταξύ των διαμερίσεων έτσι ώστε να αναδείξουμε το βάθος στο οποίο συνέβη η διαμέριση, δίνοντας μεγαλύτερα κενά σε όσες διαχωρίσεις έγιναν κοντινότερα της ρίζας του δέντρου.

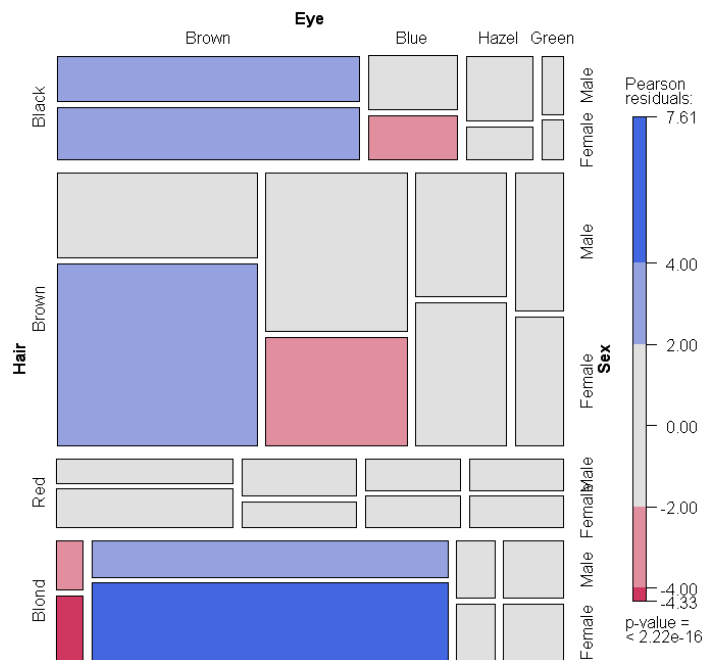
2.4.5 Δεδομένα πολλών διαστάσεων

Μωσαϊκό διάγραμμα (Mosaic Plot)

Το μωσαϊκό διάγραμμα στην ουσία είναι ένας δενδρικός χάρτης ενός δένδρου αποσύνθεσης, δηλαδή ενός δένδρου στο οποίο οι διαχωρίσεις με το ίδιο βάθος χρησιμοποιούν την ίδια κατηγορική μεταβλητή διαχώρισης και έχουν τόσα παιδιά όσα και οι κατηγορίες των δεδομένων.

Το μωσαϊκό διάγραμμα αποτελείται από ομάδες τετράγωνων πλακιδίων. Κάθε πλακίδιο αντιστοιχεί σε ένα κελί κάποιου πίνακα συνάφειας (contingency table). Η περιοχή του είναι ανάλογη του μεγέθους του κελιού. Το σχήμα και η τοποθεσία του καθορίζονται κατά την διαδικασία της κατασκευής του.

Τα μωσαϊκά διαγράμματα δεν περιορίζονται στις δύο διαστάσεις · μπορούν να επεκταθούν σε έναν αυθαίρετο αριθμό διαστάσεων. Στην πράξη όμως ο χώρος είναι ένας περιοριστικός παράγοντας.

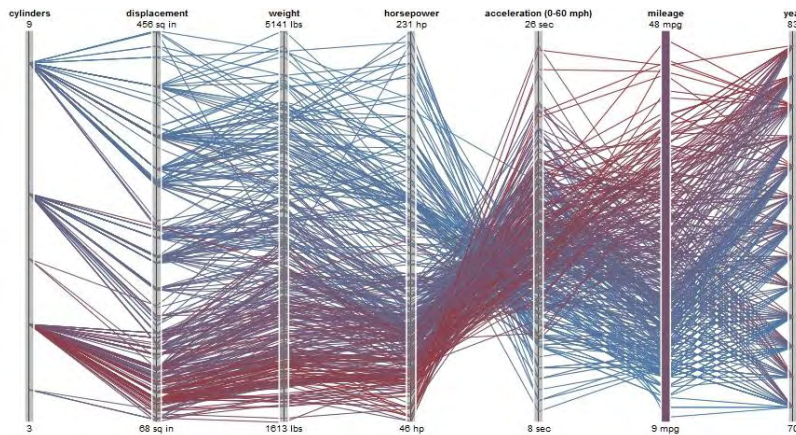


Εικόνα 2-13 Στο Mosaic plot της εικόνας τα κουτιά διαχωρίζονται ανάλογα με τις τιμές των τεσσάρων μεταβλητών: hair, sex, eye color, pearson residuals. Πηγή: <http://statmath.wu.ac.at/projects/vcd/>.

Παράλληλες συντεταγμένες (Paraller coordinates)

Κάθε παράλληλη συντεταγμένη αναπαριστά μία μεταβλητή. Παίρνοντάς την τομή κάθε γραμμής με κάθε παράλληλη συντεταγμένη βλέπουμε τις τιμές της υπό εξερεύνηση οντότητας για κάθε μεταβλητή.

Με τις παράλληλες συντεταγμένες η αναζήτηση για σχέσεις πολλαπλών μεταβλητών σε σύνολα δεδομένων πολλαπλών διαστάσεων μετασχηματίζεται σε ένα δυσδιάστατο πρόβλημα αναγνώρισης προτύπων.



Εικόνα 2-14 Κάθε παράλληλη συντεταγμένη αναπαριστά ένα μέτρο επίδοσης κάποιου αυτοκινήτου, ενώ κάθε γραμμή ένα συγκεκριμένο αυτοκίνητο. Σε ένα διαδραστικό περιβάλλον μπορούμε να φιλτράρουμε το πλήθος γραμμών προς εμφάνιση, με βάση το ποια αυτοκίνητα μας ενδιαφέρουν, για να κάνουμε ευκολότερα τις συγκρίσεις. Πηγή: <http://mbostock.github.io/protovis/ex/cars-full.html>.

2.4.6 Επιλογές επιστημονικών σχεδιασμών στην οπτικοποίηση δεδομένων

Ο τύπος των μεταβλητών προς σχεδίαση επηρεάζει σημαντικά την επιλογή του γραφήματος. Για παράδειγμα για συνεχείς μεταβλητές προτιμότερα είναι τα γραφήματα με γραμμές, τα ιστογράμματα και τα γραφήματα με κουτιά, ενώ για κατηγορικές μεταβλητές τα γραφήματα με ράβδους.

Το αν τα δεδομένα μας θα είναι συναθροισμένα ή μετασχηματισμένα εξαρτάται από την κατανομή τους και από το ποιοι είναι οι στόχοι της οπτικοποίησης.

Ο σχεδιασμός γραφημάτων πολλών μεταβλητών είναι πιο πολύπλοκος. Οι κύριες αποφάσεις που πρέπει να παρθούν είναι η μορφή της απεικόνισης και η επιλογή και η διάταξη των μεταβλητών. Γενικά μια εξαρτημένη μεταβλητή πρέπει να σχεδιάζεται τελευταία.

Το πρώτο βήμα είναι η επιλογή της κατάλληλης οπτικοποίησης. Η επόμενη πρόκληση είναι η σχεδίαση των οπτικών στοιχείων της απεικόνισης με τρόπο που το μήνυμά μας θα δηλώνεται με σαφήνεια, χωρίς να αποσπά την προσοχή σε ασήμαντα στοιχεία, αλλά αναδεικνύοντας ότι είναι το πιο σημαντικό.

Σύμφωνα με τον Edward Tufte, οι πίνακες και τα γραφήματα πρέπει να σχεδιάζονται με την χρήση δύο τύπων μελάνης: μελάνι δεδομένων (data ink) και μελάνι μη δεδομένων (non-data ink). Η αναλογία του μελανιού που χρησιμοποιείται για την απεικόνιση των δεδομένων προς το συνολικό μελάνι που εμφανίζεται στο περιβάλλον πρέπει να είναι υψηλή. Με άλλα λόγια, το μελάνι που χρησιμοποιείται για οτιδήποτε εκτός από δεδομένα θα πρέπει να μειωθεί στο ελάχιστο.

Ο Edward Tufte αναφέρθηκε επίσης και στην πυκνότητα των δεδομένων. Οι αριθμοί που εμφανίζονται σε ένα γράφημα μπορούν να οργανωθούν σε έναν πίνακα. Η πυκνότητα των δεδομένων ορίζεται ως η αναλογία του αριθμού των στοιχείων του παραπάνω πίνακα, ως προς τον χώρο που πιάνει το γράφημα. Στόχος μας είναι να μεγιστοποιήσουμε την πυκνότητα των δεδομένων, ώστε να δώσουμε ένα εννοιολογικό πλαίσιο και αξιοπιστία στα στατιστικά στοιχεία. Όταν ο χρήστης βλέπει ένα γράφημα με μικρή ποσότητα πληροφορίας μπορεί να υποπτευθεί ότι κάτι είναι κρυμμένο ή έχει παραληφθεί. Επίσης όταν έχουμε μεγάλη πυκνότητα δεδομένων μπορούμε να συγκρίνουμε διάφορα μέρη των δεδομένων παρουσιάζοντας μεγάλη ποσότητα πληροφορίας που μπορεί να γίνει αντιληπτή με μια ματιά. Όσο βέβαια η ποσότητα της πληροφορίας μεγαλώνει, τα μέτρα πρέπει να μικρύνουν, π.χ. να μικρύνουν οι τελείες σε ένα διάγραμμα σημείων ή να λεπτύνουν οι γραμμές σε ένα γράφημα χρόνου. Επίσης πρέπει να μεγαλώσει η αναλογία data-ink/no data ink. Εκτός όμως από το να αυξήσουμε την πληροφορία σε ένα γράφημα μπορούμε επίσης να μειώσουμε την περιοχή του γραφήματος (χωρίς να χαθεί πληροφορία ή να αλλοιωθεί η αναγνωσιμότητα) ώστε να αυξηθεί η πυκνότητα των δεδομένων.

2.4.7 Κλίμακες

Μια κλίμακα είναι ένας τύπος συνάρτησης που αντιστοιχίζει σύνολα μεταβλητών σε διαστάσεις.

Η επιλογή μιας κατατοπιστικής διάταξης έχει να κάνει με τον ορισμό της κλίμακας του άξονα κάποιας κατηγορικής μεταβλητής. Αυτό εξαρτάται από το τι αναπαριστούν οι κατηγορίες και τα σχετικά τους μεγέθη. Για μια συνεχή μεταβλητή ο ορισμός της κλίμακας γίνεται ακόμα πιο δύσκολος.

Στο "Grammar of Graphics" ο Wilkinson αναφέρεται (*Leland Wilkinson, 2005*) σε κάποιες ιδιότητες που μία καλή κλίμακα θα πρέπει να κατέχει: απλότητα, σταθμισμένη διασπορά, κάλυψη, ύπαρξη του μηδενός. Πρέπει να λαμβάνουμε υπόψη το τι αναπαριστούν τα δεδομένα και πως αυτά συλλέγονται. Σε κάθε περίπτωση ο χρήστης θα πρέπει να ελέγχει το αποτέλεσμα και να το τροποποιεί την κλίμακα αν αυτό κρίνεται σκόπιμο.

2.4.8 Επικεφαλίδες, λεζάντες, σχόλια

Ιδανικά, οι επικεφαλίδες πρέπει να εξηγούν πλήρως το γράφημα και να δηλώνουν την πηγή προέλευσης των δεδομένων. Οι λεζάντες περιγράφουν το πως τα σύμβολα ή τα χρώματα αντιστοιχίζονται με τις ομάδες των δεδομένων. Ο Tufte υποστήριξε ότι πρέπει να συμπεριλαμβάνονται απευθείας στο γράφημα και όχι να υπάρχουν ξεχωριστά, ώστε να μην χρειάζεται το μάτι του αναγνώστη να μετακινηθεί μπρος ή πίσω. Τα σχόλια χρησιμοποιούνται για να τονίσουν κάποια συγκεκριμένα χαρακτηριστικά του γραφήματος. Λόγω έλλειψης χώρου δεν μπορεί να είναι μεγάλα σε έκταση. Μπορούν να χρησιμοποιηθούν για να προσδιορίσουμε γεγονότα σε κάποιο διάγραμμα χρόνου ή για να τραβήξουμε την προσοχή σε συγκεκριμένα σημεία κάποιου διαγράμματος σημείων.

2.5 Το ανθρώπινο οπτικό σύστημα

Για να μπορέσουμε να παρουσιάσουμε την πληροφορία οπτικά με αποδοτικό τρόπο θα πρέπει πρώτα να κατανοήσουμε κάπως την οπτική αντίληψη του ανθρώπου. Το ανθρώπινο οπτικό σύστημα είναι μία αναζήτηση μοτίβων μεγάλης δύναμης και λεπτότητας. Το μάτι και ο οπτικός φλοιός του εγκεφάλου σχηματίζουν έναν μαζικά παράλληλο επεξεργαστή που παρέχει το κανάλι με το υψηλότερο εύρος

ζώνης στα ανθρώπινα κέντρα γνώσης. Αν κατανοήσουμε τον τρόπο με τον οποίο λειτουργεί η αντίληψή μας τότε η γνώση μας θα μπορέσει να μεταφραστεί σε κανόνες απεικόνισης πληροφορίας. Αν δεν υπακούσουμε στους κανόνες αυτούς τα δεδομένα μας θα είναι μη κατανοητά ή παραπλανητικά.

Αυτό που βλέπουμε δεν είναι απλά μια καταγραφή του τι υπάρχει στο περιβάλλον μας. Η όραση είναι μια ενεργή διαδικασία που συμπεριλαμβάνει διερμηνείες από τον εγκέφαλο μας πάνω σε δεδομένα που έχουν ανιχνευτεί από τα μάτια μας σε μια προσπάθεια να βγάλει νόημα από τα συμφραζόμενα. Ο εγκέφαλος μας αντιλαμβάνεται περισσότερο οπτικές διαφορές, παρά απόλυτες τιμές.

Τα μάτια μας ανιχνεύουν το φως που ανακλάται από μια επιφάνεια ενός αντικείμενου. Αυτό που αντιλαμβανόμαστε ως αντικείμενο έχει δημιουργηθεί ως μια σύνθεση πολλαπλών οπτικών ιδιοτήτων, τις οποίες μπορούμε να διαχωρίσουμε. Οι ιδιότητες αυτές είναι μοναδικά χαρακτηριστικά του φωτός, για τις οποίες τα μάτια μας είναι συντονισμένα να ανιχνεύουν. Κάποια από αυτά τα χαρακτηριστικά είναι η θέση στον δυσδιάστατο χώρο, το μέγεθος, το μήκος, η περιοχή, το σχήμα, το χρώμα, ο προσανατολισμός.



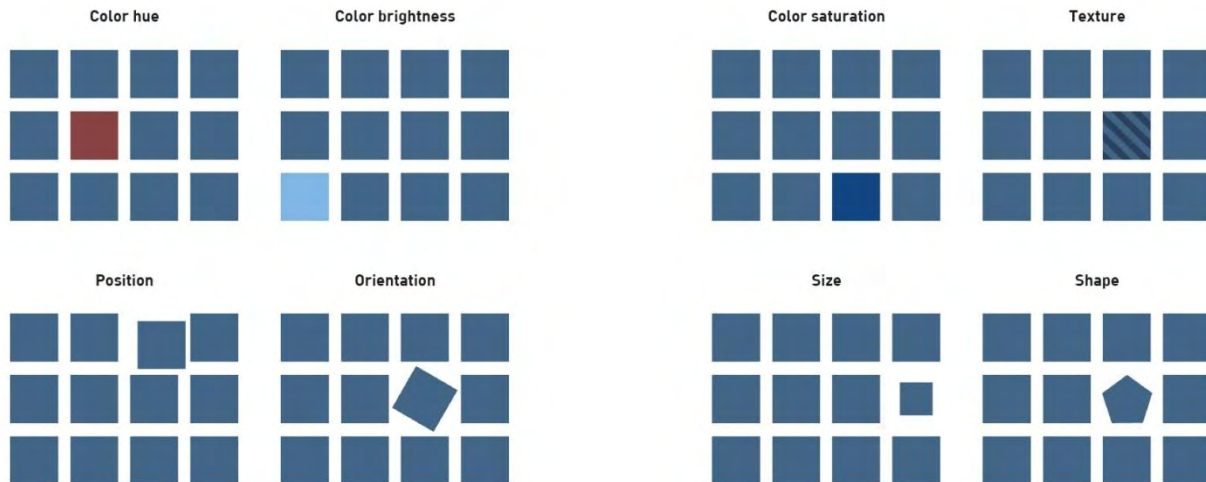
Εικόνα 2-15 Στο παράδειγμα της εικόνας, ενώ και τα τέσσερα αντικείμενα έχουν ακριβώς το ίδιο χρώμα (κατά απόλυτη τιμή), ο εγκέφαλος τα αντιλαμβάνεται ως διαφορετικά, λόγω της αλλαγής στην φωτεινότητα του φόντου. Στην συγκεκριμένη περίπτωση ο εγκέφαλος μας έχει αντιληφθεί την οπτική διαφορά του χρώματος του αντικείμενου με το χρώμα που το περιβάλλει. Μεταξύ των άλλων, ένα συμπέρασμα που μπορούμε να βγάλουμε μελετώντας το παραπάνω παράδειγμα είναι ότι πρέπει να αποφεύγουμε την χρήση χρωματικών διαβαθμίσεων (διανυσμάτων – color gradient) ως φόντο σε κάποιο γράφημα. Πηγή: Table and graph design for enlightening communication by Stephen Few. Perceptual Edge. Σελίδα 86

2.5.1 Preattentive μεταβλητές

Υπάρχουν κάποια συγκεκριμένα οπτικά χαρακτηριστικά τα οποία ο άνθρωπος αντιλαμβάνεται με την πρώτη ματιά, πριν ακόμα αυτός εστιάσει την προσοχή του και πριν ακόμα κάνει κάποια συνειδητή σκέψη. Η Jennifer Tidwell (*Jennifer Tidwell, 2010*) τις έχει ονομάσει "preattentive variables".

Με τον τρόπο αυτό και λαμβάνοντας υπ' όψιν ότι το ανθρώπινο μάτι διαθέτει εκατομμύρια υποδοχείς που δουλεύουν συντονισμένα και ταυτόχρονα, η preattentive αντίληψη είναι μία εξαιρετικά γρήγορη και παράλληλη διαδικασία. Όταν τα δεδομένα αναπαρίστανται με την χρήση κάποιων preattentive μεταβλητών, είμαστε σε θέση να αντιληφθούμε έναν μεγάλο όγκο πληροφορίας γρήγορα, αλλά και να

αποθηκεύσουμε ένα ολόκληρο αντικείμενο που αποτελείται από πολλές ιδιότητες, ως ένα ενιαίο κομμάτι στην μνήμη μας.



Εικόνα 2-16 Οι 8 preattentive μεταβλητές του Tidwell. Πηγή: *An introduction to visualizing data* by Joel Laumans (σελίδα 14).

2.5.2 Περισσότερη κωδικοποίηση (Redundant Encoding)

Αν μετά την κωδικοποίηση των κύριων διαστάσεων των δεδομένων μας, μας έχουν μείνει κάποιες οπτικές ιδιότητες αχρησιμοποίητες μπορούμε να τις χρησιμοποιήσουμε για να κωδικοποιήσουμε ήδη κωδικοποιημένες διαστάσεις. Το πλεονέκτημα αυτής της προσέγγισης είναι ότι χρησιμοποιώντας περισσότερα του ενός κανάλια μπορούμε να κατανοήσουμε την ίδια πληροφορία πιο εύκολα, γρήγορα και με περισσότερη ακρίβεια.

2.5.3 Κατηγοριοποίηση οπτικών ιδιοτήτων

Φυσική διάταξη (Natural Ordering)

Το αν μια οπτική ιδιότητα έχει φυσική διάταξη εξαρτάται από το αν οι μηχανισμοί του "λογισμικού" του εγκεφάλου μας και του οπτικού μας συστήματος, αυτόματα και χωρίς σκοπό, αποδίδουν μια ταξινόμηση ή μια κατάταξη στις διάφορες πιθανές τιμές αυτής της ιδιότητας. Το "λογισμικό" που κάνει αυτήν την κρίση βρίσκεται βαθιά ενσωματωμένο στο νου μας και αξιολογεί την σχετική διάταξη ανεξάρτητα από την γλώσσα ή την κουλτούρα.

Οπτικές ιδιότητες όπως η θέση, το μήκος, το πάχος μιας γραμμής, η γωνία, η έκταση μιας περιοχής, η πυκνότητα προτύπου, η φωτεινότητα και ο κορεσμός έχουν φυσική διάταξη, ενώ το σχήμα, η υφή (texture), το στυλ της γραμμής (συνεχής, με τελείες, διακεκομμένη) και το χρώμα δεν έχουν φυσική διάταξη.

Διακριτές τιμές (Distinct Values)

Ένας άλλος παράγοντας που πρέπει να λάβουμε σοβαρά υπ' όψιν για την επιλογή της κατάλληλης οπτικής ιδιότητας είναι πόσες διακριτές τιμές αυτή έχει ώστε ο αναγνώστης να είναι ικανός να ξεχωρίσει να τις αντιληφθεί και πιθανώς να τις θυμηθεί. Για παράδειγμα, αν και υπάρχουν άπειρα χρώματα ο αναγνώστης της οπτικοποίησης δεν μπορεί να τα διακρίνει αν είναι παρόμοια. Σε αντίθεση, μπορεί πιο εύκολα να ξεχωρίσει έναν μεγάλο αριθμό σχημάτων, θέσεων, αριθμών.

Όπως σημειώσαμε πριν το χρώμα δεν έχει φυσική διάταξη και για τον λόγο αυτό δεν πρέπει να το χρησιμοποιούμε για να ταξινομήσουμε, να διατάξουμε σε σειρά ή να κατατάξουμε τα δεδομένα μας. Μπορούμε όμως να χρησιμοποιήσουμε το χρώμα για να επισημάνουμε κατηγορικά δεδομένα. Βέβαια αυτό δεν μπορεί να συμβεί αν έχουμε μεγάλο αριθμό κατηγοριών, γιατί πολύ απλά θα χρειαστεί να χρησιμοποιήσουμε παρόμοια χρώματα που δεν μπορεί κάποιος να τα ξεχωρίσει εύκολα.

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

Εικόνα 2-17 Ο παραπάνω πίνακας των συνηθισμένων οπτικών ιδιοτήτων μπορεί να μας φανεί αρκετά χρήσιμος κατά την διαδικασία κωδικοποίησης για τους τύπους των δεδομένων μας. Πηγή: *Designing data visualizations by Noob Illinski & Julie Steele* (σελίδα 28).

2.6 Δυναμικές οπτικοποιήσεις δεδομένων

2.6.1 Διαδραστικότητα

Η διαδραστικότητα αναφέρεται στην δυνατότητα των χρηστών να ενεργούν επί της οπτικοποίησης, ελέγχοντας και επιλέγοντας τι και πως θα οπτικοποιηθεί. Όταν επιτρέπουμε στον χρήστη να αλληλεπιδράσει με τα δεδομένα, πλέον δεν έχουμε να κάνουμε μόνο με το πως οπτικοποιούνται τα δεδομένα αλλά και με το πως αυτά συμπεριφέρονται.

Υπάρχουν πολλές φόρμες διεπαφών που μπορούν να εφαρμοστούν πάνω σε οπτικοποιήσεις δεδομένων. Οι φόρμες αυτές μπορούν να ενταχθούν σε δύο κατηγορίες (*Galitz, 2002*):

- **Επιλογή δεδομένων και φιλτράρισμα:** Ο χρήστης ελέγχει ποια δεδομένα θα οπτικοποιηθούν. Αυτό τον βοηθά να εστιάσει στην αναζήτηση μόνο των δεδομένων που τον αφορούν, εμποδίζοντας με τον τρόπο αυτό την υπερφόρτωση της πληροφορίας.
- **Διευθέτηση και πλοήγηση δεδομένων:** Ο χρήστης επιλέγει τον τρόπο με τον οποίο παρουσιάζονται ή οπτικοποιούνται τα δεδομένα την επιλογή αυτή ο χρήστης μπορεί να βρει κάποιο νέο νόημα στα δεδομένα. Παρουσιάζοντας τα ίδια δεδομένα με διαφορετικό τρόπο, μπορούμε να οδηγηθούμε σε νέα συμπεράσματα αλλά και να δούμε διαφορετικές σχέσεις ανάμεσα στα δεδομένα.

Αν επιτρέψουμε τον χρήστη να ελέγχει τις δύο παραπάνω μεταβλητές, τότε έχουμε να κάνουμε με αποτελεσματικότερες οπτικοποιήσεις, αφού μπορούν να τον επικεντρώσουν στις δικές του ανάγκες.

Με βάση το πόσο διαδραστική είναι μία οπτικοποίηση μπορεί να ανήκει σε κάποια από τις παρακάτω κατηγορίες (*Joel Laumans*):

- **Στατική οπτικοποίηση:** Δεν αλλάζει καθόλου στο πέρασμα του χρόνου και δεν προσφέρει καμία διαδραστική λειτουργικότητα.
- **Κινούμενη οπτικοποίηση:** Μεταβάλλεται στο πέρασμα του χρόνου. Χρησιμοποιεί τον χρόνο ως μια επιπλέον μεταβλητή ή διάσταση.
- **Διαδραστική οπτικοποίηση:** Ο χρήστης μπορεί να αλληλεπιδράσει με την οπτικοποίηση επιλέγοντας ποια δεδομένα θα οπτικοποιηθούν και με ποιον τρόπο. Ωστόσο τα αποτελέσματα της αλληλεπίδρασης δεν είναι άμεσα ορατά.
- **Απευθείας χειρισμός (Direct Manipulation):** Ο χρήστης έχει τις ίδιες δυνατότητες όπως με μια διαδραστική οπτικοποίηση, με την μόνη διαφορά ότι στην περίπτωση του απευθείας χειρισμού τα αποτελέσματα της αλληλεπίδρασης γίνονται άμεσα ορατά.

2.6.2 Συστήματα πραγματικού χρόνου

Επειδή τα δεδομένα έρχονται από τον πραγματικό κόσμο, όπου δεν υπάρχουν απόλυτα, δεν είναι σταθερές τιμές προς ανάλυση, αντιθέτως αποτελούν κινούμενο στόχο. Για παράδειγμα, η θερμοκρασία αλλάζει συνεχώς, η επισκεψιμότητα ενός ιστότοπου αυξάνεται με την κυκλοφορία ενός νέου προϊόντο, κλπ.. Κρίνεται λοιπόν αναγκαίο να δημιουργήσουμε αναπαραστάσεις δεδομένων που αναπροσαρμόζονται με νέες τιμές κάθε εβδομάδα, ώρα, δευτερόλεπτο κ.ο.κ. .

2.6.3 Ροή Γραφήματος (Graph Stream)

Οι δυναμικοί γράφοι μοντελοποιούνται ως μια ροή γεγονότων που συμβαίνουν στον γράφο. Τα γεγονότα αυτά μπορεί να αφορούν είτε την δομή του γράφου (προσθήκη ή αφαίρεση στοιχείων) είτε τις ιδιότητες των στοιχείων. Ένα stream είναι η διασύνδεση μεταξύ μιας πηγής γεγονότων και ενός αποδέκτη. Μια πηγή γεγονότων μπορεί να είναι για παράδειγμα ένα αρχείο γεγονότων ή ένας αλγόριθμος παραγωγής γράφου.

Το Graph Stream παρέχει επίσης μερικά χαρακτηριστικά οπτικοποίησης γράφων. Η απόδοση των στοιχείων μπορεί να προσαρμόζεται μέσω κανόνων CSS μορφοποίησης. Με τον τρόπο αυτό ο θεατής της οπτικοποίησης δεν βλέπει απλά μια στατική οπτικοποίηση αλλά έναν γράφο που αλλάζει συνεχώς. Ο θεατής κοιτώντας τον γράφο μπορεί να παρατηρήσει την ακολουθία των γεγονότων και να βγάλει συμπεράσματα για το πως το δίκτυο εξελίσσεται στον χρόνο.

Για την υλοποίηση του Graph Stream μπορούμε να χρησιμοποιήσουμε την ομώνυμη βιβλιοθήκη της Java. Στο λογισμικό Gerhi, που θα παρουσιάσουμε αναλυτικά στην επόμενη ενότητα, μπορούμε να χρησιμοποιήσουμε το πρόσθετο λογισμικό Graph Streaming API.

2.7 Εργαλεία οπτικοποίησης δεδομένων

2.7.1 SVG, HTML5 Canvas

Αμφότερα τα Scalable Vector Graphics (SVG) και το Canvas της HTML 5, είναι τεχνολογίες που επιτρέπουν την δημιουργία πλούσιων γραφικών τα οποία μπορούν να αναπαρασταθούν μέσα στον φυλλομετρητή.

SVG

Το SVG είναι ένα φορμάτ αρχείου αναπαράστασης διανυσματικών γραφικών, το οποίο έχει ως βάση την XML. Το περιεχόμενο του SVG μπορεί να είναι στατικό, δυναμικό διαδραστικό ή κινούμενο. Μπορούμε να μορφοποιήσουμε το SVG με χρήση CSS αλλά και να του δώσουμε δυναμική συμπεριφορά χρησιμοποιώντας το SVG DOM. Με το SVG μπορούμε να αναπτύξουμε υψηλά διαδραστικές εφαρμογές στο web με χρήση scripting, εξελιγμένων γεγονότων κίνησης, φίλτρων και άλλων στοιχείων.

HTML 5 Canvas

Οι προδιαγραφές του HTML5 Canvas είναι ένα ευέλικτο Javascript API που μας επιτρέπει να προγραμματίσουμε πράξεις σχεδιασμού. Μέσα στο αντικείμενο του καμβά που ορίζεται από την ετικέτα `<canvas>` της HTML5 μπορούμε να σχεδιάσουμε με κώδικα script το περιεχόμενο είτε στις δύο διαστάσεις είτε στις 3 διαστάσεις (με χρήση WebGL). Δεν υπάρχει κάποιο φορμάτ αρχείου ούτε κάποιο ιεραρχικό σχήμα DOM για τα σχήματα που σχεδιάζουμε. Όλα είναι στην επιφάνεια ως pixel.

Αξίζει να σημειωθεί ότι δεν υπάρχουν χαρακτηριστικά που δίνουν κίνηση στα σχέδια, απλά εμείς σχεδιάζουμε όποτε και όπως θέλουμε. Μέσα στην επιφάνεια του Canvas μπορούμε να αναπτύξουμε πράξεις διαχείρισης pixel, όπως φίλτρα εικόνας. Μπορούμε επίσης να εισάγουμε εικόνες π.χ. σε μορφή png ή jpg. Η έξοδος του Canvas μπορεί να εξαχθεί σε μορφή εικόνας.

Σύγκριση Html5 Canvas με SVG

Αν σχεδιάσουμε έναν κύκλο με Html5 Canvas και SVG θα δούμε ότι στην δεύτερη περίπτωση αναπτύξαμε αρκετά λιγότερες γραμμές κώδικα. Θα δούμε επίσης ότι η SVG έκδοση γίνεται πιο εύκολα κατανοητή. Σε περίπτωση που πρέπει να αναπτύξουμε ένα κινούμενο σχέδιο φαίνεται ότι αυτό μπορεί να γίνει πιο εύκολα με SVG αφού αυτή υποστηρίζει δηλωτικά κινούμενα σχέδια, ενώ με το Canvas θα έπρεπε να χρησιμοποιήσουμε μετρητές χρόνου.

Από την άλλη, όταν θέλουμε να αναπαραστήσουμε δυναμική πληροφορία, όπως διαδραστικά γραφήματα ή διαδραστική επεξεργασία εικόνας (π.χ. με χρήση φίλτρων) πιο κατάλληλη είναι η χρήση του Canvas.

2.7.2 Βιβλιοθήκες Javascript

Η javascript είναι η πιο δημοφιλής scripting language στο web. Όλες οι μοντέρνες ιστοσελίδες χρησιμοποιούν την javascript για να προσθέσουν λειτουργικότητα όπως αξιολόγηση φόρμας, επικοινωνία με εξυπηρετητές, εντοπισμός γεωγραφικής θέσης, sockets, canvas, ήχος/βίντεο, offline αποθήκευση, web workers. Αφού ο φυλλομετρητής φορτώσει τα αρχεία html και css μία σελίδας στην συνέχεια φορτώνει τους κώδικες σε javascript.

Στην συνέχεια θα εξετάσουμε κάποιες βιβλιοθήκες για την javascript που χρησιμοποιούνται για την οπτικοποίηση των δεδομένων:

D3.js Data-Driven-Documents

Με την d3 μπορούμε να δώσουμε ζωή στα δεδομένα μας χρησιμοποιώντας html, canvas και css. Η d3 δίνει έμφαση στα web standards και για τον λόγο αυτό εκμεταλλεύεται όλες τις δυνατότητες των μοντέρνων φυλλομετρητών. Συνδυάζει συστατικά δυναμικών οπτικοποιήσεων και μια προσέγγιση για τον χειρισμό του DOM ([Document Object Model](#)) οδηγούμενη από τα δεδομένα.

Η d3 μας επιτρέπει να συνδέσουμε αυθαίρετα δεδομένα σε ένα DOM και εν συνεχεία να εφαρμόσουμε μετασχηματισμούς οδηγούμενους από τα δεδομένα στο έγγραφο μας. Για παράδειγμα, μπορούμε να χρησιμοποιήσουμε την d3 για να δημιουργήσουμε έναν html πίνακα από ένα αριθμητικό διάλυμα ή να χρησιμοποιήσουμε το διάλυμα αυτό για να δημιουργήσουμε ένα διαδραστικό διάγραμμα ράβδων.

Η d3 λύνει το πρόβλημα της αποδοτικής διαχείρισης εγγράφων βασισμένων σε δεδομένα, είναι γρήγορη, διευκολύνει την επαναχρησιμοποίηση κώδικα, υποστηρίζει σύνολα δεδομένων μεγάλου μεγέθους και δυναμικές συμπεριφορές για αλληλεπίδραση και κινούμενα σχέδια.

InfoVis Toolkit

Αναπτύχθηκε από τον [Nicolas Garcia Belmonte](#). Είναι γραμμένη σε Java και παρέχει εργαλεία για την δημιουργία διαδραστικών οπτικοποιήσεων δεδομένων και γραφικών στο web. Κύρια χαρακτηριστικά της είναι:

- Ενοποιημένη δομή δεδομένων. Η κύρια δομή της είναι ένας πίνακας κάθε στήλη του οποίου περιέχει αντικείμενα ομοιογενών τύπων όπως ακέραιοι ή αλφαριθμητικά. Τα δέντρα και οι γράφοι προέρχονται από τους πίνακες.
- Μικρές απαιτήσεις σε μνήμη, λόγω της χρήσης ομοιογενών στηλών, όπως περιγράφηκε από πάνω.

- Ενοποιημένο σύνολο διαδραστικών στοιχείων. Το διαδραστικό φιλτράρισμα, δηλαδή στην ουσία οι δυναμικές επερωτήσεις μπορούν να πραγματοποιηθούν με τα ίδια αντικείμενα ελέγχου και δομικά συστατικά, ανεξάρτητα της δομής δεδομένων. Κάτι τέτοιο απλοποιεί την επαναχρησιμοποίηση ήδη υπαρχόντων δομικών συστατικών και την σχεδίαση πιο γενικών συστημάτων
- Υψηλή ταχύτητα. Η βιβλιοθήκη στηρίζεται στο API της OpenGL για γραφικά επιταχυνόμενα από το υλικό. Έτσι σε ένα μηχάνημα που υποστηρίζει την επιτάχυνση μέσω του υλικού κάποιες οπτικοποιήσεις μπορούν να απεικονίζονται ακόμα και 100 φορές πιο γρήγορα, συγκριτικά με μία κανονική εφαρμογή Java2D.
- Επεκτασιμότητα.

Highcharts.js

Βιβλιοθήκη για εύκολη δημιουργία διαγραμμάτων, η οποία προσφέρει παράλληλα έναν εύκολο τρόπο για να ενσωματώσει κάποιος τα διαγράμματα αυτά στην σελίδα ή στην web εφαρμογή του. Δεν απαιτεί ειδικές γνώσεις προγραμματισμού για ρύθμιση των επιλογών. Οι επιλογές δίνονται σε μία δομή σημειογραφίας αντικειμένων javascript που συντάσσεται ως ένα σύνολο αντιστοιχίσεων κλειδιών – τιμών, συνδεδεμένων με άνω-κάτω τελεία, χωρισμένων με κόμμα και ομαδοποιημένων με αγκύλες. Είναι επίσης δυναμική, με την έννοια ότι μετά την δημιουργία του διαγράμματος, μπορούμε εύκολα να προσθέσουμε, να διαγράψουμε και να τροποποιήσουμε σημεία, γραμμές, ακόμα και άξονες. Με την highcharts.js μπορούμε εύκολα να προσθέσουμε επικεφαλίδες, λεζάντες και διαδραστικά σχόλια στα διαγράμματα μας. Επιπλέον, έχουμε την δυνατότητα να φορτώσουμε δεδομένα από έναν πίνακα javascript ο οποίος μπορεί να ορίζεται τοπικά, σε διαφορετικό αρχείο ή ακόμη και σε κάποιον άλλον ιστότοπο. Τα δεδομένα μπορούν να έχουν οποιαδήποτε μορφή και με μια συνάρτηση επιστροφής κλήσης (callback function) να αναλύσουμε τα δεδομένα αυτά σε έναν πίνακα.

Υποστηρίζει μια πληθώρα διαγραμμάτων, υποστηρίζεται από όλους τους μοντέρνους φυλλομετρητές αλλά και σε κινητές συσκευές και είναι δωρεάν για μη εμπορικούς σκοπούς.

Arbor.js

Βιβλιοθήκη οπτικοποίησης γράφων που χρησιμοποιεί web workers και jQuery. Η βιβλιοθήκη αυτή δεν έχει σχεδιαστεί ως ένα πλαίσιο που συμπεριλαμβάνει όλες τις λειτουργικότητες, αντιθέτως παρέχει έναν αποδοτικό και κατευθυνόμενο με δυνάμεις αλγόριθμο διάταξης. Επιπλέον περιλαμβάνει αφαιρέσεις για οργάνωση του γράφου και διαχείριση της ανανέωσης των όσων παρουσιάζονται στην οθόνη.

Ένας χρήστης της βιβλιοθήκης μπορεί, με παράλληλη χρήση Canvas, SVG ή HTML στοιχείων, να απεικονίσει οποιαδήποτε προσέγγιση είναι κατάλληλη για την εργασία του και ικανοποιεί τις ανάγκες του. Ως αποτέλεσμα, με την χρήση της Arbor μπορεί κάποιος να εστιαστεί στα δεδομένα του γράφου και το οπτικό στυλ, χωρίς να χάνει χρόνο στον καθορισμό και επεξεργασία της φυσική και μαθηματική δομή της σχεδίασης.

Sigma.js

Ελαφριά βιβλιοθήκη ανοιχτού κώδικα για σχεδίαση γράφων, χρησιμοποιώντας το στοιχείο canvas. Η βιβλιοθήκη αυτή έχει σχεδιαστεί κυρίως για οπτικοποίηση διαδραστικά στατικών γράφων τα οποία έχουν εξαχθεί από ένα λογισμικό οπτικοποίησης γράφων όπως το *Gerphi*, αλλά και για οπτικοποίησης

δυναμικών γράφων που παράγονται καθώς ο γράφος κινείται. Είναι εύκολη στην χρήση και υψηλά παραμετροποιήσιμη.

Cubism.js

Πρόσθετο λογισμικό για την d3 για οπτικοποίηση χρονολογικών δεδομένων.

RickShaw

Εργαλειοθήκη για οπτικοποίηση διαδραστικών γραφημάτων χρόνου, βασισμένη στην d3. Με την βιβλιοθήκη αυτή μπορεί κάποιος να δημιουργήσει εύκολα στοιχεία αλληλεπίδρασης όπως λεζάντες, επικάλυψη ποντικίου, επιλογείς εύρους κτλ.

Raphael

Απλοποιεί εργασίες που έχουν να κάνουν με διανυσματικά γραφικά στο web. Χρησιμοποιεί την SVG W3C Recommendation και την VML σαν βάση δημιουργίας γραφικών. Αυτό σημαίνει ότι κάθε γραφικό αντικείμενο που δημιουργείται είναι επίσης ένα αντικείμενο DOM και έτσι μπορούμε να επισυνάψουμε άλλες javascript βιβλιοθήκες ή να το τροποποιήσουμε αργότερα. Σκοπός της Raphael είναι να παρέχει έναν προσαρμογέα που θα κάνει την σχεδίαση διανυσματικής τέχνης συμβατή με όλους τους φυλλομετρητές και εύκολη.

Leaflet

Βιβλιοθήκη ανοιχτού κώδικα για δημιουργία διαδραστικών χαρτών φιλικών σε κινητές συσκευές.

2.7.3 Η γλώσσα προγραμματισμού R

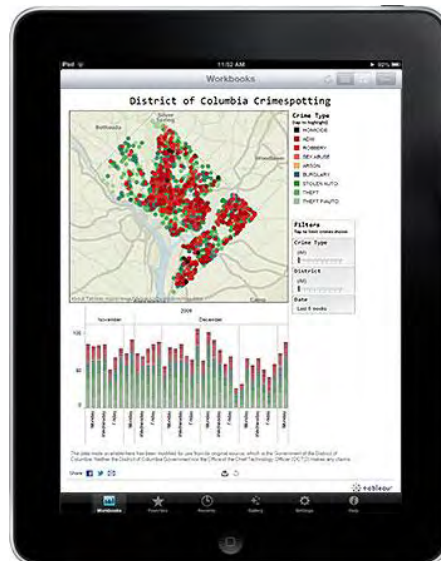
Η R είναι μία ελεύθερη γλώσσα προγραμματισμού και ένα προγραμματιστικό περιβάλλον για την υλοποίηση στατιστικών υπολογισμών και γραφημάτων. Η γλώσσα R χρησιμοποιείται ευρέως από τους στατιστικολόγους και τους αναλυτές δεδομένων για ανάπτυξη λογισμικών στατιστικής και ανάλυσης δεδομένων.

Η R παρέχει μία μεγάλη ποικιλία στατιστικών και γραφικών τεχνικών όπως γραμμική ή μη γραμμική μοντελοποίηση, κλασικά στατιστικά τεστ, ανάλυση time-series, κατηγοριοποίηση, συσταδοποίηση κ.α. . Είναι εύκολα επεκτάσιμη μέσω συναρτήσεων ή άλλων επεκτάσεων και έχει ένα μεγάλο και ενεργή κοινότητα χρηστών. Η γλώσσα R είναι περισσότερο αντικειμενοστραφής από κάθε άλλη γλώσσα στατιστικών υπολογισμών και χρησιμοποιεί κανόνες τοπικής εμβέλειας για τα ονόματα των μεταβλητών, έχει το δικό της φORMÁT για μαθηματικές εκφράσεις και σύμβολα, εμπεριέχει στατικά γραφήματα και επιτρέπει την εγκατάσταση πρόσθετων πακέτων λογισμικού για την υλοποίηση δυναμικών και διαδραστικών γραφημάτων.

2.7.4 Tableau

Με το Tableau Public μπορεί κάποιος να δημιουργήσει καλαίσθητες και διαδραστικές οπτικοποιήσεις δεδομένων χωρίς να έχει κάποια γνώση προγραμματισμού. Μέσω της καλά σχεδιασμένης διεπαφής του μπορεί κάποιος εύκολα να εισάγει δεδομένα. Κατόπιν θα αναγνωριστούν οι διαστάσεις των δεδομένων, τις οποίες ο χρήστης μπορεί να κωδικοποιήσει αποδίδοντας τις τα οπτικά χαρακτηριστικά που εκείνος ορίζει. Οι οπτικοποιήσεις μπορούν να ενσωματωθούν σε μια ιστοσελίδα με παρόμοιο τρόπο όπως ενσωματώνεται ένα βίντεο του Youtube. Το Tableau Public παρέχει μια μεγάλη συλλογή τύπων οπτικοποιήσεων.

Διατίθενται επίσης για εμπορική χρήση και εμπορικά λογισμικά όπως το Tableau Desktop και το Tableau Server με τα οποία κάποιος μπορεί με εύκολο τρόπο να συνδέει τις οπτικοποιήσεις του, σε πραγματικό χρόνο, με δεδομένα που βρίσκονται σε διάφορες πηγές.

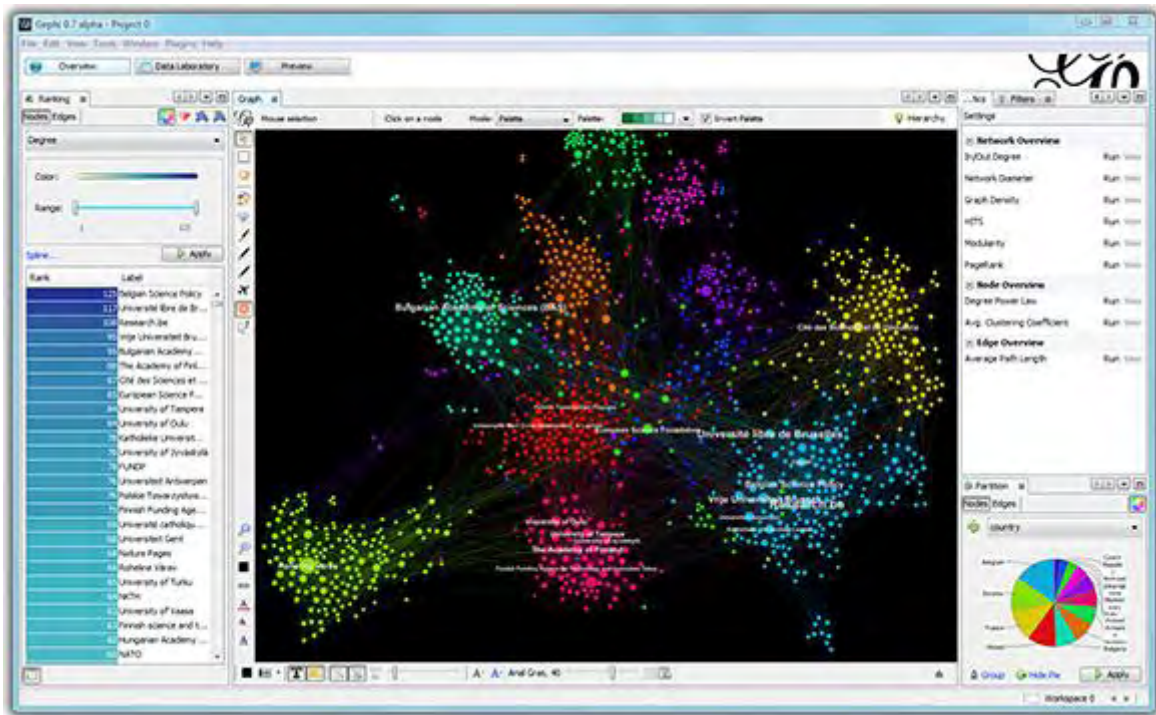


2-18 Με το Tableau Public μπορεί να δημιουργήσει κανείς, αυτόματα, οπτικοποιήσεις βελτιστοποιημένες για οθόνης αφής και χρήση μέσω iPad ή Android Tablets. Πηγή: www.tableausoftware.com.

2.7.5 Gephi

Το Gephi είναι ένα λογισμικό ανοικτού κώδικα για ανάλυση και οπτικοποίηση δικτύων και γράφων. Αποτελεί μία διαδραστική πλατφόρμα εξερεύνησης, οπτικοποίησης, ανάλυσης, χωροθέτησης, φιλτραρίσματος και διαχείρισης για όλα τα είδη των δικτύων, ιεραρχικών και δυναμικών γράφων. Ο χρήστης της πλατφόρμας μπορεί να διαχειριστεί δομές, σχήματα και χρώματα για να αναδείξει κάποιες κρυμμένες ιδιότητες. Το Gephi χρησιμοποιεί μια τρισδιάστατη μηχανή απόδοσης για να αναπαραστήσει μεγάλους γράφους σε πραγματικό χρόνο και για να κάνει γρήγορη την εξερεύνηση.

Στο Gephi υπάρχουν ενσωματωμένοι αλγόριθμοι για απόδοση μορφής και σχήματος στον γράφο (π.χ. τεχνικές κατευθυνόμενες με δυνάμεις και αλγόριθμοι πολλών επιπέδων). Παρέχει επίσης ένα πρόγραμμα-πλαίσιο (framework) για στατιστικές και μετρήσεις, δυναμική ανάλυση του δικτύου και graph streaming, δυνατότητες εξαγωγής του γράφου σε διάφορες μορφές, δυναμικό φιλτράρισμα, δυνατότητα εξερεύνησης.



2-19 Το περιβάλλον εργασίας στο Overview του Gephi.

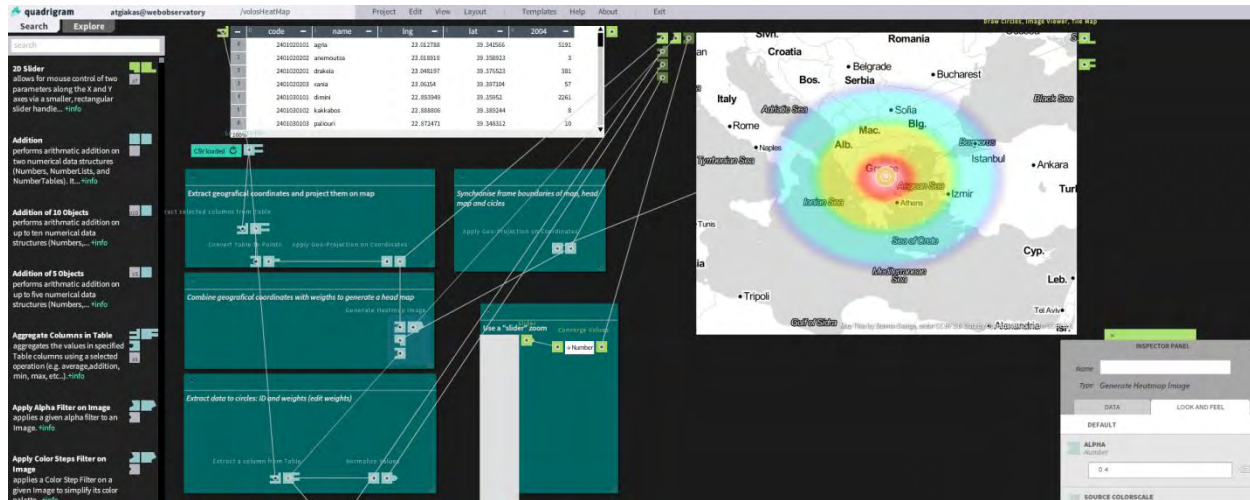
2.7.6 Quadrigram

Το Quadrigram είναι ένα εμπορικό λογισμικό και προγραμματιστικό περιβάλλον με την χρήση του οποίου μπορούμε να δημιουργήσουμε, να συλλέξουμε και να διαμοιραστούμε διαδραστικές και αρκετά εξατομικευμένες οπτικοποιήσεις δεδομένων. Σκοπός του είναι η γρήγορη δημιουργία και δημοσίευση οπτικοποιήσεων δεδομένων.

Το Quadrigram αποτελείται από λειτουργικές μονάδες οι οποίες συνίστανται από κάποιες παραμέτρους οι τιμές των οποίων ορίζονται από τον χρήστη, δέχονται μία ή περισσότερες εισόδους κάνουν κάποια πράξη επί των εισόδων και παράγουν μία ή περισσότερες εξόδους. Οι λειτουργικές μονάδες είναι συνδεδεμένες μεταξύ τους και με τον τρόπο αυτό παράγουν ένα διάγραμμα ροής εργασιών μέσα σε έναν μαύρο καμβά.

Για να χρησιμοποιήσει κάποιος το Quadrigram θα πρέπει να κάνει έναν λογαριασμό και να αγοράσει το προϊόν. Μόλις το κάνει αυτό θα του δοθεί πρόσβαση στην web εφαρμογή που μπορεί να φορτώσει από κάποιον φυλλομετρητή που έχει εγκατεστημένο κάποια πρόσφατη έκδοση του flash player. Η επιφάνεια εργασίας της εφαρμογής αποτελείται από τον απεριόριστο μαύρο καμβά πάνω στον οποίο ο χρήστης μπορεί να τοποθετήσει και να ενώσει τις λειτουργικές μονάδες, από δύο μενού, από τον εξερευνητή λειτουργικών μονάδων και από τον πίνακα επιθεώρησης (inspector panel) στον οποίον ο χρήστης μπορεί να χειριστεί τις παραμέτρους και τις εισόδους. Ο χρήστης επιλέγει τις πηγές των δεδομένων του που μπορεί να είναι μη ομοιογενείς, κάποια αρχεία σε κάποιους εξυπηρετητές, ένα API, ή μια βάση δεδομένων. Με τις λειτουργικές μονάδες ο χρήστης μπορεί να επεξεργαστεί τα δεδομένα και μέσω των οπτικοποιητών μπορεί να εμφανίσει τα αποτελέσματα της κάθε επί μέρους επεξεργασίας

σε κάποιο γράφημα και να επιλέξει αν θέλει να το δημοσιοποιήσει εξαγοντας το στον Web Server του Quadrigram.



2-20 Το περιβάλλον εργασίας του Quadrigram.

2.7.7 Visual.ly, infoogr.am

Στις ιστοσελίδες visual.ly και infoogr.am μπορεί κάποιος να χρησιμοποιήσει τις web εφαρμογές για εύκολη και γρήγορη δημιουργία γραφημάτων πληροφοριών.

Στην πρώτη περίπτωση ο χρήστης επιλέγει από μία συλλογή infographics κάποια από τις διαθέσιμες ιστορίες που βρίσκονται σε μορφή προτύπου και κατόπιν αφού συνδεθεί μέσω facebook ή twitter μπορεί να χρησιμοποιήσει κάποια εφαρμογή που θα του επιτρέψει να προσαρμόσει την επιλεγείσα ιστορία στα δικά του μέτρα είτε χειροκίνητα είτε αυτόματα αφού η εφαρμογή μπορεί να αντλεί στοιχεία του προφίλ του.

Στην δεύτερη περίπτωση ο χρήστης πάλι επιλέγει ένα πρότυπο (σε αυτήν την περίπτωση το πρότυπο αφορά μόνο τον σχεδιασμό και όχι την ιστορία) και κατόπιν μέσω μιας web εφαρμογής μπορεί να δηλώσει την πηγή των δεδομένων του, να εισάγει δεδομένα χειροκίνητα, να επεξεργαστεί τα δεδομένα, να παραμετροποιήσει τα κείμενα, τα χρώματα, τις γραμματοσειρές και γενικά το στυλ και να χρησιμοποιήσει κάποια από τα περίπου 30 είδη γραφημάτων που εμπεριέχονται στην εφαρμογή.

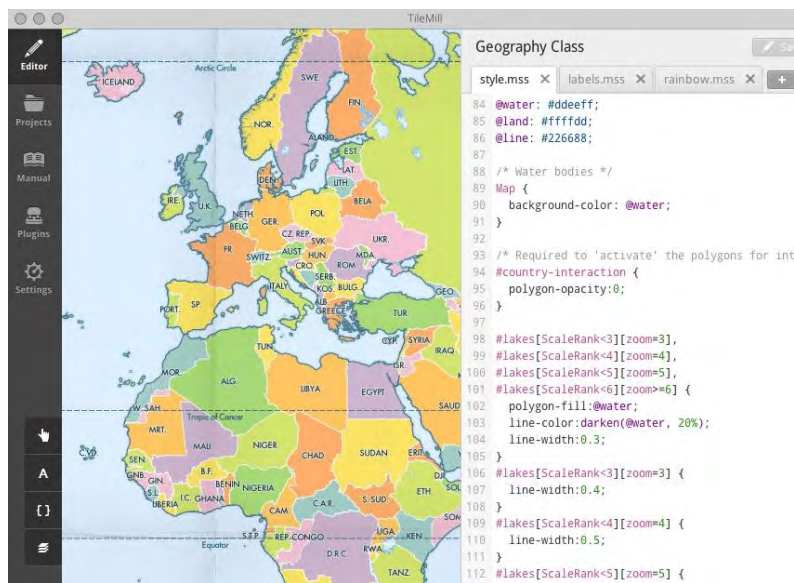
2.7.8 Map Box

Το εργαλείο αυτό έχει ως σκοπό την εύκολη δημιουργία και δημοσίευση προσαρμοσμένων χαρτών. Χρησιμοποιεί τον υψηλής ακρίβειας και συνεχώς ενημερωμένο παγκόσμιο χάρτη του Open Street Map Project, το οποίο είναι μια ανοιχτή ομαδική εθελοντική προσπάθεια για δημιουργία ενός χάρτη που είναι ελεύθερα διαθέσιμος προς όλους. Ο χάρτης αυτός έχει σχεδιαστεί έτσι ώστε να μπορεί κάποιος εύκολα να αλλάξει τα χρώματα, να δείξει ή να κρύψει κάποια χαρακτηριστικά, να επεξεργαστεί τις λεζάντες, να προσαρμόσει το σχήμα του εδάφους ή να προσθέσει κάποιο texture πάνω σε αυτό.

Με το Map Box γίνεται εύκολη και η ενσωμάτωση χαρτών σε εφαρμογές κινητών συσκευών.

Το Map Box διαθέτει ελεύθερα ένα ανοιχτού κώδικα Javascript API το οποίο κάνει εύκολη και γρήγορη την υλοποίηση χαρτών στο web. Το MapBox.js είναι πρόσθετο κομμάτι λογισμικού της Leaflet που είδαμε παραπάνω.

Το Map Box διαθέτει επίσης και ένα ανοιχτό και ελεύθερο λογισμικό, το TileMill το οποίο είναι ένα στούντιο σχεδίασης χωρικών δεδομένων. Με το TileMill και την χρήση μιας καινοτόμας, εύκολης στην χρήση γλώσσα μορφοποίησης βασισμένη στην CSS, την CryptoCSS, μπορούμε να δημιουργήσουμε εξελιγμένες οπτικοποιήσεις. Το TileMill, βέβαια, είναι κάτι παραπάνω από στούντιο σχεδίασης καθώς με αυτό μπορούμε να αναλύσουμε δεδομένα ράστερ, όπως δορυφορικά σήματα υπέρυθρης ζώνης. Αφού ολοκληρωθεί η οπτικοποίησή μας μπορούμε πολύ εύκολα να την ανεβάσουμε στο Map Box και να την δημοσιεύσουμε.



2-21 Το περιβάλλον εργασίας του TileMill

Κεφάλαιο 3

3 Οπτικοποίηση αποτελεσμάτων απογραφής πληθυσμού

3.1 Περιγραφή του προβλήματος

Η Ελληνική απογραφή του 2011 διενεργήθηκε από την Ελληνική Στατιστική αρχή και αποτέλεσε την τριακοστή απογραφή στην ιστορία του σύγχρονου Ελληνικού κράτους. Η προηγούμενη απογραφή διενεργήθηκε το 2001. Τα αποτελέσματα των τελευταίων απογραφών που δημοσιεύτηκαν σε

ιστοσελίδες δημόσιων οργανισμών έχουν συνήθως την μορφή ενός στατικού πίνακα δύο στηλών, με το όνομα της κάθε περιοχής και τον πληθυσμό της. Μια περιοχή μπορεί να είναι ένας οικισμός, μία δημοτική ενότητα, ένας δήμος ένας νομός ή μια περιφέρεια. Πολλές φορές βλέπουμε ότι οι περιοχές ομαδοποιούνται ανάλογα με το αν είναι οικισμοί που ανήκουν στην ίδια δημοτική ενότητα ή αν είναι δημοτικές ενότητες που ανήκουν στον ίδιο δήμο και ούτω καθεξής. Σε μερικές περιπτώσεις υπάρχουν και άλλες στήλες που εμφανίζουν αποτελέσματα προηγούμενων απογραφών.

Λαμβάνοντας ως μελέτη περίπτωσης τα αποτελέσματα της απογραφής για τον Δήμο Βόλου, ανατρέξαμε στην ιστοσελίδα του Δήμου και εκεί εντοπίσαμε έναν πίνακα σε μορφοποίηση excel που παρουσιάζει τα αποτελέσματα της απογραφής με τον τρόπο που περιγράφεται παραπάνω.

Μελετώντας τον πίνακα, βλέπουμε ότι είναι αρκετά δύσκολο να βγάλουμε κάποια πληροφορία απλά και μόνο διαβάζοντας τους αριθμούς. Για παράδειγμα, αν θέλουμε να εκτελέσουμε ακόμα και μία απλή πράξη σύγκρισης του πληθυσμού δύο οικισμών, θα πρέπει πρώτα να ψάξουμε ποιες είναι οι γραμμές που αντιστοιχούν σε αυτούς τους δύο οικισμούς και κατόπιν να συγκρίνουμε τις δύο αριθμητικές τιμές. Ας σκεφτούμε επίσης ότι μέσω του πίνακα είναι αδύνατο να αποκτήσουμε άμεσα σύνθετη γνώση για το πως κατανέμεται συνολικά ο πληθυσμός στην γεωγραφική περιοχή του δήμου.

Ενώ λοιπόν έχουμε στην διάθεση μας κάποια χρήσιμα δεδομένα, αναρωτιόμαστε πως θα μπορέσουμε να τα δομήσουμε και να τα παρουσιάσουμε για μπορέσουμε να εξερευνήσουμε σχέσεις και αλήθειες που κρύβονται πίσω από αυτά. Και βέβαια, η απάντηση σε αυτό είναι η ανάπτυξη των κατάλληλων οπτικοποιήσεων.

3.2 Στόχος

Από ότι φαίνεται λοιπόν στόχος μας είναι η οπτικοποίηση των αποτελεσμάτων της απογραφής για τον Δήμο Βόλου. Ρίχνοντας ακόμα μία ματιά στα δεδομένα μας μπορούμε να πούμε ότι έχουμε να κάνουμε με ποσοτικά (πληθυσμός), χωρικά (περιοχή), ιεραρχικά (διοικητική διαίρεση σύμφωνα με το Πρόγραμμα Καλλικράτη) δεδομένα. Θα πρέπει λοιπόν στην οπτικοποίησή μας :

- να συμπεριλάβουμε χαρακτηριστικά απεικόνισης ποσότητας ώστε να μπορεί κάποιος να αποκτήσει αίσθηση για το μέγεθος του πληθυσμού,
- να αποτυπώσουμε τις περιοχές στον χάρτη,
- να δομήσουμε ιεραρχικά τις περιοχές με ρίζα τον Δήμο Βόλου, δεύτερο επίπεδο τις δημοτικές ενότητες, τρίτο επίπεδο τις τοπικές ενότητες και τέταρτο τους οικισμούς.

Στην τελευταία περίπτωση μας ενδιαφέρει εκτός από την ανάλυση σε τρία επίπεδα και η ξεκάθαρη αποτύπωση των διοικητικών διαμερισμών.

Απώτερος σκοπός μας είναι, λοιπόν, η ανάπτυξη μιας διερευνητικής οπτικοποίησης, που θα μας δώσει μια ξεκάθαρη εικόνα για την χρονική και χωρική κατανομή του πληθυσμού στον Δήμο μας, αλλά και που ενδεχομένως θα μας οδηγήσει σε κάποια συμπεράσματα για ορισμένες περιοχές (π.χ. ακραίες τιμές) .

Για να επιτύχουμε κάτι τέτοιο, αλλά και για να δώσουμε την δυνατότητα σε οποιονδήποτε άλλον θέλει να δει τα αποτελέσματα της απογραφής και να αντλήσει την πληροφορία και την γνώση που θέλει, θα πρέπει να προσθέσουμε διαδραστικά στοιχεία στην οπτικοποίησή μας.

71	2401	ΔΗΜΟΣ ΒΟΛΟΥ (Έδρα: Βόλος,ο)	144.449
72	240102	ΔΗΜΟΤΙΚΗ ΕΝΟΤΗΤΑ ΑΓΡΙΑΣ	5.632
73	24010201	Δημοτική Κοινότητα <u>Αγριάς</u>	5.191
74	24010201	▶ Αγριά,η	5.191
75	24010202	Τοπική Κοινότητα <u>Δρακείας</u>	441
76	24010202	▶ Αγεμούτσα,η	3
77	24010202	▶ Δράκεια,η	381
78	24010202	▶ Χάνια,τα	57
79	240103	ΔΗΜΟΤΙΚΗ ΕΝΟΤΗΤΑ ΑΙΣΩΝΙΑΣ	3.249
80	24010301	Δημοτική Κοινότητα <u>Διμηνίου</u>	2.279
81	24010301	▶ Διμήνι,το	2.261
82	24010301	▶ Κάκκαβος,ο	8
83	24010301	▶ Παλιούρι,το	10
84	24010302	Τοπική Κοινότητα <u>Σέσκλου</u>	970
85	24010302	▶ Σέσκλο,το	862
86	24010302	▶ Χρυσή Ακτή Παναγίας,η	108
87	240104	ΔΗΜΟΤΙΚΗ ΕΝΟΤΗΤΑ ΑΡΤΕΜΙΔΑΣ	4.145
88	24010402	Τοπική Κοινότητα <u>Αγίου Βλασίου</u>	515
89	24010402	▶ Άγιος Βλάσιος,ο	322

3-1 Μέρος του συνόλου των δεδομένων μας, τα οποία αντλήθηκαν από την ιστοσελίδα του Δήμου Βόλου σε μορφή Excel.

3.3 Προδιαγραφές

Κάποιες προδιαγραφές που πρέπει να έχει η υλοποίησή μας ώστε να μεταδίδει την πληροφορία στον χρήστη με τον πλέον άμεσο και καταλληλότερο τρόπο είναι οι εξής:

- Να μπορεί εύκολα να κατανοεί το μέγεθος της πληθυσμιακής πυκνότητας που αντιστοιχεί σε κάθε περιοχή.
- Να μπορεί να επιλέγει και να εστιάζει στον χάρτη τις περιοχές που τον ενδιαφέρουν
- Να μπορεί να επιλέγει το επίπεδο στο οποίο θέλει να κάνει τις συγκρίσεις του (δήμος, δημοτικές ενότητες, οικισμοί).

3.4 Σχεδιασμός

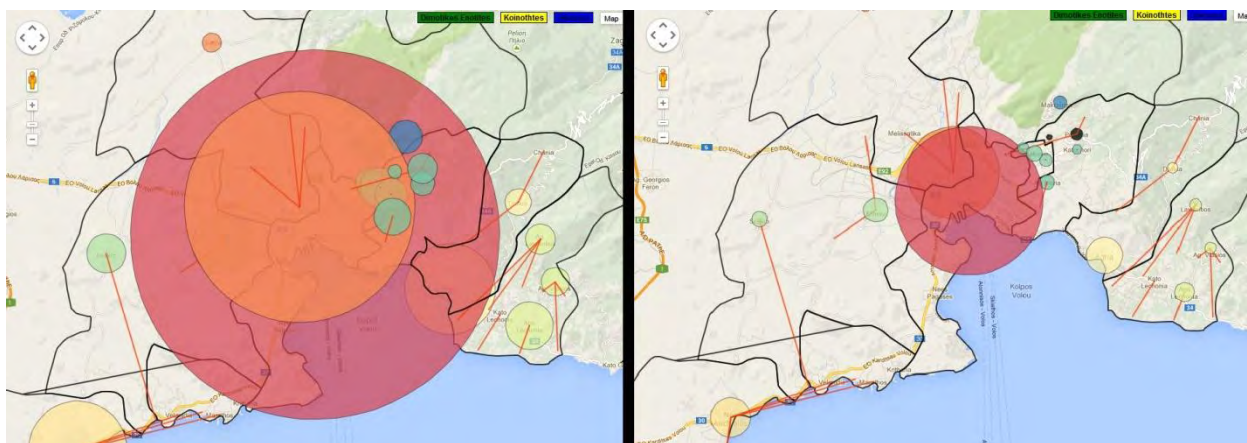
Για να επιτύχουμε τους στόχους μας, πληρώντας τις παραπάνω προδιαγραφές αρχικά μελετήσαμε τα δεδομένα μας. Παρατηρούμε ότι η οπτικοποίησή μας θα αποτελείται από τέσσερις διαστάσεις κάθε μία από τις οποίες θα αντιστοιχεί σε μία μεταβλητή. Επειδή, όπως αναφέραμε πριν τα δεδομένα μας είναι και ιεραρχικά θα πρέπει να υλοποιήσουμε ένα σύστημα απόδοσης των συνδέσμων που υπάρχουν για περιοχές που ανήκουν στην ίδια ενότητα (σε ένα δενδρικό σύστημα, έχουν κοινό γονέα).

Όσον αφορά την διάσταση του χώρου μπορούμε να χρησιμοποιήσουμε έναν χάρτη και να απεικονίσουμε τις περιοχές σαν σημεία πάνω στον δισδιάστατο χώρο του χάρτη. Επειδή όμως οι τιμές της χωρικής μεταβλητής εκφράζονται με ονομασίες και όχι με γεωγραφικά μήκη και πλάτη, θα πρέπει να χρησιμοποιήσουμε μία υπηρεσία Geolocation. Το Google API v3, που θα παρουσιάσουμε αναλυτικά στην συνέχεια, προσφέρει αυτήν την υπηρεσία, εμείς απλά στέλνουμε ως ερώτηση τις ονομασίες και λαμβάνουμε ως απάντηση τα γεωγραφικά μήκη και πλάτη που αντιστοιχούν στην κάθε ονομασία.

Για να απεικονίσουμε τα όρια της κάθε δημοτικής ενότητας χρησιμοποιήσαμε δεδομένα από την ιστοσελίδα <http://www.geodata.gov.gr/geodata> του Ινστιτούτου Πληροφοριακών Συστημάτων / Ερευνητικό Κέντρο «Αθηνά», τα οποία τα πήραμε σε KML μορφή. Κατόπιν χρησιμοποιώντας έναν [μετατροπέα](#) φέραμε τα δεδομένα μας από KML στην βολικότερη για εμάς JSON μορφή, που θα αναλύσουμε στην ενότητα 3.6.2. Δίνοντας ως όρισμα την JSON μεταβλητή μέσω του API της Google εναποθέσαμε στην υπηρεσία Google Maps τον σχεδιασμό των πολυγώνων που αφορούν τα όρια της κάθε ενότητας. Τέλος, προσθέσαμε event listeners σε κάθε πολύγωνο που σχηματίστηκε ώστε αν ο χρήστης κλικάρει πάνω σε κάποια ενότητα να γίνεται αυτόματη εστίαση.

Προσθέτοντας τα όρια της κάθε δημοτικής ενότητας απεικονίζουμε και την ιεραρχική φύση των μεταβλητών μας. Εύκολα καταλαβαίνει παραδειγματος χάριν ότι η δημοτική κοινότητα της Άλλης Μεριάς ανήκει στην δημοτική ενότητα Πορταριάς αφού το κέντρο του κύκλου βρίσκεται μέσα στα αντίστοιχα όρια. Επιπλέον για να απεικονίσουμε την ιεραρχική σχέση των οικισμών ως προς τις τοπικές κοινότητες χρησιμοποιούμε γραμμές οι οποίες συνδέουν περιοχές που ανήκουν στην ίδια γεωγραφική ενότητα. Επειδή το χρώμα της φυσαλίδας μας έχει μείνει ως αχρησιμοποίητη οπτική ιδιότητα εφαρμόζουμε περιττή κωδικοποίηση (βλ. 2.5.2) και χρωματίζουμε με το ίδιο χρώμα φυσαλίδες που ανήκουν στην ίδια δημοτική ενότητα, ώστε να κατανοήσουμε ευκολότερα και γρηγορότερα την διοικητική διαίρεση.

Με βάση τα όσα αναφέραμε στο προηγούμενο κεφάλαιο, για να κωδικοποιήσουμε την ποσοτική μεταβλητή θα χρειαστούμε μια οπτική ιδιότητα φυσικής διάταξης και διακριτών τιμών. Η οπτική αυτή ιδιότητα θα πρέπει να τοποθετηθεί σε διάφορα σημεία του χάρτη. Για τις τοπικές κοινότητες και τους οικισμούς, χρησιμοποιούμε φυσαλίδες, που είναι κύκλοι με κέντρο την θέση της περιοχής. Προσέχουμε ότι ο πληθυσμός δεν θα πρέπει να κωδικοποιηθεί ως μέγεθος ακτίνας κύκλου R αλλά ως εμβαδόν κύκλου πR^2 . Χρωματίζουμε, τέλος τις φυσαλίδες ανάλογα με την κατηγορική μεταβλητή δηλαδή με το σε ποια δημοτική ενότητα ανήκουν. Παρατηρούμε ότι η οπτικοποίησή μας μοιάζει με το διάγραμμα σημείων που παρουσιάσαμε στην ενότητα 2.4.1 με την διαφορά ότι εδώ ο οριζόντιος και κάθετος άξονας έχουν αντικατασταθεί από το γεωγραφικό μήκος και πλάτος και στο φόντο της οπτικοποίησης εμφανίζεται γεωγραφικός χάρτης. Για τις δημοτικές ενότητες χρησιμοποιούμε ένα χρωματικό διάνυσμα όπου τα κρύα χρώματα (μπλε) κωδικοποιούν χαμηλές τιμές ενώ τα θερμά (κόκκινα) υψηλές.



3-2 Στην αριστερή εικόνα βλέπουμε ότι η φυσαλίδα του Δήμου Βόλου καλύπτει πολλές από τις γειτονικές περιοχές. Μεγαλώνουμε το μέγεθος της κλίμακας και στα δεξιά βλέπουμε ότι πλέον δεν υπάρχει το πρόβλημα αυτό.

Χρησιμοποιώντας τα κατάλληλα εργαλεία που θα αναφέρουμε παρακάτω υλοποιούμε την οπτικοποίηση. Το πρώτο πράγμα που παρατηρούμε είναι ότι έχουμε πολλές ακραίες τιμές. Για παράδειγμα, ενώ οικισμός Ανεμούτσα έχει μόνο 3 μόνιμους κατοίκους και η αποτύπωση του στον χάρτη είναι σχεδόν άορατη, η πόλη του Βόλου έχει πληθυσμό 86.046 και η φυσαλίδα που αντιστοιχεί σε αυτήν υπερκαλύπτει τις γειτονικές της λόγω ακραίου μεγέθους. Για να λύσουμε το πρόβλημα της υπερκάλυψης θα χρησιμοποιήσουμε μία κατάλληλη κλίμακα. Με δοκιμές διαιρούμε τις εκτάσεις των φυσαλίδων με κάποιον αριθμό ώσπου επιλέγουμε την κλίμακα για τα εμβιά τους.

3.5 Λειτουργίες

Η οπτικοποίηση που αναπτύξαμε είναι διαδραστική αφού έχουμε ενσωματώσει μια διεπαφή χρήστη με τις εξής λειτουργίες:

- Ο χρήστης έχει την δυνατότητα να εστιάσει (zoom) στην περιοχή που τον ενδιαφέρει μέσω του κουμπιού ελέγχου που παρέχει το Javascript API του Google Map, είτε επιλέγοντας επίπεδο από τα τρία κουμπιά στο πάνω δεξιά μέρος της οπτικοποίησης, είτε επιλέγοντας με αριστερό κλικ κάποια δημοτική ή τοπική ενότητα. Στην τελευταία περίπτωση γίνεται αυτόματη εστίαση και το viewport αλλάζει έτσι ώστε να περιέχει όλο το πολύγωνο μιας περιοχής με τον μέγιστο δυνατό ζουμ.
- Προβολή του αριθμού του πληθυσμού. Μόλις ο χρήστης περάσει τον κέρσορα του ποντικιού πάνω από κάποια περιοχή εμφανίζεται ένα μικρό παράθυρο στο οποίο αναγράφεται το όνομα της περιοχής και ο πληθυσμός της.

3.6 Εργαλεία

3.6.1 Google Maps Javascript API V3

Το Google Maps Javascript API V3 αυτή τη στιγμή είναι το επίσημο API της Google για χρήση μιας πληθώρας γεωγραφικών διαδικτυακών υπηρεσιών όπως η διαχείριση των χαρτών Google, η προσθήκη περιεχομένου σε αυτούς, αλλά και η ενσωμάτωσή τους σε εφαρμογές ιστού. Η τρίτη έκδοση, είναι σχεδιασμένη να έχει μεγαλύτερη ταχύτητα και να είναι περισσότερο εφαρμόσιμη για κινητές συσκευές, σε βαθμό ίδιο με μια παραδοσιακή desktop εφαρμογή φυλλομετρητή. Η υπηρεσία είναι ελεύθερη για χρήση σε κάθε ιστοσελίδα που είναι ελεύθερη στους καταναλωτές.

Στην δική μας υλοποίηση χρησιμοποιήσαμε αρκετές διεπαφές όπως η υπηρεσία Geolocation, η αλλαγή επιπέδου εστίασης (map.setzoom), η ενεργοποίηση γεγονότος αυτόματης βέλτιστης εστίασης στο αριστερό κλικ εντός της περιοχής που σχηματίζουν τα πολυγωνικά όρια (fitBounds), η σχεδίαση των φυσαλίδων (Circle), των γραμμών (Polyline) και των ορίων (Polygon), η εμφάνιση παραθύρου πληροφοριών (infoWindow) καθώς περνάμε με το ποντίκι πάνω από (mouseover) μία δημοτική ενότητα.

3.6.2 JSON

Η γλώσσα JSON (**J**ava**S**cript **O**bject **N**otation) είναι μια μορφοποίηση αναγνώσιμη από τον άνθρωπο για εναλλαγή δεδομένων - κειμένου. Η γλώσσα είναι αυτοπεριγραφική (self-describing) και εύκολη στην κατανόηση. Παρόλο που χρησιμοποιεί συντακτικό της Javascript για να περιγράψει αντικείμενα δεδομένων, είναι από μόνη της μία γλώσσα ανεξάρτητη πλατφόρμας και ως εκ τούτου διαθέσιμη σε πολλές γλώσσες προγραμματισμού. Το φαρμάκι της γλώσσας πολλές φορές χρησιμοποιείται για σειριοποίηση και μεταφορά δομημένων δεδομένων σε μια σύνδεση δικτύου. Αρχικά χρησιμοποιήθηκε για να μεταφέρει δεδομένα μεταξύ του εξυπηρετητή και της web εφαρμογής, ως εναλλακτική της XML.

Την γλώσσα JSON την χρησιμοποιήσαμε προκειμένου να αναπαραστήσουμε τα δεδομένα μας, που αρχικά ήταν με την μορφή πίνακα, με ιεραρχικά δομημένο τρόπο. Προσθέσαμε επίσης τα γεωγραφικά μήκη και πλάτη (lat: latitude, lng: longitude). Παρακάτω βλέπουμε πως αναπαραστήσαμε την δημοτική ενότητα Αγριάς ως αντικείμενο JSON.

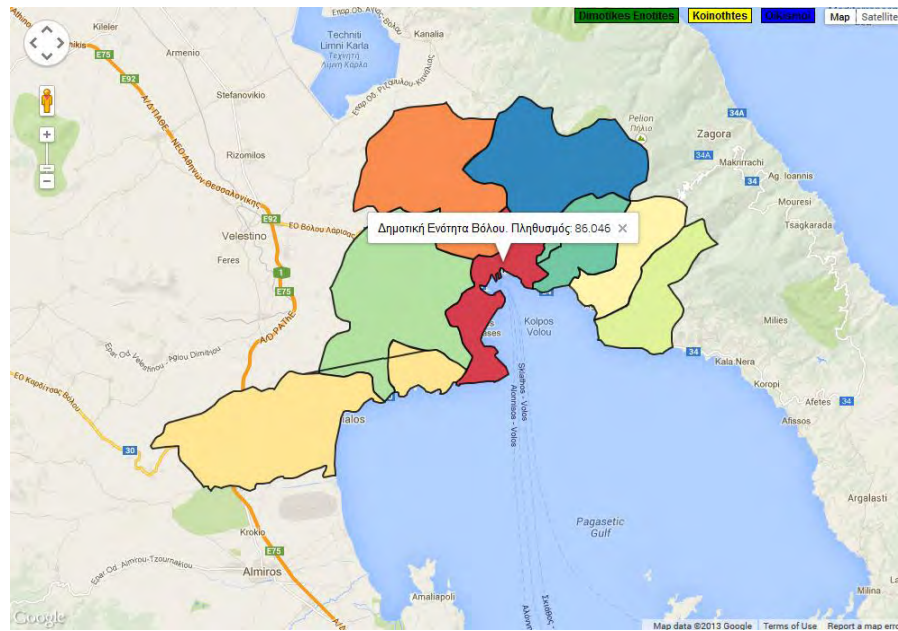
```
var database= [
  {"code":240102,"name":"ΔΗΜΟΤΙΚΗ ΕΝΟΤΗΤΑ ΑΓΡΙΑΣ", "population":5632, "koinothtes": [
    {"code":24010201,"name":"Τοπική Κοινότητα Αγριάς", "population":5191,
    "lat":39.341566, "lng":23.012788, "oikismoi": [
      {"code":2401020101,"name":"Αγριά,η", "population":5191,
      "lat":39.341566, "lng":23.012788}
    ]},
    {"code":24010202,"name":"Τοπική Κοινότητα Δρακείας", "population":441,
    "lat":39.376523, "lng":23.048197, "oikismoi": [
      {"code":2401020202,"name":"Ανεμούτσα,η", "population":3,
      "lat":39.358923, "lng":23.018918},
      {"code":2401020201,"name":"Δράκεια,η", "population":381,
      "lat":39.376523, "lng":23.048197},
      {"code":2401020203,"name":"Χάνια,τα", "population":57, "lat":39.397104,
      "lng":23.06154}
    ]}
  ]}
]
```

Με τον τρόπο αυτό μπορέσαμε εύκολα να ανατρέξουμε επαναληπτικά ή αναδρομικά στα υποσύνολα των δεδομένων που μας ενδιαφέρουν κατά περίπτωση και να πάρουμε από αυτά τα χαρακτηριστικά που μας ενδιαφέρουν.

Επίσης, σε JSON μορφή αναπαραστήσαμε τα όρια των δημοτικών ενότητων. Για την κάθε ενότητα δημιουργήσαμε μια JSON μεταβλητή με τιμές τις τετμημένες και τεταγμένες της κάθε κορυφής του πολυγώνου που αντιστοιχεί στα παραπάνω όρια.

3.7 Επίδειξη

3.7.1 Επίπεδο Δημοτικών Ενοτήτων

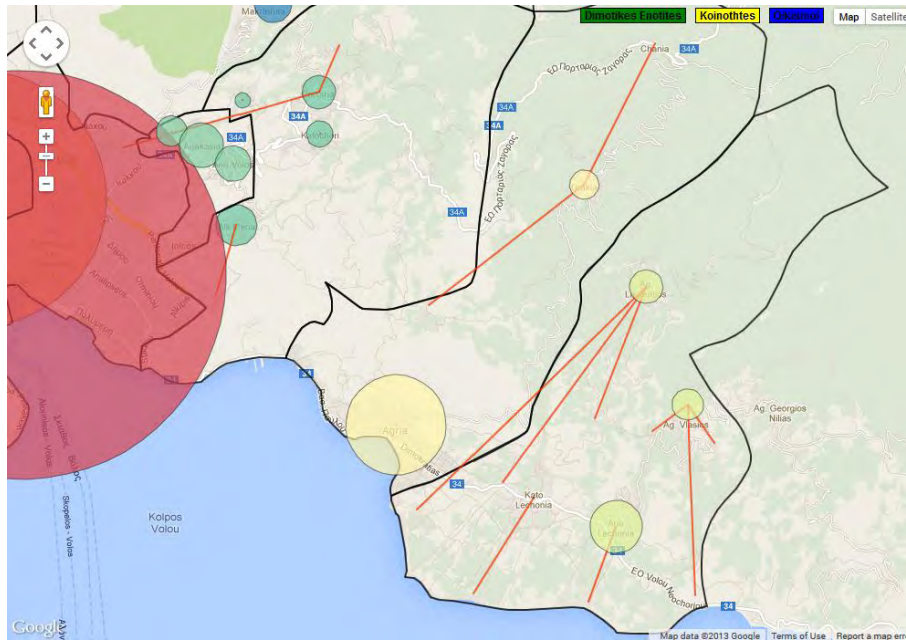


3-3 Πληθυσμός ανά δημοτική ενότητα

Ο χρήστης της εφαρμογής μας αρχικά βλέπει τον χάρτη των δημοτικών ενοτήτων. Παρατηρούμε ότι η Δημοτική Ενότητα Βόλου έχει πολύ πιο θερμό χρώμα από ότι οι υπόλοιπες και άμεσα συμπεραίνουμε ότι είναι η ενότητα με τον μεγαλύτερο πληθυσμό. Για να συγκρίνουμε τα μεγέθη, ελέγχουμε την λεζάντα με την κλίμακα που αντιπροσωπεύει το χρωματικό διάγραμμα. Μετακινώντας το ποντίκι μέσα στα όρια της ενότητας βλέπουμε την τιμή του πληθυσμού. Κλικάροντας σε κάποια από τις δημοτικές ενότητες βλέπουμε ότι ο χάρτης εστιάζεται σε αυτήν την ενότητα. Έχουμε περάσει πλέον στο επίπεδο των Δημοτικών Κοινοτήτων.

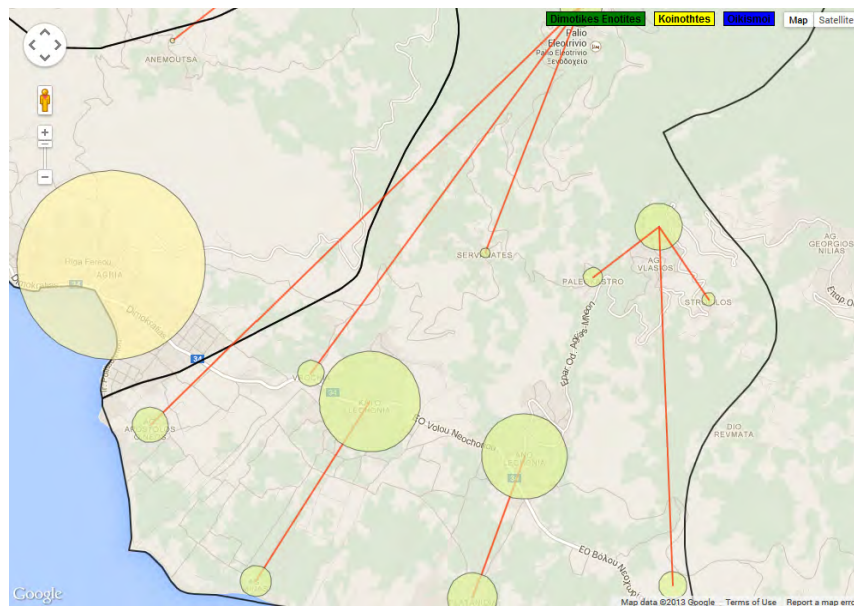
3.7.2 Επίπεδο Δημοτικών Κοινοτήτων

Εδώ βλέπουμε ότι σε κάθε Δημοτική Κοινότητα αντιστοιχεί μία φυσαλίδα. Το κέντρο της φυσαλίδας συμπίπτει με το κεντρικό σημείο της κοινότητας, ενώ το εμβαδόν της τον πληθυσμό. Βλέπουμε επίσης τις γραμμές οι οποίες δείχνουν προς τους οικισμούς που ανήκουν στην Δημοτική Κοινότητα. Για το επίπεδο των Κοινοτήτων έχουμε διαθέσιμα δύο επίπεδα εστίασης.



3-4 Ο πληθυσμός των κοινοτήτων αναπαρίσταται ως φυσαλίδα.

3.7.3 Επίπεδο Οικισμών



3-5 Ο πληθυσμός των οικισμών αναπαρίσταται ως φυσαλίδα

Όπως και πριν, αναπαριστούμε τον πληθυσμό των οικισμών με φυσαλίδες, αυτή τη φορά μια για κάθε οικισμό.

Κεφάλαιο 4

4 Παρατηρητήριο ερευνητικών δραστηριοτήτων για το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Π.Θ.

4.1 Παρατηρητήρια Ερευνητικών Δραστηριοτήτων

Αναμφίβολα, η επιστημονική έρευνα αποτελεί ένα από τα κυριότερα εργαλεία ανάπτυξης και προόδου μιας κοινωνίας. Ως εκ τούτου η καταγραφή, η εξαγωγή συμπερασμάτων και η μελέτη των αποτελεσμάτων της έρευνας που πραγματοποιούν διάφοροι οργανισμοί στην υφήλιο θα μπορούσε να μας δώσει πολύτιμη γνώση για το ποιες ερευνητικές κοινότητες προσδεύουν, πως εξελίσσεται η επιστημονική παραγωγή ανά χώρα, πως συνεργάζονται οι οργανισμοί σε παγκόσμιο επίπεδο κλπ. Κάτι τέτοιο αποτελεί θεμελιώδη παράγοντα για το σχεδιασμό πολιτικών έρευνας και καινοτομίας, που θα μπορούσε να συμβάλλει αποφασιστικά στη διαμόρφωση νέων αναπτυξιακών μοντέλων, θέτοντας τη γνώση ως ανταγωνιστικό πλεονέκτημα.

Η μελέτη των αποτελεσμάτων είναι επίσης εξαιρετικά χρήσιμη για τους ερευνητές ώστε να μπορέσουν να επεκτείνουν την γνώση τους και να σχηματίσουν ένα δίκτυο γνώσης που θα οδηγήσει σε νέα ερευνητικά αποτελέσματα.

Όσον αφορά τα πανεπιστήμια, η μελέτη των ερευνητικών δραστηριοτήτων και των αποτελεσμάτων της που πραγματοποιείται στα πλαίσια του κάθε ιδρύματος, δίνει την δυνατότητα στο να αναδείξουν και να προβάλλουν τα επιστημονικά πεδία στα οποία διαθέτουν πραγματογνωμοσύνη, αλλά και να δουν τα πεδία στα οποία εμφανίζουν αδυναμίες, ώστε να επιδιώξουν βελτίωση. Επίσης, δίνει την δυνατότητα να προβάλλουν το έργο τους σε άλλα πανεπιστήμια, ιδρύματα ή ερευνητές και μέσω συσχέτισης των ερευνών να επιδιώξουν συνεργασίες. Τέλος τους δίνει την δυνατότητα να προβάλλουν τα γενικότερα αποτελέσματα και την υπάρχουσα γνώση και τεχνογνωσία στο ευρύ κοινό.

Το ερώτημα που γεννάται είναι πώς και από πού θα αντλήσουμε την πληροφορία που θα μας οδηγήσει στα παραπάνω συμπεράσματα. Τα αποτελέσματα μιας έρευνας που πραγματοποιεί μια ομάδα επιστημόνων δημοσιεύονται, μετά από συνήθως αυστηρή κρίση, ως **άρθρα σε κάποιο επιστημονικό περιοδικό** (ή πρακτικά συνεδρίων). Οι δημοσιεύσεις σε επιστημονικά περιοδικά, το κυριότερο μέσο για τη διάχυση των ερευνητικών αποτελεσμάτων στην επιστημονική κοινότητα, αποτελούν μια σημαντική πηγή για την καταγραφή και αποτίμηση του ερευνητικού έργου. Σκοπός μας λοιπόν είναι να συγκεντρώσουμε δεδομένα που προκύπτουν από τα παραπάνω άρθρα. Για να το κάνουμε αυτό, μπορούμε να χρησιμοποιήσουμε κάποια μηχανή αναζήτησης ερευνητικών δραστηριοτήτων ή κάποια διεθνή βιβλιογραφική βάση.

4.1.1 Βάσεις δεδομένων επιστημονικών δημοσιεύσεων

Για να αντλήσουμε λοιπόν τα δεδομένα που θα μας οδηγήσουν στην γνώση που περιγράφηκε από πάνω, μπορούμε να χρησιμοποιήσουμε βάσεις δεδομένων που περιλαμβάνουν βιβλιογραφικές εγγραφές επιστημονικών δημοσιεύσεων σε παγκόσμιο επίπεδο και στοιχεία για τις αναφορές μεταξύ τους. Παρακάτω θα αναφέρουμε κάποιες από τις πιο γνωστές μηχανές αναζήτησης ερευνητικών δραστηριοτήτων:

Το *Scopus* της εταιρίας Elsevier αποτελεί μία από τις μεγαλύτερες βάσεις δεδομένων για ερευνητικές δραστηριότητες. Περιλαμβάνει αναλυτικά στοιχεία και πληροφορίες για τις επιστημονικές δημοσιεύσεις και τις αναφορές σε αυτές και υποστηρίζει το ομώνυμο διαδικτυακό εργαλείο της Elsevier. Στην βάση δεδομένων του Scopus εμπεριέχονται περίπου 49 εκατομμύρια εγγραφές. Περιέχει επίσης έξυπνα εργαλεία για παρατήρηση, ανάλυση και οπτικοποίηση των αποτελεσμάτων της αναζήτησης. Στο νεότερο σύστημα Scopus, ευρετηριάζονται πάνω από 18.500 τίτλοι επιστημονικών περιοδικών, οι οποίοι διευρύνονται συνεχώς.

Το *Web Of Science* της εταιρίας Thomson Reuters παρέχει πρόσβαση σε βάσεις δεδομένων που καλύπτουν τα επιστημονικά περιοδικά με υψηλή επιρροή παγκοσμίως. Το σύστημα Web of Science (WoS) είναι η παλαιότερη βάση δεδομένων επιστημονικών δημοσιεύσεων με εγγραφές που ξεκινούν από το 1900. Αντλεί δεδομένα από περισσότερα από 12.000 περιοδικά τα οποία υπόκεινται σε αξιολόγηση κριτών (peer-review). Η λειτουργικότητα του είναι παρόμοια με αυτή του Scopus: ένας χρήστης μπορεί να κάνει κάποια επερώτηση ορίζοντας κάποια πεδία αναζήτησης όπως το όνομα του ερευνητή, τον τίτλο της δημοσίευσης, τις λέξεις κλειδιά και ούτω καθεξής. Η σημαντικότερη διαφορά με το Scopus είναι ότι το αποτέλεσμα που επιστρέφεται από το Scopus είναι στραμμένο περισσότερο στο προφίλ του ερευνητή.

Το *Google Scholar* αντλεί δεδομένα από επιστημονικά άρθρα, διατριβές, εκθέσεις, περιλήψεις, γνώμες δικαστηρίου που μπορεί να βρίσκονται σε ιστότοπους ακαδημαϊκών δημοσιεύσεων, σε online αποθήκες, σε ιστοσελίδες πανεπιστημίων και σε άλλους διαδικτυακούς τόπους. Στο Google Scholar μία αναζήτηση γίνεται όπως και στην κανονική αναζήτηση της Google αλλά τα αποτελέσματα αφορούν επιστημονικές δραστηριότητες και δείκτες μέτρησης τους.

Το σύστημα Google Scholar, παρά τον τεράστιο αριθμό πηγών που περιλαμβάνει, δεν είναι κατάλληλο για βιβλιομετρικές αναλύσεις που αναφέρονται σε επίπεδο χωρών ή οργανισμών, λόγω της έλλειψης μεταδεδομένων που απαιτούνται για την ταυτοποίηση των δημοσιεύσεων και της απουσίας κριτηρίων που διασφαλίζουν την ποιότητα των δημοσιεύσεων που περιλαμβάνονται στο σύστημα.

Στα δεδομένα του Google Scholar βασίζεται και το ελεύθερο λογισμικό "Publish or Perish" (*Anne-Wil Harzing, 2010*) το οποίο υπολογίζει μια σειρά από βιβλιομετρικούς δείκτες.

Όσον αφορά τα συστήματα Web of Science και Scopus, και τα δύο διασφαλίζουν τη διάθεση αναλυτικών μεταδεδομένων και την ποιότητα των δημοσιεύσεων που περιλαμβάνουν.

Για την ανάπτυξη της εφαρμογής μας χρησιμοποιήσαμε τη διεθνή βιβλιογραφική βάση Scopus.

Το *Incites* είναι άλλη μία υπηρεσία της Thomson Reuters που επιτρέπει την ανάλυση της παραγωγικότητας των ιδρυμάτων και των οργανισμών, μέσα από οπτικοποιήσεις δεδομένων και εργαλεία δημιουργίας αναφορών. Για την ανάλυση αυτή χρησιμοποιεί κυρίως δεδομένα και δείκτες μετρήσεων που προκύπτουν από το Web of Science και αποσκοπεί σε λήψη συμπερασμάτων και αποφάσεων σε ένα γενικότερο πεδίο.

Η μηχανή αναζήτησης της Microsoft, *Microsoft Academic Search* παρέχει μια πολύ καλή κατηγοριοποίηση των δημοσιεύσεων όπως επίσης και έναν εξαιρετικά καλαίσθητο τρόπο παρουσίασης των αποτελεσμάτων μέσα από πλούσιες και διαδραστικές οπτικοποιήσεις δεδομένων.

Το *CiteULike* βασίζεται στις αρχές της κοινωνικής διαμοίρασης σελιδοδεικτών (social bookmarking) στοχεύοντας στην προώθηση και ανάπτυξη της διαμοίρασης των επιστημονικών αναφορών μεταξύ των ερευνητών. Ένας ερευνητής μπορεί να εισάγει στο CiteULike μια αναφορά μιας δημοσίευσης του που βρίσκεται σε κάποια online βάση δεδομένων. Αυτόματα το CiteULike θα αναζητήσει τα μεταδεδομένα που αφορούν την δημοσίευση και θα τα καταχωρήσει στην δική του βάση δεδομένων. Στην συνέχεια θα ψάξει για αναφορές που αφορούν την δημοσίευση.

4.1.2 Συστήματα CRIS

Η ανάγκη για συστηματική προτυποποιημένη καταγραφή και προβολή της ερευνητικής δραστηριότητας και την ευρύτερη αξιοποίηση των αποτελεσμάτων της αποτελεί το βασικό κίνητρο για την ανάπτυξη των σύγχρονων συστημάτων διαχείρισης πληροφορίας για την ερευνητική δραστηριότητα (Current Research Information Systems - CRIS). Ένα σύγχρονο σύστημα CRIS καταγράφει βάσει προτύπων, επεξεργάζεται και παρουσιάζει, στοιχεία και δείκτες σχετικά με την ερευνητική δραστηριότητα σε ολόκληρο τον κύκλο ζωής της. Ιδιαίτερη έμφαση αποδίδεται στην καταγραφή των αποτελεσμάτων της ερευνητικής δραστηριότητας (δημοσιεύσεις, πατέντες, ερευνητικά δεδομένα) και τη σύνδεσή τους με το περιβάλλον από το οποίο αυτά προέρχονται (ερευνητές, οργανισμοί, χρηματοδοτούμενα προγράμματα και έργα, ερευνητικές υποδομές, υπηρεσίες).

Τα συστήματα διαχείρισης της πληροφορίας για την ερευνητική δραστηριότητα λειτουργούν σήμερα με επιτυχία στον διεθνή χώρο ως αυτόνομα συστήματα που παρέχουν αξιόπιστη πληροφόρηση για την ερευνητική δραστηριότητα τόσο σε επίπεδο ερευνητών όσο και σε επίπεδο κρατών, οργανισμών, χρηματοδοτικών πρωτοβουλιών και προγραμμάτων κ.λπ. Η χρήση τους έχει διαδοθεί ιδιαίτερα τα τελευταία χρόνια, με αποτέλεσμα να θεωρούνται απαραίτητο εργαλείο για τη διαχείριση, διάχυση και αξιοποίηση των αποτελεσμάτων της ερευνητικής δραστηριότητας.

4.1.3 Βιβλιομετρικοί δείκτες μέτρησης της ερευνητικής δραστηριότητας

Αντικείμενο της βιβλιομετρικής ανάλυσης είναι η καταγραφή και επεξεργασία δεδομένων που σχετίζονται με τις επιστημονικές δημοσιεύσεις και η εξαγωγή των σχετικών "βιβλιομετρικών δεικτών", όπως ο αριθμός των δημοσιεύσεων, οι αναφορές σε αυτές από άλλες δημοσιεύσεις (citations) κ.λπ.

Οι βιβλιομετρικοί δείκτες μέτρησης της ερευνητικής συμβάλλουν στη διαμόρφωση μιας όσο το δυνατόν αντικειμενικά μετρήσιμης εικόνας των συστημάτων έρευνας, τεχνολογικής ανάπτυξης και

καινοτομίας. Χρησιμοποιούνται για την αξιολόγηση ερευνητικών οργανισμών, ομάδων και ερευνητών, καταγράφουν τα ερευνητικά πεδία στα οποία δραστηριοποιείται η επιστημονική κοινότητα, αποτυπώνουν τα νέα ερευνητικά πεδία που αναδύονται καθώς και τα επιστημονικά δίκτυα που δημιουργούνται για την υλοποίηση κοινών ερευνητικών στόχων.

Η βιβλιομετρική ανάλυση αποτελεί σημαντικό εργαλείο για τον προσδιορισμό χαρακτηριστικών και τάσεων της ερευνητικής παραγωγής σε επίπεδο οργανισμού, χώρας ή ευρύτερου συνόλου χωρών, την εκτίμηση της απήχησης του επιστημονικού έργου, την αξιολόγηση της ερευνητικής δραστηριότητας και την ανάδειξη εθνικών και πολυεθνικών δικτύων μεταξύ επιστημόνων και επιστημονικών κλάδων. Οι μελέτες που στηρίζονται στη βιβλιομετρική ανάλυση αυξάνονται συνεχώς τα τελευταία χρόνια στο διεθνή χώρο. Οι βιβλιομετρικοί δείκτες χρησιμοποιούνται για την αξιολόγηση ερευνητικών συστημάτων ή οργανισμών και συνεισφέρουν στη διαμόρφωση εθνικών πολιτικών για την έρευνα.

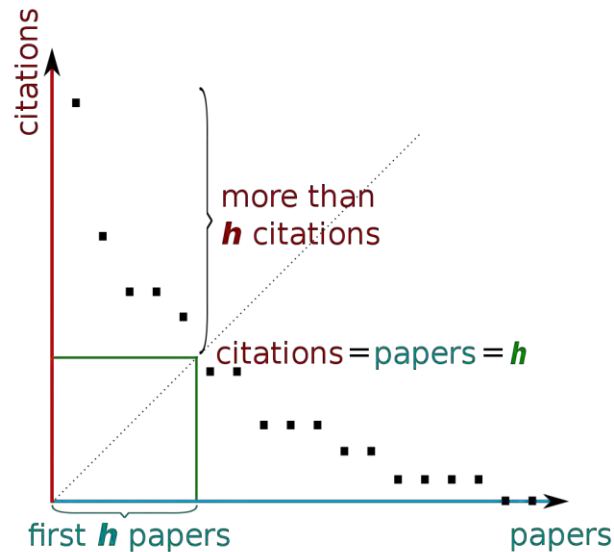
Στη βιβλιογραφία καταγράφονται μειονεκτήματα και περιορισμοί στον υπολογισμό και τη χρήση των δεικτών, όπως π.χ. οι διαφορές στην πρακτική δημοσιεύσεων και αναφορών στα επιστημονικά πεδία (π.χ. ιατρικές σε σχέση με ανθρωπιστικές επιστήμες) οι οποίες επηρεάζουν τους δείκτες απήχησης. Επιπλέον, αναφέρονται προβλήματα που σχετίζονται με τον «καθαρισμό» των πρωτογενών δεδομένων και την ταυτοποίηση των δημοσιεύσεων, η αδυναμία απόδοσης άλλων σημαντικών συνιστωσών της ερευνητικής δραστηριότητας κ.ά. Οι προβληματισμοί αυτοί δεν αναιρούν τη σημασία των βιβλιομετρικών δεικτών ως πολύτιμη πηγή δεδομένων και, όπως άλλωστε ισχύει με την ερμηνεία των περισσότερων δεικτών, μπορούν να ξεπεραστούν όταν οι βιβλιομετρικοί δείκτες ερμηνευτούν στο σωστό πλαίσιο.

Για τον προσδιορισμό των κύριων χαρακτηριστικών της επιστημονικής παραγωγής σε δημοσιεύσεις υιοθετήθηκαν βιβλιομετρικοί δείκτες οι οποίοι είναι καθιερωμένοι στη διεθνή βιβλιογραφία, και χρησιμοποιούνται σε αντίστοιχες μελέτες.

Στην εφαρμογή που θα αναπτύξουμε παρακάτω, θα χρησιμοποιήσουμε τους παρακάτω δείκτες:

- **Αριθμός εγγράφων (Number of Documents).** Ο αριθμός των δημοσιεύσεων που αντιστοιχεί στον κάθε ερευνητή. Μια δημοσίευση μπορεί να είναι: Άρθρο σε επιστημονικό περιοδικό, Review, Editorial, Conference, Erratum.
- **Αριθμός αναφορών (Number of References)**
- **Αριθμός αναφορών σε άλλες δημοσιεύσεις (Number of Citations).** Ένας πολύ σημαντικός δείκτης του πόσο σημαντική είναι μία δημοσίευση είναι κατά πόσο άλλες δημοσιεύσεις αναφέρονται σε αυτήν.
- **Αριθμός αναφορών ανά δημοσίευση (Number of Citations by Document)**
- **H index.** Ο h-index είναι ένας βιβλιομετρικός δείκτης που επινοήθηκε το 2005 από τον J.E. Hirsch. Όπως το διατυπώνει ο ίδιος: "Ένας συγγραφέας εμφανίζει δείκτη h όταν h από τις N_p εργασίες του έχουν τουλάχιστον h αναφορές η καθεμία, ενώ οι υπόλοιπες (N_p-h) εργασίες του έχουν λιγότερες από h αναφορές η καθεμία. Παράδειγμα: Αν μια δοσμένη ομάδα εργασιών ή οι δοσμένες εργασίες κάποιου συγγραφέα εμφανίζουν $h=12$, αυτό σημαίνει ότι σε αυτή την ομάδα των εργασιών υπάρχουν 12 εργασίες που έχουν τουλάχιστον 12 (ή και περισσότερες) αναφορές η καθεμία. Δηλαδή, ο h-index εξάγεται πρακτικά με κατάταξη της δοσμένης ομάδας των εργασιών κατά φθίνουσα σειρά ξεκινώντας από αυτή που έχει τις περισσότερες αναφορές. Καθώς προχωράμε προς τα κάτω, στο σημείο που ο αριθμός των αναφορών που έχει λάβει κάποια εργασία γίνει μικρότερος από τον αύξοντα αριθμό της

εργασίας, τραβάμε μία γραμμή και λαμβάνουμε ως h-index τον αμέσως προηγούμενο αύξοντα αριθμό. Εξ' ορισμού η τιμή του h-index εξαρτάται από τη βάση δεδομένων στην οποία θα γίνει η αναζήτηση.



4-1 Υπολογισμός του h index. Πηγή: Wikipedia

4.1.4 Επιστημονικά πεδία

Μεταξύ των δεδομένων που παρέχουν οι διεθνείς βάσεις επιστημονικών δημοσιεύσεων, είναι και η κατηγοριοποίηση των δημοσιεύσεων σε συγκεκριμένες επιστημονικές περιοχές. Στις βάσεις του Scopus, οι δημοσιεύσεις διαχωρίζονται θεματικά σε εξειδικευμένες επιστημονικές περιοχές με κριτήριο το περιοδικό στο οποίο δημοσιεύονται. Μία δημοσίευση μπορεί να ανήκει σε περισσότερες από μία κατηγορίες.

Οι μέχρι στιγμής δημοσιεύσεις του Τ.Η.Μ.Μ.Υ. εντάσσονται σε κάποιες από τις παρακάτω κατηγορίες:

AGRI	Agricultural and Biological Sciences
ARTS	Arts and Humanities
BIOC	Biochemistry, Genetics and Molecular Biology
BUSI	Business, Management and Accounting
CENG	Chemical Engineering
CHEM	Chemistry
COMP	Computer Science
DECI	Decision Sciences
DENT	Dentistry
EART	Earth and Planetary Sciences

ECON	Economics, Econometrics and Finance
ENER	Energy
ENGI	Engineering
ENVI	Environmental Science
HEAL	Health Professions
IMMU	Immunology and Microbiology
MATE	Materials Science
MATH	Mathematics
MEDI	Medicine
NEUR	Neuroscience
NURS	Nursing
PHAR	Pharmacology, Toxicology and Pharmaceutics
PHYS	Physics and Astronomy
PSYC	Psychology
SOCI	Social Sciences
VETE	Veterinary
MULT	Multidisciplinary

Όσον αφορά τον τύπο της δημοσίευσης, αυτή μπορεί να είναι άρθρο σε επιστημονικό περιοδικό, δημοσίευση σε συνέδριο, διορθωτικό (Erratum), κύριο άρθρο (Editorial) ή κριτική.

4.1.5 Διαδικτυακά Παρατηρητήρια (Web Observatories)

Ένα διαδικτυακό παρατηρητήριο στοχεύει στην δημιουργία μιας κατανεμημένης αρχειοθήκης δεδομένων στο Web, που αφορά κάποιου είδους δραστηριότητα, καθώς επίσης και εκείνων των εργαλείων και των μηχανισμών που θα μας δώσουν την δυνατότητα να ανακαλύψουμε το πως αυτή η δραστηριότητα εξελίσσεται στον χρόνο, να εξετάσουμε σε ποια κατάσταση βρίσκεται την παρούσα στιγμή και να σχεδιάσουμε πως αυτή μπορεί να βελτιωθεί στο μέλλον.

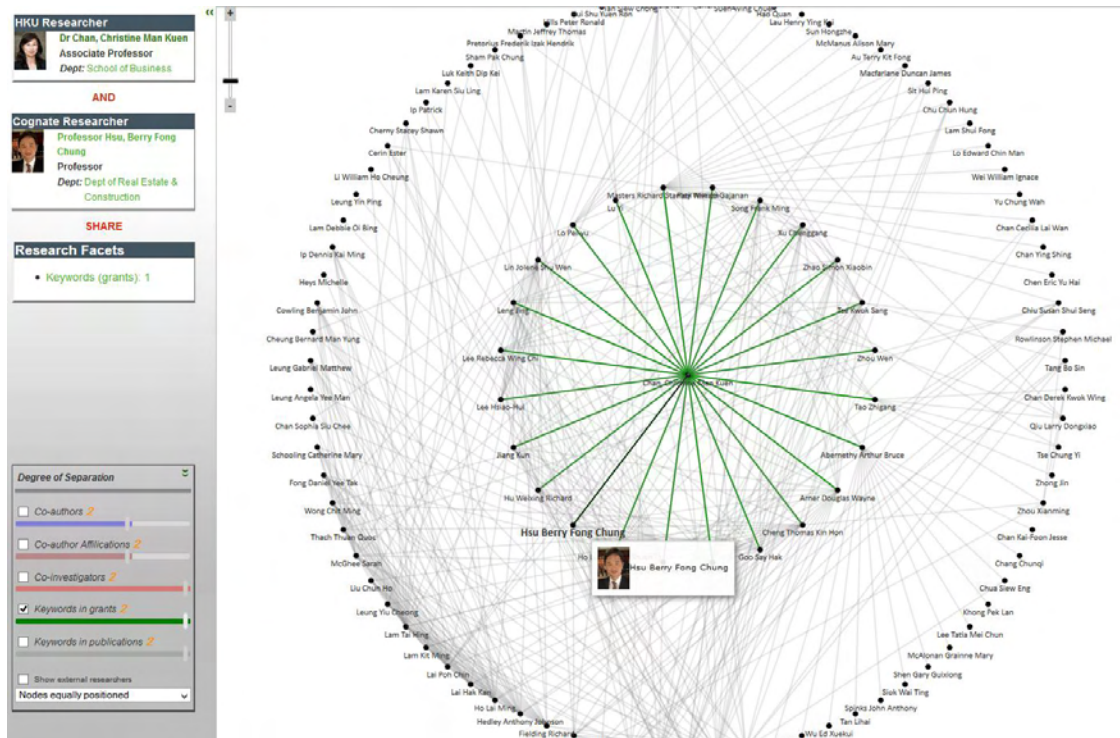
Το διαδικτυακό παρατηρητήριο παρέχει:

- Ζωντανό έλεγχο της κατάστασης του Ιστού σε όρους τοπολογίας, συνδέσεων και δραστηριότητας. Αυτό συμπεριλαμβάνει την διαθεσιμότητα των πηγών του Ιστού (έγγραφα και δεδομένα) τις σχέσεις μεταξύ των πηγών και την δραστηριότητα στο πλαίσιο των κοινωνικών δικτύων των επιχειρήσεων και της κυβέρνησης. Μία τέτοια εργασία θα πρέπει να περιλαμβάνει υπερσύγχρονες οπτικοποιήσεις, ώστε να διευκολύνει την περαιτέρω ανάλυση και κατανόηση.
- Προσομοίωση της κατάστασης του Ιστού σε συγκεκριμένα σημεία αναφοράς στο παρελθόν, βασισμένα στην πληροφορία που βρίσκεται στην αρχειοθήκη του παρατηρητηρίου, αλλά και στο μέλλον, βασισμένα σε κάποιο μοντέλο πρόβλεψης της εξέλιξης του Ιστού. Και σε αυτή τη περίπτωση οι αναλύσεις και οι προβολές θα πρέπει να είναι ισχυρά οπτικοποιημένες ώστε να επιτρέπουν συγκρίσεις και να παρέχουν ενοράσεις.

4.1.6 Παρατηρητήρια ερευνητικών δραστηριοτήτων σε πανεπιστήμια της υφελίου

Όπως αναφέραμε στην εισαγωγή, σκοπός μας είναι η δημιουργία ενός παρατηρητηρίου ερευνητικών δραστηριοτήτων για τα μέλη Δ.Ε.Π. του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας. Πριν προχωρήσουμε στην εν λόγω υλοποίηση αναζητήσαμε πανεπιστήμια που έχουν ήδη προχωρήσει σε μία αντίστοιχη υλοποίηση και μελετήσαμε κάποιες καλές πρακτικές, τις οποίες και θα παρουσιάσουμε εν συντομία παρακάτω.

Το *Πανεπιστήμιο του Hong Kong (HKU)* διαθέτει μία βάση δεδομένων που κρατά όλη την πληροφορία σχετικά με τις ερευνητικές του δραστηριότητες. Η πρόσβαση σε αυτή την πληροφορία είναι ανοιχτή και διαθέσιμη στον ιστότοπο <http://hub.hku.hk/>. Εκεί μπορεί κάποιος να αναζητήσει την πληροφορία που θέλει ορίζοντας τα πεδία αναζήτησης που τον ενδιαφέρουν. Η επιλογή του είδους της αναζήτησης γίνεται από το κεντρικό μενού: αναζήτηση ερευνητή, δημοσίευσης, thesis, grant, πατέντας. Στην περίπτωση της αναζήτησης ερευνητή ο χρήστης αποκτά πρόσβαση στο προφίλ του ερευνητή που τον ενδιαφέρει το οποίο εμπεριέχει την λίστα με τις δημοσιεύσεις του, βιβλιομετρικούς δείκτες που τον αφορούν, το δίκτυο συνεργατών του το οποίο απεικονίζεται μέσα από οπτικοποιήσεις δεδομένων δικτύου και διάφορες άλλες πληροφορίες που αφορούν τον ερευνητή (στοιχεία επικοινωνίας, βιογραφικό κτλ.). Στην περίπτωση της αναζήτησης δημοσίευσης, ανάλογα με τα κριτήρια αναζήτησης, παρουσιάζεται μία λίστα δημοσιεύσεων. Επιλέγοντας κάποια από αυτές ο χρήστης αποκτά πρόσβαση σε όλη την πληροφορία που την αφορά. Μελετώντας το παρατηρητήριο του HKU, διαπιστώσαμε ότι χρησιμοποιεί ένα εύχρηστο και σαφές σύστημα αναζήτησης. Ένα σημαντικό μειονέκτημα, είναι ότι δεν υπάρχουν αρκετές οπτικοποιήσεις δεδομένων (εκτός από την περίπτωση του δικτύου συνεργατών) που θα έδιναν την δυνατότητα στον χρήστη να βγάλει πολύτιμα συμπεράσματα.



4-2 Οπτικοποίηση Δικτύου Συνεργατών στο παρατηρητήριο του HKU.

Το [Πανεπιστήμιο Κονκόρντια](http://portico.concordia.ca/researchtest/) που βρίσκεται στο Μοντρεάλ διαθέτει ένα παρατηρητήριο δραστηριοτήτων, το οποίο δίνει έμφαση στις λέξεις κλειδιά που χρησιμοποιεί ο κάθε ερευνητής στις δημοσιεύσεις του. Ένας χρήστης της εφαρμογής η οποία είναι προσβάσιμη μέσω του ιστότοπου <http://portico.concordia.ca/researchtest/> μπορεί να πληκτρολογήσει στην φόρμα αναζήτησης το όνομα του ερευνητή ή κάποια λέξη κλειδί ή να επιλέξει ένα θεματικό αντικείμενο από μία λίστα που εμφανίζεται στο δεξί μέρος της αρχικής σελίδας. Αν ο χρήστης αναζητήσει κάποιον ερευνητή, θα μπορέσει να δει το προφίλ του, μια οπτικοποίηση με τις λέξεις κλειδιά που χρησιμοποιεί συχνότερα σε μορφή tag-cloud και μια λίστα με τους συνεργάτες ερευνητές. Εκτός από την λίστα ερευνητών στα δεξιά είναι εμφανείς οι κοινές λέξεις κλειδιά με τον κάθε συνεργάτη. Αν η αναζήτηση έχει γίνει με βάση κάποια λέξη κλειδί η λίστα θα εμπεριέχει τους ερευνητές που χρησιμοποιούν την ίδια λέξη και στα δεξιά ποιες άλλες λέξεις χρησιμοποιούν. Χρησιμοποιώντας την εφαρμογή του Πανεπιστημίου Κονκόρντια βλέπουμε ότι μπορούμε εύκολα και γρήγορα, μέσω απλών οπτικοποιήσεων να δούμε με τι θεματικό αντικείμενο ασχολείται ο κάθε ερευνητής και κατ' επέκταση το ίδρυμα. Από την άλλη, δεν μπορούμε να πάρουμε κάποια άλλη πληροφορία πέρα από τις λέξεις κλειδιά.

The screenshot shows the Concordia University research test interface. On the left, there is a navigation menu with options like 'new search', 'research', 'learn more', 'connections', and 'contact'. The main content area displays a profile for 'Lata Narayanan' with a list of keywords: computation (47), networks (15), optimization (14), telecommunication (13), wireless (08), algorithms (07), distributed systems (04), mobile computing (04), optical networks (04), wireless sensor networks (03), mobile ad hoc networks (03), wireless networks (02), routing algorithms, randomized algorithms, parallel algorithms, and theoretical computer science. Below the profile, there is a list of research collaborators: Chadi Assi (wireless networks, optimization, modeling), Brigitte Jaumard (optimization, telecommunication, networks), Roch Glitho (end-user services architectures, virtualization, cloud co...), Ferhat Khendek (software, telecommunication, service engineering), John Xiupu Zhang (quantum dot lasers, InP broadband photodetectors, radio...), Abdel R. Sebak (antennas, telecommunication, electromagnetics), M. Reza Soleymani (wireless, networks, telecommunication), and Thomas G. Fevens (medical imaging, computational geometry, computer gra...).

4-3 Προφίλ ερευνητή στο παρατηρητήριο του Πανεπιστημίου του Κονκόρντια.

4.2 Στόχος

Στόχος μας είναι ο σχεδιασμός και η ανάπτυξη ενός διαδικτυακού παρατηρητηρίου για τις ερευνητικές δραστηριότητες των μελών ΔΕΠ του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας. Το παρατηρητήριο θα παρέχει αυτού του είδους την πληροφορία συμπεριλαμβάνοντας τις επιστημονικές δημοσιεύσεις, τις ερευνητικές εργασίες, τις συνδέσεις μεταξύ των ερευνητών, τις πατέντες και τις διακρίσεις. Όσον αφορά τις συνεργασίες, θέλουμε να απεικονίσουμε τόσο τις συνεργασίες που γίνονται μέσα στα όρια του τμήματος, όσο και το ποιες

συνεργασίες αναπτύσσονται έξω από τα όρια του, δείχνοντας παράλληλα τον βαθμό της εξωστρέφειας του κάθε μέλους. Χρησιμοποιώντας βιβλιομετρικούς δείκτες μετρήσεων αλλά και τις κατάλληλες οπτικοποιήσεις δεδομένων θέλουμε να απεικονίσουμε με άμεσο και σαφή τρόπο το πώς εξελίσσεται η ερευνητική δραστηριότητα του κάθε μέλους στον χρόνο και τις θεματικές περιοχές με τις οποίες ασχολείται.

Στο σημείο αυτό πρέπει να τονισθεί ότι η παραπάνω περιγραφείσα προσπάθεια είναι αποτελεί μια προφανώς ατελή και σαφώς αρχική προσπάθεια. Η υλοποίηση ιδιαίτερα έχει την έννοια του proof-of-concept. Πλήρεις σχετικές production level υλοποιήσεις απαιτούν επιπρόσθετη προσπάθεια ευρύτερης μορφής και ισχυρότερης έντασης.

Υλοποιώντας τα παραπάνω στοχεύουμε στην υποστήριξη της διάδοσης της επιστημονικής γνώσης αλλά και την αξιοποίηση των αποτελεσμάτων της έρευνας, αναδεικνύοντας με τον τρόπο αυτό την τεχνολογική ανάπτυξη και την καινοτομία. Παράλληλα προσπαθούμε να βοηθήσουμε στην προσπάθεια που κάνει το T.H.M.M.Y. για εξωστρέφεια, διαφάνεια και προβολή του ερευνητικού του έργου.

Ενδεικτικά οφέλη για διάφορες κατηγορίες χρηστών μπορούν να συνοψιστούν ως εξής:

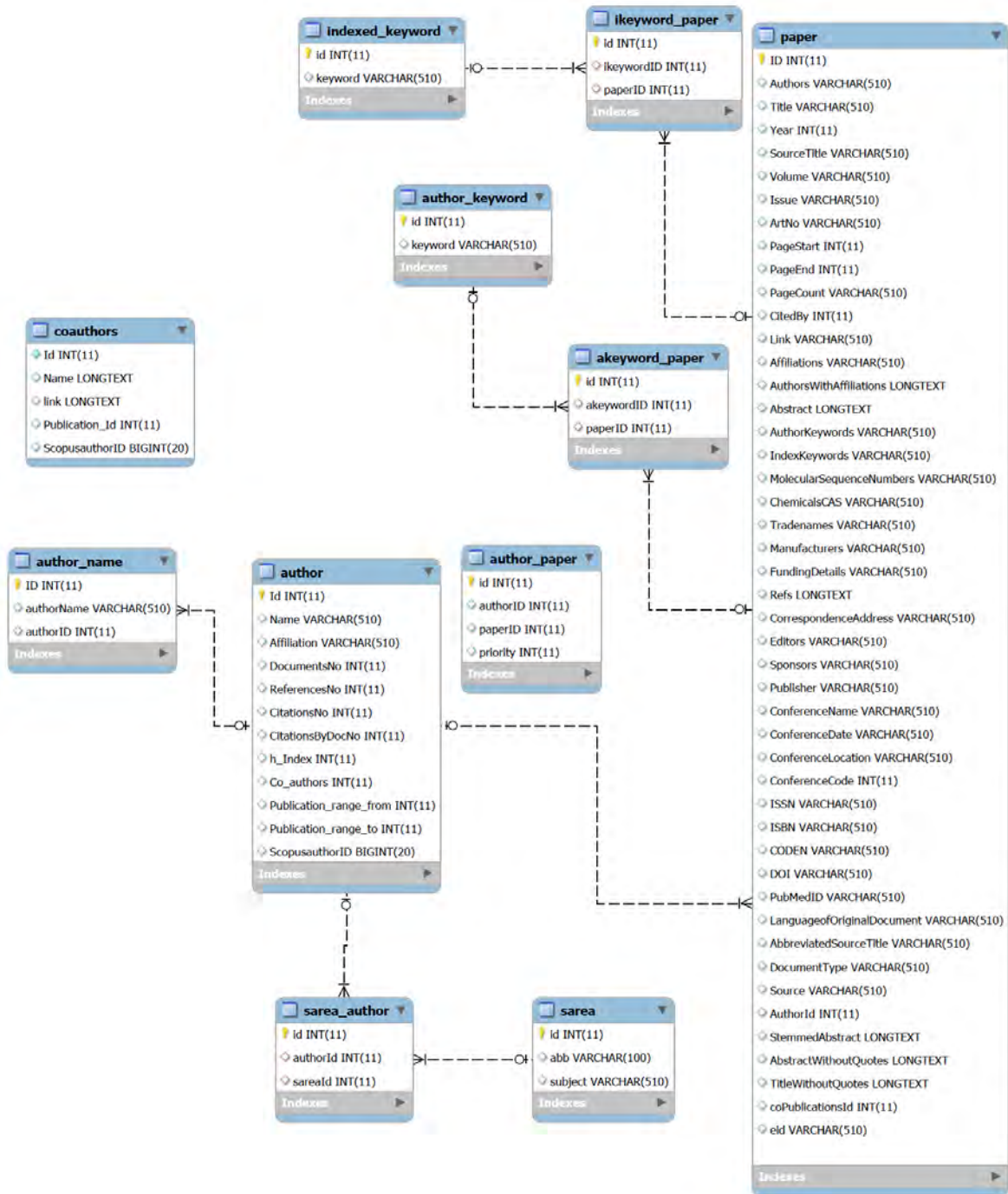
- Ένας απλός χρήστης μπορεί να αποκτήσει πρόσβαση σε αξιόπιστη πληροφορία για τους συντελεστές και τα αποτελέσματα της έρευνας που αναπτύσσεται μέσα στο T.H.M.M.Y.
- Ένας ερευνητής θα έχει στην διάθεση του ένα πολύτιμο εργαλείο που θα τον βοηθήσει στην παρακολούθηση των δραστηριοτήτων και των επιτευγμάτων του αλλά και να συγκεντρώσει έγκυρη πληροφορία για το ευρύτερο ερευνητικό περιβάλλον στο οποίο ανήκει.
- Οι διοικήσεις ερευνητικών φορέων έχουν στα χέρια τους συγκεντρωτικά στοιχεία, δομημένα στη βάση προτύπων, ειδικές αναφορές και συγκριτικές πληροφορίες για την ερευνητική δραστηριότητα του T.H.M.M.Y.
- Τα επιστημονικά συμβούλια, οι φορείς χάραξης ερευνητικής πολιτικής και οι φορείς χρηματοδότησης της έρευνας αποκτούν την απαραίτητη πληροφόρηση για τη διαμόρφωση πολιτικών στη βάση στοιχείων (evidence based policy), τη συστηματική παρακολούθηση και αποτίμηση των αποτελεσμάτων των πολιτικών αυτών, τον καθορισμό προτεραιοτήτων, το συντονισμό των ερευνητικών προσπαθειών σε εθνικό και περιφερειακό επίπεδο.
- Οι επιχειρήσεις και οι φορείς τεχνολογικής ανάπτυξης μπορούν να εντοπίζουν τους ανταγωνιστές τους και τους πιθανούς συνεργάτες τους στην ερευνητική τους δραστηριότητα.

4.3 Σχεδιασμός

4.3.1 Επεξεργασία δεδομένων και σχηματισμός βάσης δεδομένων

Για να επιτύχουμε την υλοποίηση των στόχων μας βασιζόμαστε σε πρωτογενή δεδομένα, τα οποία συγκεντρώθηκαν από το Scopus και αναλύθηκαν από την υποψήφια διδάκτωρ του T.H.M.M.Y., Μαρία Μάρκου.

Η βάση μας αποτελείται από τους παρακάτω πίνακες:



4-4 Σχηματική απεικόνιση της βάσης δεδομένων μας

- Paper: Ο πίνακας αυτός εμπεριέχει όλες τις δημοσιεύσεις και τα μεταδεδομένα τους. Κάποια από τα πεδία του πίνακα είναι οι συντάκτες της δημοσίευσης, ο τίτλος, το έτος δημοσίευσης, η σελίδα του περιοδικού, ο αριθμός των αναφορών (cites), σύνδεσμος της δημοσίευσης στο Scopus, η περίληψη, author keywords, index keywords, σπόνσορες, εκδότης, τύπος

δημοσίευσης (άρθρο, συνέδριο, διορθωτικό κτλ.) όνομα, ημερομηνία και τοποθεσία συνεδρίου (για όσες δημοσιεύσεις ανήκουν στην κατηγορία "συνέδριο" - στις υπόλοιπες NULL), ISBN, γλώσσα.

- Author_Name: Ο πίνακας αυτός περιέχει όλες τις μορφές ονομάτων των μελών όπως εμφανίζονται στα μεταδεδομένα του Scopus. Από τον πίνακα αυτόν προέκυψε ο πίνακας
- Author ο οποίος περιέχει τα μέλη ΔΕΠ του THMMY. Κάθε μέλος εμφανίζεται μία μοναδική φορά με ένα μοναδικό όνομα. Κάποια άλλα πεδία του πίνακα author : αριθμός των δημοσιεύσεων του κάθε μέλους, αριθμός αναφορών, αριθμός co-authors, h-index, έτος πρώτης δημοσίευσης, έτος τελευταίας δημοσίευσης. Ο πίνακας Author σχετίζεται με τον πίνακα Paper με βάση το ξένο κλειδί AuthorID του πίνακα Paper.
- sarea και sarea_author: Περιέχουν τα επιστημονικά πεδία και σε ποια από αυτά δραστηριοποιείται το κάθε μέλος.
- Author_Keyword και Akeyword_paper περιέχουν τα author keywords και το id της δημοσίευσης στην οποία εμφανίζονται και αντίστοιχα
- indexed_keyword και ikeyword_paper για τα index keywords.

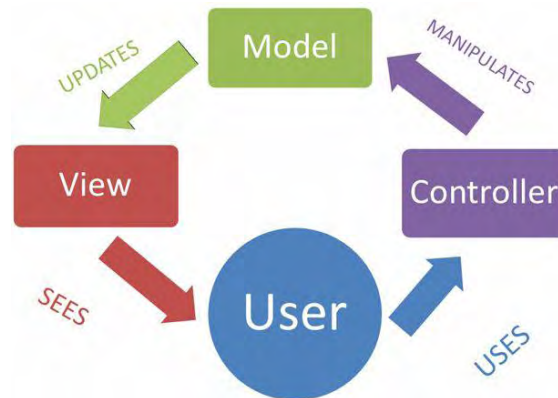
4.3.2 Model View Controller (MVC)

Αφού μελετήσαμε την βάση δεδομένων μας, επιλέξαμε την αρχιτεκτονική MVC (Model View Controller) για την υλοποίηση της εφαρμογής μας τόσο από την πλευρά του πελάτη όσο και από την πλευρά του εξυπηρετητή. Σκοπός του MVC είναι να διαιρέσει ένα συστατικό σε τρία λογικά μέρη το μοντέλο (model), την αναπαράσταση (view) και την διαχείριση (controller) καθιστώντας ευκολότερη την διαδικασία τροποποίησης του κάθε μέρους. Παρακάτω θα δούμε αναλυτικότερα τα τρία αυτά λογικά μέρη:

- Στο model τοποθετούμε τις λειτουργίες της εφαρμογής που σχετίζονται με την πρόσβαση στη βάση δεδομένων. Οι λειτουργίες αυτές είναι με τη μορφή function (μεθόδων στον προγραμματισμό). Είναι κάποιες συναρτήσεις με τις οποίες εκτελούμε διάφορες λειτουργίες διαχείρισης των δεδομένων που λαμβάνουμε από τη βάση. Για παράδειγμα, για να εμφανίσουμε κάποιο προφίλ ενός ερευνητή, όπως αυτό είναι καταχωρημένο στην βάση δεδομένων, το πρώτο βήμα είναι ότι στο model των ερευνητών (θα μιλήσουμε σε λίγο γι αυτό) υπάρχει κάποια συνάρτηση για παράδειγμα η `public List getProfile(String lastname)` η οποία περιέχει κώδικα που μιλάει με τη βάση και τραβάει τα δεδομένα που θέλουμε.
- Μέσα στη view υπάρχει το HTML της σελίδας της εφαρμογής μας. Είναι αυτό που βλέπουμε. Τις περισσότερες φορές μία view μιλάει με ένα controller και αφού ο controller κάνει τις διάφορες επεξεργασίες των δεδομένων στέλνει στη view συγκεκριμένα δεδομένα να εμφανίσει.
- Ο controller είναι ο μεσίτης μεταξύ model και view. Ελέγχει το πώς "τρέχει" η εφαρμογή. Μιλάει με το model, παίρνει τα δεδομένα που ζητά και εν συνεχεία και αφού τα επεξεργαστεί τα στέλνει πίσω στη view για απεικόνιση.

Στην αρχιτεκτονική MVC συμβαίνουν τα εξής: Ο χρήστης δίνει κάποια είσοδο στην εφαρμογή. Για παράδειγμα συμπληρώνει μία φόρμα και πατάει το κουμπί "submit". Στη συνέχεια ο controller έχοντας λάβει την είσοδο του χρήστη μιλάει με το model χρησιμοποιώντας την είσοδο του χρήστη σαν

μεταβλητή και ζητάει δεδομένα (από το model), τα οποία όταν τα λαμβάνει τα προσαρμόζει ανάλογα με αυτό που ζήτησε ο χρήστης. Έπειτα, με τα νέα δεδομένα πλέον, ανανεώνει την view.



4-5 Σχηματική απεικόνιση της MVC αρχιτεκτονικής. Πηγή: <http://www.c-sharpcorner.com/UploadFile/3d39b4/Asp-Net-mvc-introduction/Images/MVC-Introduction2.jpg>

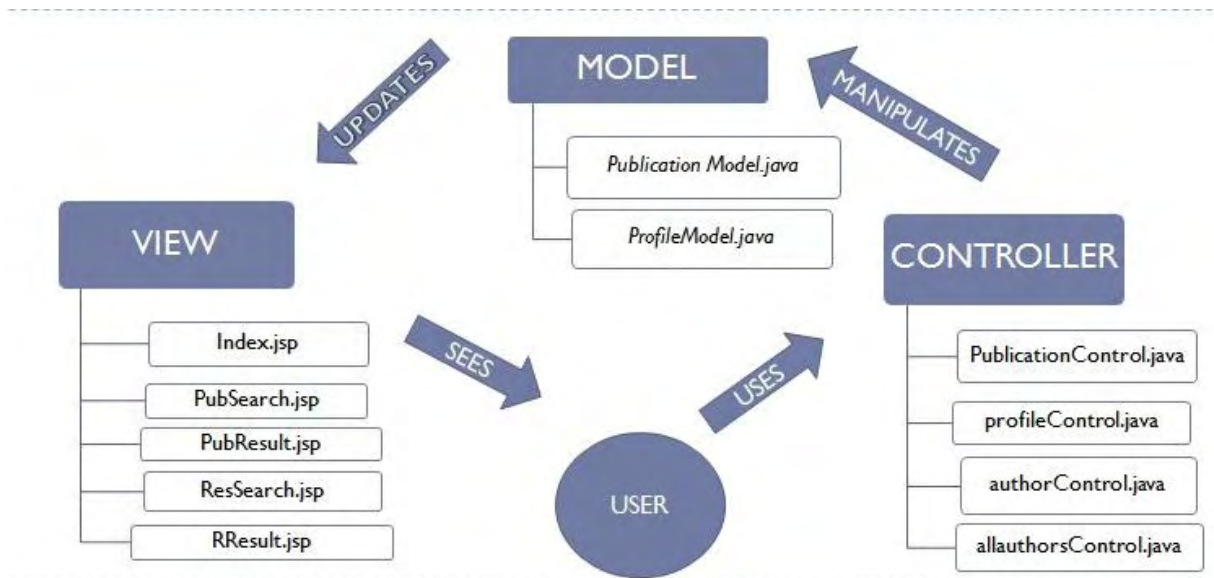
Πιο συγκεκριμένα στην περίπτωση μας, όταν κάποιος χρήστης της εφαρμογής μας πραγματοποιήσει μία αναζήτηση για δημοσιεύσεις που συνέβησαν μέσα σε ένα συγκεκριμένο χρονικό διάστημα, ο controller των δημοσιεύσεων, θα επικοινωνήσει με την συνάρτηση `getPublications` του model των δημοσιεύσεων, ζητώντας όλες τις δημοσιεύσεις που ανήκουν μέσα στο χρονικό διάστημα που ορίζεται από τις παραμέτρους `yearfrom` και `yearto`, που προκύπτουν από την είσοδο του χρήστη. Το model θα εντοπίσει τις δημοσιεύσεις αυτές στην βάση δεδομένων και θα τα επιστρέψει στον controller, ο οποίος θα τα εμφανίσει παραδίδοντάς τα στο view κομμάτι της αρχιτεκτονικής μας.

Για την υλοποίηση της εφαρμογής μας αρχικά δημιουργήσαμε ένα πρότυπο για την ιστοσελίδα στην οποία ενσωματώνονται οι εφαρμογές μας, καθορίζοντας τις θέσεις και τις περιοχές τις επικεφαλίδας, του μενού, του κυρίου σώματος. Κατόπιν, δημιουργήσαμε το μενού και καθορίσαμε τα πεδία αναζήτησης τόσο για την περίπτωση της αναζήτησης ερευνητή όσο και για την περίπτωση την αναζήτησης δημοσίευσης. Αφού προστέθηκαν και οι απαραίτητοι έλεγχοι για τις δυνατές τιμές σε κάθε πεδίο της φόρμας αναζήτησης, στην πλευρά του πελάτη, εφαρμόσαμε μια αρχιτεκτονική MVC για την επικοινωνία του πελάτη με τον εξυπηρετητή.

Στο VIEW κομμάτι υπάρχουν τα αρχεία που καθορίζουν την μορφή της κάθε σελίδας της εφαρμογής μας. Χρησιμοποιούμε επίσης 4 controllers:

- Τον `allauthorsControl` που ζητάει από το MODEL την λίστα με τα ονόματα όλων των μελών ΔΕΠ
- Τον `authorControl` που ζητάει τα ονόματα των μελών που ταιριάζουν με τα κριτήρια αναζήτησης που εισήγαγε ο χρήστης στο VIEW κομμάτι,
- Τον `ProfileControl` που ζητάει τα στοιχεία του προφίλ κάποιου μέλους
- Τον `PublicationControl` που ζητάει την λίστα με τις δημοσιεύσεις που ταιριάζουν στα κριτήρια αναζήτησης που εισήγαγε προηγουμένως ο χρήστης.

Όσον αφορά το MODEL μέρος, αυτό αποτελείται από δύο αρχεία ένα μοντέλο δημοσιεύσεων και ένα μοντέλο ερευνητών, τα οποία δέχονται τις αιτήσεις των παραπάνω controller, επικοινωνούν με την βάση δεδομένων και απαντούν σε αυτούς δίνοντας τους τα αποτελέσματα.



Εικόνα 6 Σχηματική απεικόνιση της MVC αρχιτεκτονικής της υλοποίησής μας

Χτίζοντας μία εφαρμογή με MVC έχουμε τα εξής βασικά πλεονεκτήματα:

- Διαχωρισμός Προβλημάτων (Separation of Concerns): Αυτό είναι και το πιο βασικό πλεονέκτημα του MVC. Ουσιαστικά δημιουργείται μία εφαρμογή η οποία έχει τρία επίπεδα, το επίπεδο των models, το επίπεδο των controllers και το επίπεδο των views -που θα αναλυθούν παρακάτω- και το κάθε επίπεδο επιτελεί ξεχωριστό έργο και ταυτόχρονα συνεργάζεται με τα άλλα επίπεδα. Μία σωστή MVC εφαρμογή είναι εκείνη που τα τρία επίπεδα είναι ξεκάθαρα καθορισμένα και δεν συμπλέκονται. Για παράδειγμα είναι λάθος στο επίπεδο των View να υπάρχει κώδικας που επικοινωνεί με την βάση δεδομένων και τραβάει δεδομένα.
- Επεκτασιμότητα: Το δεύτερο πλεονέκτημα της MVC αρχιτεκτονικής είναι πολύ σημαντικό επίσης. "Επεκτασιμότητα" είναι η δυνατότητα που διαθέτει μία εφαρμογή, κατά την οποία μπορούμε μελλοντικά να προσθέσουμε λειτουργίες σε αυτή ή να αλλάξουμε κάποιες από τις ήδη υπάρχουσες λειτουργίες και να έχουμε άλλα αποτελέσματα. Τα προγράμματα που είναι φτιαγμένα με MVC αρχιτεκτονική έχουν βασικό χαρακτηριστικό ότι είναι επεκτάσιμα.
- Ελεγχιμότητα (Testability): Οι MVC εφαρμογές έχουν την δυνατότητα να είναι ελέγξιμες και με τον τρόπο αυτό συντηρούνται πιο εύκολα. Για παράδειγμα, αν θέλουμε να ελέγξουμε αν η αναζήτηση ερευνητή λειτουργεί σωστά, ελέγχουμε τον Controller authorControl ο οποίος περιέχει κώδικα που διαχειρίζεται τα δεδομένα που εισήχθησαν από τον χρήστη. Αυτός ο controller θεωρείται μία μονάδα. Στο MVC μπορούμε με πολλή ευκολία να γράψουμε απλό κώδικα με τον οποίο τεστάρουμε αυτόν τον controller αλλά και κάθε μία από τις λειτουργίες του. Παίρνουμε τα αποτελέσματα και βλέπουμε αν η συγκεκριμένη μονάδα της εφαρμογής μας λειτουργεί σωστά.
- "Καθαρά" URLs. Τα περισσότερα MVC frameworks για web applications δίνουν τη δυνατότητα να έχουμε "καθαρά" urls.

4.3.3 Οπτικοποίηση της πληροφορίας

Η πληροφορία που μεταβιβάζει ο controller στο view μέρος είναι αναγκαστικά σε μορφή κειμένου αφού προέκυψε από την επικοινωνία του model με την βάση δεδομένων. Για παράδειγμα, στο προφίλ κάποιου μέλους εμφανίζουμε την λίστα με τα ονόματα των συνεργατών του, τα περισσότερα μέλη έχουν αναπτύξει πολλές συνεργασίες και έτσι το αποτέλεσμα που αρχικά βλέπουμε είναι μια λίστα μεγάλης έκτασης. Το μόνο που μπορούμε να κάνουμε με την λίστα αυτή είναι να την διαβάσουμε σειριακά. Με τον τρόπο αυτό όμως δεν μπορούμε να κάνουμε καμία εκτίμηση για το είδος των συνεργασιών του κάθε μέλους. Αν όμως με την χρήση των κατάλληλων εργαλείων επεξεργαστούμε τα δεδομένα που έχει επιστρέψει ο controller στο view και τα οπτικοποιήσουμε τότε θα μπορούσαμε να αντλήσουμε πολύτιμη γνώση.

Στόχος μας είναι τόσο η οπτικοποίηση όλων των αποτελεσμάτων που έχουν προκύψει από τις επερωτήσεις του χρήστη στην βάση δεδομένων μας, όσο και η υλοποίηση κάποιων άλλων οπτικοποιήσεων που αφορούν την συνολική εικόνα του T.H.M.M.Y..

Αρχικά, μελετήσαμε προσεκτικά τα δεδομένα μας ως προς την φύση, την δομή, το είδος και τα χαρακτηριστικά τους. Κατόπιν, έχοντας υπ' όψιν τους παραπάνω στόχους προσπαθήσαμε να σκεφτούμε τους καλύτερους τρόπους για να τα αναπαραστήσουμε οπτικά. Στο κεφάλαιο 5 θα αναφερθούμε χωριστά σε κάθε οπτικοποίηση που αναπτύξαμε.

4.4 Λειτουργίες

Κάποιος χρήστης της Web εφαρμογής μας θα έχει την δυνατότητα να εκτελέσει ένα σύνολο λειτουργιών:

- Δυνατότητα αναζήτησης ερευνητή είτε μέσω επιλογής από μία λίστα είτε μέσω συμπλήρωσης των πεδίων στην φόρμα αναζήτησης.
- Δυνατότητα αναζήτησης δημοσιεύσεων που πληρούν κάποια από (ή όλα) τα κριτήρια που έχει καθορίσει ο χρήστης στην φόρμα αναζήτησης.
- Προβολή προφίλ ερευνητή. Στο προφίλ απεικονίζονται πληροφορίες όπως ο αριθμός των δημοσιεύσεων, το H-index, ο αριθμός των αναφορών, ένας χάρτης που δείχνει σε ποια συνέδρια έχει λάβει μέρος, οι συνεργάτες του μέλους, οι θεματικές περιοχές ενδιαφέροντος, οι συχνότερες λέξεις κλειδιά που χρησιμοποιεί στις δημοσιεύσεις του και οι λίστα με όλες τις δημοσιεύσεις.
- Προβολή της λίστας των δημοσιεύσεων. Προσφέρεται η δυνατότητα ταξινόμησης τους, αν ο χρήστης κάνει κλικ στην επικεφαλίδα κάποιας στήλης. Η λίστα είναι σελιδοποιημένη και αρχικά εμφανίζονται 10 αποτελέσματα. Μπορεί όμως ο χρήστης να επιλέξει πόσα αποτελέσματα να εμφανίζονται στην κάθε σελίδα.
- Εξερεύνηση της πληροφορίας μέσα από διαδραστικές οπτικοποιήσεις δεδομένων. Ο χρήστης λαμβάνει τον μεγαλύτερο όγκο της πληροφορίας άμεσα μέσα από τις οπτικοποιήσεις και με τον τρόπο αυτό λαμβάνει γνώση σχετική με τους συντελεστές και τα αποτελέσματα της σύγχρονης έρευνας που αναπτύσσεται μέσα στο THMMY. Μπορεί επίσης να αλληλεπιδράσει

με τις οπτικοποιήσεις για να εστιάσει στο συγκεκριμένο πεδίο γνώσης που τον ενδιαφέρει αλλά και να προχωρήσει σε συγκρίσεις, αξιολογώντας παράλληλα τα αποτελέσματα.

4.5 Εργαλεία

Για την υλοποίηση της εφαρμογής μας χρησιμοποιήσαμε τα εργαλεία οπτικοποίησης δεδομένων που αναλύσαμε στην ενότητα 2.7. Εκτός όμως από αυτά, για την υλοποίηση της Web Εφαρμογής μας τόσο από την πλευρά του πελάτη όσο και από την πλευρά του εξυπηρετητή κάναμε χρήση των εργαλείων που θα αναλύσουμε παρακάτω.

4.5.1 HTML 5, CSS3 , JAVASCRIPT, JQuery

Η **HTML5** είναι η τελευταία έκδοση της προδιαγραφής της html. Το προηγούμενο πρότυπο (html4) υπάρχει από το 1999 όταν το web ήταν τελείως διαφορετικό από ότι είναι σήμερα. Η html5 βρίσκεται ακόμα σε εξέλιξη, παρόλα αυτά οι πιο δημοφιλείς φυλλομετρητές υποστηρίζουν τα περισσότερα από τα νέα στοιχεία της html5 και των API.

Με την χρήση της HTML5 και όσον αφορά την γενικότερη προσπάθειά μας επιτυγχάνουμε:

- Καλύτερη ενσωμάτωση των πολυμέσων, απελευθέρωση από επιπρόσθετα λογισμικά (π.χ. Flash), με την χρήση της εγγενούς υποστήριξης για ήχο και βίντεο.
- Περισσότερο περιγραφική επισήμανση
- Με την χρήση του canvas, μετασχηματισμών και javascript μπορούμε να δημιουργήσουμε εξαιρετικά καλαίσθητες διεπαφές και κινούμενα σχέδια.
- Λειτουργικότητες αποθήκευσης από την πλευρά του πελάτη και ως επακόλουθο offline πρόσβαση στο web.
- Δυνατότητα χρήσης εργατών (web workers) για να δώσουμε ευαισθησία προσαρμογής διαστάσεων (responsiveness) στην σελίδα μας.
- Ευκολότερος σχεδιασμός εφαρμογών για κινητές συσκευές.

Η **CSS3** είναι το νέο πρότυπο για την css. Τα πιο σημαντικά από τα καινούργια στοιχεία που συμπεριλαμβάνονται στην CSS3 είναι:

- Νέοι επιλογείς (selectors), δηλαδή πρότυπα που επιλέγουν το στοιχείο που θέλουμε να δώσουμε στυλ.
- Καλύτερο μοντέλο κουτιού (Box model) και επιλογές διάταξης.
- Φόντο και σύνορα
- Δυσδιάστατοι και τρισδιάστατοι μετασχηματισμοί
- Κινούμενα σχέδια
- Διάταξη πολλαπλών στηλών
- Περισσότερες δυνατότητες σχεδίασης διεπαφών χρήστη

Για να επιτύχει η νέα προδιαγραφή της HTML τους σκοπούς της απαραίτητη είναι η χρήση μιας σειράς τεχνολογιών όπως: Javascript, Canvas, web workers, local storage, geolocation. Οι τεχνολογίες αυτές εξελίσσονται ώστε να εξυπηρετήσουν τις νέες απαιτήσεις που προκύπτουν από την καθιέρωση της HTML5 ως προγραμματιστικό πρότυπο για εφαρμογές Ιστού.

Η *JavaScript (JS)* είναι μια διερμηνευμένη γλώσσα προγραμματισμού. Αρχικά αποτέλεσε μέρος της υλοποίησης των φυλλομετρητών Ιστού ώστε τα σενάρια από την πλευρά του πελάτη (client-side scripts) να μπορούν να επικοινωνούν με τον χρήστη, να ανταλλάσσουν δεδομένα ασύγχρονα και να αλλάζουν δυναμικά το περιεχόμενο του εγγράφου που εμφανίζεται.

Η JavaScript είναι μια γλώσσα σεναρίων που βασίζεται στα πρωτότυπα (prototype-based), είναι δυναμική, με ασθενείς τύπους και έχει συναρτήσεις ως αντικείμενα πρώτης τάξης. Η σύνταξή της είναι επηρεασμένη από τη C. Η JavaScript αντιγράφει πολλά ονόματα και συμβάσεις ονοματοδοσίας από τη Java, αλλά γενικά οι δύο αυτές γλώσσες δε σχετίζονται και έχουν πολύ διαφορετική σημασιολογία. Η γλώσσα υποστηρίζει αντικειμενοστραφές, προστακτικό και συναρτησιακό στυλ προγραμματισμού.

Οι νεότερες εικονικές μηχανές και πλαίσια ανάπτυξης για JavaScript (όπως το Node.js) έχουν επίσης κάνει τη JavaScript πιο δημοφιλή για την ανάπτυξη εφαρμογών Ιστού στην πλευρά του διακομιστή (server-side).

Η *jQuery* είναι μια βιβλιοθήκη Javascript, σχεδιασμένη με σκοπό να απλοποιήσει την υλοποίηση σεναρίων (scripting) στη πλευρά του πελάτη (client-side) της HTML. Η jQuery είναι ένα ελεύθερο λογισμικό που χρησιμοποιείται σε πάνω από το 65% των 10.000 ιστοτόπων με τη μεγαλύτερη επισκεψιμότητα. Χαρακτηριστικά της jQuery αποτελούν τα εξής:

- DOM διάσχιση και τροποποίηση (υποστηρίζοντας CSS 1-3)
- Χειρισμός DOM βασισμένος σε CSS επιλογείς που χρησιμοποιεί τα id και class σαν κριτήρια για να κατασκευάσει επιλογείς.
- Events
- Εφέ και κινητά στοιχεία
- AJAX
- Επεκτασιμότητα μέσω plug-ins
- Εργαλεία όπως πληροφορίες user-agent, ανίχνευση χαρακτηριστικών.
- Μεθόδους συμβατότητας που είναι εγγενώς διαθέσιμες σε σύγχρονα προγράμματα περιήγησης.

4.5.2 JSP, JAVA SERVLETS

Οι *JSPs (JavaServer Pages)* είναι μια τεχνολογία που έχει δημιουργηθεί από την εταιρεία Sun Microsystems για να μπορεί να δημιουργεί δυναμικό περιεχόμενο (dynamic content) στο Web. Πρόκειται για HTML έγγραφα (ιστοσελίδες) τα οποία αναμειγνύονται με τη γλώσσα προγραμματισμού Java, η οποία και έχει τη δυνατότητα να παρέχει (δημιουργεί) αυτό το δυναμικό περιεχόμενο.

Οι JSPs είναι μια εφαρμογή στην πλευρά του server (server-side application), που σημαίνει ότι δέχονται μια αίτηση (request) και παράγουν μια απόκριση ή απάντηση (response). Σε γενικές γραμμές, οι αιτήσεις γίνονται από έναν Web client και η απόκριση είναι ένα παραγόμενο HTML έγγραφο (ιστοσελίδα) το οποίο στέλνεται πίσω στον Web client. Επειδή οι JSPs είναι μια εφαρμογή στην πλευρά του server, έχουν πρόσβαση σε πηγές (resources) στον server, όπως είναι τα Servlets, JavaBeans, EJBs, αλλά και σε βάσεις δεδομένων.

Υπάρχουν πολλά πλεονεκτήματα από τη χρήση των JavaServer Pages. Επειδή οι JSPs χρησιμοποιούν τη γλώσσα προγραμματισμού Java, ακολουθούν την πολιτική "write-once, run-anywhere". Αυτό σημαίνει

ότι μια JSP μπορεί να εκτελεσθεί σε έναν οποιονδήποτε application server ο οποίος υποστηρίζει τις JSPs χωρίς να χρειασθεί κάποια τροποποίηση στον κώδικα. Ένα άλλο πλεονέκτημα των JSPs είναι η χρήση των tag libraries. Οι JSPs χρησιμοποιούν τα tags, τα οποία είναι παρόμοια μ' αυτά της HTML και της XML, για να εισάγουν δυναμικό περιεχόμενο (dynamic content). Τα tag libraries ορίζουν επιπλέον tags τα οποία μπορούν να χρησιμοποιηθούν για να αντικαταστήσουν τμήματα κώδικα. Ένα άλλο σημαντικό πλεονέκτημα των JSPs είναι ο διαχωρισμός των ρόλων. Οι προδιαγραφές των JSPs επιτρέπουν να μοιραστεί το φορτίο σε δύο κατηγορίες : στο γραφικό περιεχόμενο της σελίδας και στο δυναμικό περιεχόμενο της σελίδας. Αυτό σημαίνει στην πράξη ότι η ομάδα που δεν γνωρίζει τη γλώσσα προγραμματισμού Java μπορεί να δημιουργήσει το γραφικό περιεχόμενο της σελίδας και ένας προγραμματιστής της Java να δημιουργήσει το δυναμικό περιεχόμενο της σελίδας.

Όταν γράφουμε ένα JSP, είναι ευκολότερο να γράψουμε πρώτα τον HTML κώδικα και μετά να εισάγουμε τον κώδικα της Java για να δημιουργήσουμε το δυναμικό περιεχόμενο.

Οι ιστοσελίδες που περιέχουν κώδικα JSP (JavaServer Pages), δηλ. ανάμιξη κώδικα HTML με κώδικα Java, μετατρέπονται (μεταφράζονται ή μεταγλωττίζονται) σε [Servlets](#) πριν εκτελεσθούν στον server. Ένα Servlet, στη γενική του μορφή, είναι μια τάξη (class) της Java που υλοποιεί (implements) το interface Servlet, δέχεται αιτήσεις (requests) και παράγει (δημιουργεί) αποκρίσεις (responses). Οι αιτήσεις μπορεί να προέρχονται από τάξεις της Java, από Web clients ή και από άλλα Servlets.

Όταν υλοποιούμε ένα interface λέμε ότι η τάξη μάς παρέχει υλοποιήσεις για τις μεθόδους που είναι δηλωμένες στο interface. Συνεπώς, όταν υλοποιούμε το interface Servlet δηλώνουμε ότι ο κώδικάς μας θα παρέχει υλοποιήσεις για τις μεθόδους που βρίσκονται στο interface Servlet. Θα ασχοληθούμε μ' ένα μόνο συγκεκριμένο είδος Servlet, το HttpServlet, το οποίο δέχεται HTTP requests και παράγει HTTP responses. Όταν γράφουμε το δικό μας HttpServlet, δεν υλοποιούμε το interface Servlet απευθείας, αλλά επεκτείνουμε (extend) την τάξη HttpServlet.

Υπάρχουν πολλές μέθοδοι που υπερκαλύπτουμε (override) όταν γράφουμε ένα HttpServlet, όπως :

- void init(ServletConfig sc) throws ServletException;
- void Service(HttpServletRequest request, HttpServletResponse response) throws ServletException, IOException;
- void doGet(HttpServletRequest request, HttpServletResponse response) throws ServletException, IOException;
- void doPost(HttpServletRequest request, HttpServletResponse response) throws ServletException, IOException;

Η μέθοδος [init\(\)](#) καλείται για να αποδώσει αρχικές τιμές σε ορισμένες παραμέτρους. Η μέθοδος [Service\(\)](#) χρησιμοποιείται για να ελέγξουμε τις καταχωρήσεις του χρήστη όταν δεν μας ενδιαφέρει αν η αίτησή (request) του προήλθε από μια μέθοδο GET ή μια μέθοδο POST. Αν, όμως, έχει σημασία το αν η HTML φόρμα στάλθηκε με μια μέθοδο GET ή μια μέθοδο POST, τότε θα πρέπει να υπερκαλύψουμε μια από τις μεθόδους doGet() ή doPost().

Μια HTML φόρμα διαθέτει την ιδιότητα (attribute) METHOD η οποία ορίζει τον τρόπο με τον οποίο θα σταλούν τα δεδομένα στον server. Η μέθοδος GET προσαρτά τα δεδομένα στο URL και τα στέλνει στον server μ' αυτόν τον τρόπο, ενώ η μέθοδος POST συσκευάζει τα δεδομένα σ' ένα πακέτο και στέλνει το πακέτο στον server.

Αφού τα JSPs μετατρέπονται σε Servlets, τότε θα μπορεί να ρωτήσει κάποιος γιατί να πρέπει να μάθουμε και να χρησιμοποιούμε και τα δύο; Ο βασικότερος λόγος για το γιατί πρέπει να χρησιμοποιούμε τα JSPs είναι ότι δεν υπάρχει λόγος το servlet μας να δημιουργεί το στατικό μέρος μίας σελίδας με `println`, γιατί μια τέτοια διαδικασία είναι εξαιρετικά χρονοβόρα και επίπονη για τον προγραμματιστή. Μία JSP σελίδα περιέχει κανονικό HTML κώδικα για τα στατικά μέρη της σελίδας, και ενσωματωμένο κώδικα Java για τα δυναμικά μέρη.

Στην εφαρμογή μας χρησιμοποιούμε JSP για την υλοποίηση της πλευράς του πελάτη, που αφορά το πως θα εμφανίζεται το δυναμικό περιεχόμενο της ιστοσελίδας. Από την άλλη, χρησιμοποιούμε Servlets για να υλοποιήσουμε τον μηχανισμό επικοινωνίας του πελάτη με τον εξυπηρετητή, υπερκαλύπτοντας τις μεθόδους GET και POST. Με τον τρόπο αυτό το Servlet στέλνει τις αιτήσεις μας στον εξυπηρετητή, ο οποίος τις επεξεργάζεται και πάλι μέσω Servlet δίνει την απάντηση στον πελάτη. Τέλος, στην πλευρά του πελάτη με χρήση της JSP εμφανίζουμε το δυναμικό περιεχόμενο.

4.5.3 Apache Tomcat

Αναφερθήκαμε παραπάνω για τα Java Servlets χωρίς όμως να επισημάνουμε πως αυτά επικοινωνούν με τον εξυπηρετητή. Η επικοινωνία αυτή γίνεται μέσω ενός container όπως είναι το Apache Tomcat για τον Apache Web Server.

Το εργαλείο Apache Tomcat είναι μια ανοιχτού κώδικα υλοποίηση της τεχνολογίας των σελίδων διακομιστών Java (Servlet και JSP), τρέχει java εφαρμογές Ιστού και μπορεί να σερβίρει JSPs/servlets.

Με το εργαλείο παρέχεται η δυνατότητα:

- ρύθμισης του διαχειριστή ασφάλειας της Java
- χρησιμοποίησης πηγών δεδομένων μέσω JDBC
- υποστήριξης του πρωτοκόλλου SSL για κρυπτογραφημένες συνδέσεις
- χρήσης διαμεσολαβητών (proxy)
- εικονικής λειτουργίας (virtual hosting)
- εκτέλεσης σεναρίων CGI.

4.5.4 MySQL

Η MySQL είναι ένα σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων. Το πρόγραμμα τρέχει έναν εξυπηρετητή παρέχοντας πρόσβαση πολλών χρηστών σε ένα σύνολο βάσεων δεδομένων. Η MySQL ανήκει και χρηματοδοτείται από την κερδοσκοπική εταιρία, MySQL AB, η οποία σήμερα ανήκει στην Oracle. Η MySQL είναι δημοφιλής βάση δεδομένων για διαδικτυακά προγράμματα και ιστοσελίδες. Χρησιμοποιείται σε κάποιες από τις πιο διαδεδομένες διαδικτυακές υπηρεσίες, όπως το Flickr, το Youtube, η Wikipedia, η Google, το Facebook και το Twitter.

4.5.5 NetBeans

Το NetBeans είναι ένα ερευνητικό έργο ανοιχτής πηγής (open source) με μεγάλο αριθμό χρηστών, που ιδρύθηκε από την Sun Microsystems. Το NetBeans IDE είναι ένα περιβαλλοντικό ανάπτυγμα IDE - ένα εργαλείο στους προγραμματιστές για να γράψουν, να κάνουν compile, debug και να αναπτύξουν

προγράμματα. Είναι γραμμένο σε Java - αλλά μπορεί να υποστηρίξει όλες τις γλώσσες προγραμματισμού. Υπάρχει επίσης ένας μεγάλος αριθμός λειτουργικών μονάδων που βοηθάνε στην επέκταση της λειτουργικότητας του NetBeans IDE. Το NetBeans IDE είναι ένα ελεύθερο προϊόν δίχως περιορισμούς στον τρόπο χρησιμοποίησής του.

4.5.6 Ωκεανός (Okeanos)

Στον κόσμο μας ο ωκεανός αντιπροσωπεύει την αφθονία. Οι ωκεανοί έχουν μετασηματίσει τον κόσμο μας. Συλλαμβάνουν, αποθηκεύουν και διοχετεύουν ενέργεια, οξυγόνο και ζωή στον πλανήτη μας. Οι ωκεανοί είναι οι αστέρευτες πηγές, των γήινων πόρων και αυτό αποτελεί και την σημαντικότερη διαφορά με τους πλανήτες που δεν υπάρχει ζωή.

Η ΕΔΕΤ Α.Ε. (GRnet) που έχει ως αντικείμενο τη διαχείριση του Εθνικού Δικτύου Έρευνας και Τεχνολογίας κατά το πρότυπο των αντίστοιχων Ερευνητικών και Εκπαιδευτικών Δικτύων της Ευρωπαϊκής Ένωσης, δραστηριοποιείται ενεργά, τόσο σε εθνικό όσο και σε διεθνές επίπεδο και αποτελεί περιβάλλον ανάπτυξης καινοτόμων υπηρεσιών, στη διάθεση της Εκπαιδευτικής και Ερευνητικής Κοινότητας της χώρας. Με βάση την παραπάνω ιδέα, η ΕΔΕΤ ανέπτυξε μία υπηρεσία σύννεφου, τον Ωκεανό (Okeanos) ο οποίος είναι ελεύθερος προς χρήση για την ελληνική ακαδημαϊκή και ερευνητική κοινότητα.

Ο Ωκεανός είναι μία καινούργια IaaS υπηρεσία που αυτή τη στιγμή μόλις έχει ολοκληρώσει την λειτουργία σε επίπεδο Alpha και σύντομα ξεκινά την λειτουργία σε έκδοση Beta με περισσότερες λειτουργίες, ενώ ταυτόχρονα θα γίνει ανοιχτά διαθέσιμη σε οποιοδήποτε μέλος ενός ελληνικού δημοσίου πανεπιστημίου, απλά με την εισαγωγή των ακαδημαϊκών του διαπιστευτηρίων (όνομα χρήστη, κωδικός πρόσβασης).

Η συντόμευση «IaaS», που σημαίνει «Infrastructure as a Service» αναφέρεται στην ουσία στην δυνατότητα που έχει κάποιος να δημιουργήσει τον δικό του υπολογιστή, που θα είναι πάντα συνδεδεμένος στο διαδίκτυο, χωρίς να ανησυχεί για αποτυχίες του υλικού, προβλήματα συνδεσιμότητας, καλώδια και προβλήματα λογισμικού.

Με τον Ωκεανό κάποιος βρίσκεται μόλις ένα κλικ μακριά από την δικιά του εικονική μηχανή και τα δικά του εικονικά δίκτυα. Μπορεί να τα διαχειριστεί, να τα καταστρέψει, να τα συνδέσει και να πραγματοποιήσει χρήσιμες ενέργειες επί αυτών, και όλα αυτά μέσω του αγαπημένου του φυλλομετρητή. Μπορεί επίσης να αποθηκεύσει τα αρχεία του online στο σύννεφο, να τα μοιραστεί με τους φίλους του και να έχει πρόσβαση σε αυτά οποιαδήποτε στιγμή, από οποιοδήποτε μέρος του κόσμου. Η πρόσβαση στα αρχεία μπορεί να γίνει και μέσω των εικονικών μηχανών.



Εικόνα 7 Ο ωκεανός ως υπηρεσία τριών σύννεφων, ένα σύννεφο για πρόσβαση σε μηχανές μεγάλης υπολογιστικής ισχύος, ένα σύννεφο-υπηρεσία αποθηκευτικού χώρου, και ένα σύννεφο πρόσβασης σε εικονικά δίκτυα.

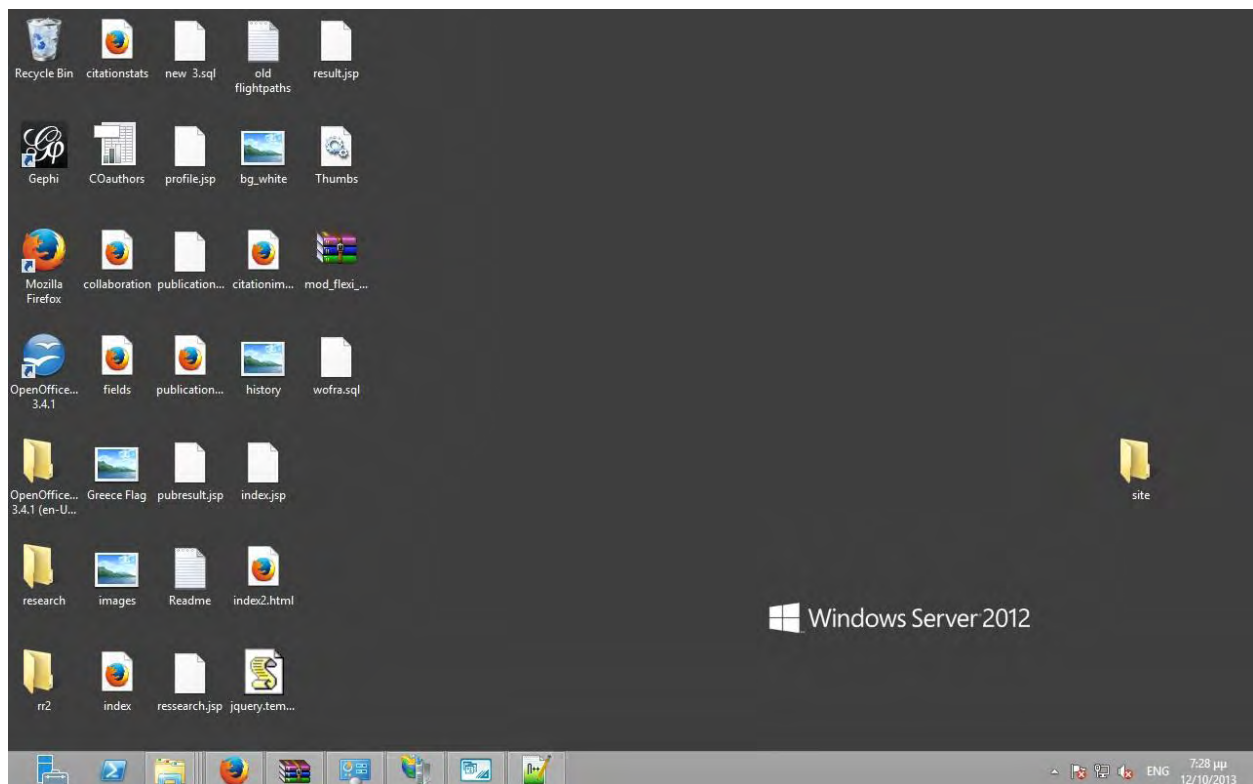
Στον Ωκεανό στήσαμε σε μία εικονική μηχανή έναν Windows Server 2012, αποδίδοντας του ως πόρους 100 GB χωρητικότητας, 8 επεξεργαστές, 8 GB RAM. Κατόπιν εγκαταστήσαμε τα λογισμικά που χρησιμοποιούμε όπως ο Apache Tomcat και η Mysql και μεταφέραμε τα αρχεία των υλοποιήσεων. Στην συνέχεια ορίσαμε ποιες από τις εφαρμογές μας και με ποιο τρόπο θα είναι προσβάσιμες στο Web. Η εγκατάσταση των λογισμικών και η ρύθμιση του εξυπηρετητή έγινε μέσω απομακρυσμένης σύνδεσης στην επιφάνεια εργασίας της εικονικής μας μηχανής σαν να την είχαμε μπροστά μας, μπορούμε να χρησιμοποιήσουμε το παραθυρικό περιβάλλον της διεπαφής των Windows 8. Εγκαταστήσαμε επίσης επιπλέον προγράμματα όπως το Net Beans και το Gephi για απευθείας επεξεργασία των αρχείων μας μέσω απομακρυσμένης σύνδεσης, από οποιονδήποτε σταθμό εργασίας και αν βρισκόμαστε. Αν και με τον παραπάνω τρόπο δεν χρειάζεται να δουλεύουμε αρχικά τοπικά και μετά να μεταφέρουμε τα αρχεία που άλλαξαν στον εξυπηρετητή, εγκαταστήσαμε και έναν ftp server, ως εναλλακτικό τρόπο μεταφοράς των αρχείων.

Οι εφαρμογές μας μπορούν να εντοπιστούν στον Ωκεανό από τα παρακάτω URLs:

Απογραφή: <http://83.212.108.14/apograhf>

Παρατηρητήριο Ερευνητικών Δραστηριοτήτων: <http://83.212.108.14/research/>

Δίκτυο Συνεργασιών: <http://83.212.108.14/infcoauthors>



Εικόνα 8 Η επιφάνεια εργασίας της εικονικής μας μηχανής

4.6 Επίδειξη

Όπως είπαμε και παραπάνω, ένας χρήστης της εφαρμογής μας έχει την δυνατότητα να αναζητήσει πληροφορία που αφορά κάποιον ερευνητή είτε να αναζητήσει δημοσιεύσεις που πληρούν κάποια συγκεκριμένα χαρακτηριστικά σύμφωνα με τα δικά του κριτήρια.

Στην πρώτη περίπτωση ο χρήστης μπορεί να επιλέξει από το μενού "Researcher Search". Από εκεί μπορεί είτε να επιλέξει από την λίστα στα αριστερά κάποιον ερευνητή είτε να αναζητήσει κάποιον ερευνητή συμπληρώνοντας μέρος του ονόματος του ή αν δεν το γνωρίζει να επιλέξει κάποιο τμήμα ή κάποιο επιστημονικό ενδιαφέρον. Αφού ο χρήστης πραγματοποιήσει την αναζήτηση θα λάβει μια λίστα με τα ονόματα των ερευνητών που ικανοποίησαν τα κριτήρια της εν λόγω αναζήτησης. Από αυτήν την λίστα μπορεί να επιλέξει κάποιο όνομα και να πλοηγηθεί έτσι στην σελίδα που περιέχει το προφίλ του επιλεγθέντος ερευνητή.

Home Researcher Search Publication Search

Select researcher from the list

Search Researcher

▶ Akritas, A.G.

▶ Antonopoulos, C.D.

▶ Argyriou, A.

▶ Bellas, N.

▶ Bozanis, P.

▶ Daskalopulu, A.

▶ Evmorfopoulos, N.

▶ Hadjidakis, A.

Department: Τ.Μ.Η.Υ.Τ.Α

Faculty

Researcher Name: Name

Research Interest:

Reset Values Search Researchers

4-9 Αναζήτηση ερευνητή

Στο προφίλ του ερευνητή μπορούμε να δούμε τον αριθμό των δημοσιεύσεων του, τον συνολικό αριθμό αναφορών, το h-Index, έναν χάρτη που απεικονίζει συνέδρια στα οποία αυτός συμμετείχε, τους συνεργάτες του, τα επιστημονικά πεδία στα οποία δραστηριοποιείται, τις λέξεις κλειδιά που χρησιμοποιεί περισσότερο και τέλος έναν πίνακα με όλες τις δημοσιεύσεις του.

Παρατηρώντας τον πίνακα βλέπουμε ότι είναι σελιδοποιημένος με 10 δημοσιεύσεις ανά σελίδα. Ο χρήστης μπορεί να αλλάξει τον αριθμό των δημοσιεύσεων ανά σελίδα. Μπορεί επίσης να ταξινομήσει τις δημοσιεύσεις επιλέγοντας μία στήλη, αφού κάνει κλικ πάνω στην επικεφαλίδα της στήλης. Στην περίπτωση αυτή η ταξινόμηση θα γίνει ως προς την στήλη που επέλεξε ο χρήστης. Κάνοντας κλικ πάνω σε μία δημοσίευση οδηγείται στην αντίστοιχη σελίδα της δημοσίευσης στο Scopus.

Στην περίπτωση που ο χρήστης επιλέξει από το μενού "Publication Search" θα οδηγηθεί σε μια φόρμα αναζήτησης δημοσιεύσεων. Κατόπιν μπορεί να συμπληρώσει κάποια από τα πεδία της φόρμας

γράφοντας κάποια λέξη και επιλέγοντας τι είδους αναζήτηση θέλουμε να κάνουμε για την λέξη αυτή. Με άλλα λόγια, ο χρήστης δηλώνει ότι θέλει το σύστημά μας να εντοπίσει τις δημοσιεύσεις εκείνες οι οποίες περιέχουν στον τίτλο, στην περιληψη ή στις λέξεις κλειδιά την λέξη που αυτός όρισε. Μπορεί επίσης να ορίσει ένα όνομα ερευνητή ή ένα επιστημονικό πεδίο. Στην φόρμα μας έχουμε 3 διαθέσιμα πεδία αναζήτησης και ο χρήστης πρέπει να συμπληρώσει τουλάχιστον ένα για να λάβει κάποια αποτελέσματα. Προαιρετικά μπορεί να συμπληρώσει και τα άλλα δύο πεδία, ορίζοντας παράλληλα αν οι επιμέρους επερωτήσεις σχετίζονται με την σχέση του λογικού «ΚΑΙ» ή το λογικού «Η». Τέλος, μπορεί να φιλτράρει εκ των προτέρων τα αποτελέσματα περιορίζοντας την αναζήτηση σε κάποιο χρονολογικό διάστημα (Limited Search by Year), σε ένα συγκεκριμένο επιστημονικό πεδίο (Limited Search by Type) ή σε ένα κάτω όριο αριθμού αναφορών.

Bozanis, P.

Documents Num.: 28
Citations Num.: 78
H index: 5

Conferences

Co Authors

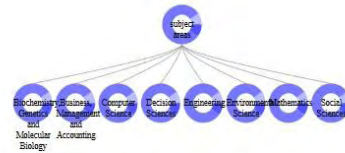
Bibi, S. (3) Gkoulalas-Divanis, A. (1) **Akritidis, L. (10)** Martini, A. (1)
Zaharis, A. (1) Foteinos, P. (1) **Fevgas, A. (3)** Fraggogiannis, N. (1) Drasidis, G. (1) Tsiouvolos, A. (1)
(1) Tsakiridis, F. (1) Vasaitis, V. (1) Verykiotis, V.S. (1) **Kitsios, N. (3)** **Nanopoulos, A. (3)**
Tsakalidis, A. (4) **Makris, C. (4)** Manolopoulos, Y. (2) **Bozanis, P. (28)** Houstis, E.N. (1) **Katsaros, D. (11)** Tounidou, E. (1)
Tsompanopoulou, P. (3)

4-10 Σελίδα προφίλ ερευνητή.

Για παράδειγμα, όπως βλέπουμε και στην εικόνα 4-10, μπορούμε να αναζητήσουμε τις δημοσιεύσεις εκείνες που έχουν είτε στον τίτλο τους, είτε στις λέξεις κλειδιά και την περιληψη, την λέξη "data", δημοσιεύθηκαν από το 2000 έως το 2013 και έχουν πάνω από 10 αναφορές.

Μόλις ο χρήστης πατήσει το κουμπί με την ένδειξη "Search Publications" το σύστημα μας εκτελεί τις επερωτήσεις στην βάση δεδομένων και εμφανίζει στον χρήστη έναν πίνακα με τα αποτελέσματα της αναζήτησης, δηλαδή την λίστα των δημοσιεύσεων που πληρούσαν τα κριτήρια του χρήστη. Στο

Research Interests / Co-investigators



Keywords

Dominance Knowledge Discovery Partial rebuilding Persistence Rectangle enclosure Security Security & privacy Spatial DBMS Spatial data Trajectory K-anonymity Computational geometry Indexing

Publications

select paper.id,title,citedby,authors,year,AuthorKeywords,link,author_paper.authorid from wofoa.paper inner join wofoa.author_paper on paper.id = author_paper.paperid where wofoa.author_paper.AuthorID = 5

Publication Title	Cited By	Authors	Year	Keywords
Computing scientometrics in large-scale academic search engines with InspecRelevance	0	Akritidis, L., Bozanis, P.	2012	null
Positional data organization and compression in web-oriented indexes	0	Akritidis, L., Bozanis, P.	2012	null
A network aware attack model for online requests in trajectory data	6	Okoulas-Ouafis, A., Verykiotis, V.S., Bozanis, P.	2009	Indexing, Knowledge Discovery, Security & privacy, Spatial DBMS, Trajectory, K-anonymity
WebFrees	2	Bozanis, P., Fotinos, P.	2007	Indexing, Partial rebuilding, Spatial data
Indexing Frees	1	Vasaitis, V., Nanopoulos, A., Bozanis, P.	2004	null
A new approximate data structure for the 1-1.2-dimensional range query problem	0	Bozanis, P.	2003	null
Categorical range queries in large databases	2	Nanopoulos, A., Bozanis, P.	2003	null

παραπάνω παράδειγμα η επερώτηση που στην ουσία δημιούργησε ο ίδιος ο χρήστης μέσω της συμπλήρωσης της φόρμας, σε SQL θα έχει ως εξής:

```
SELECT id,title,citedby,authors,year,AuthorKeywords,sourcetitle,link,authorid
FROM paper
WHERE (AuthorKeywords LIKE '%data%' AND abstract LIKE '%data%' OR title LIKE '%data%') AND year >=
'2000' AND year <= '2013' AND citedBy >= '10'
```

Ο πίνακας με τα αποτελέσματα της αναζήτησης δημοσιεύσεων έχει όμοια μορφή και λειτουργικότητα με αυτόν που είδαμε προηγουμένως στο προφίλ του ερευνητή. Εκτός από τον πίνακα αυτόν εμφανίζεται και η λίστα των κορυφαίων ερευνητών στα συγκεκριμένα αποτελέσματα αναζήτησης, δηλαδή ποιοι ερευνητές εμφανίζονται συχνότερα στις δημοσιεύσεις που το σύστημά μας επέστρεψε ως αποτελέσματα της αναζήτησης.

Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών & Δικτύων

Home Researcher Search Publication Search

Publication Search

Search

AND

OR

Limited search by Year: From To

Limited search by

Limited Search Show results that are cited by more than: times

4-11 Παράδειγμα επερώτησης μέσω συμπλήρωσης της φόρμας αναζήτησης δημοσίευσης.

Publication Title	Cited By	Authors	Year	Keywords
PYTHIA-II: A Knowledge/Database System for Managing Performance Data and Recommending Scientific Software	32	Houstis, E.N., Catlin, A.C., Rice, J., Verykios, V.S., Ramakrishnan, N., Houstis, C.,	2000	Data mining; Inductive logic programming; Knowledge discovery in databases; Knowledge-based systems; Performance evaluation; Recommender systems; Scientific software
Prefetching in content distribution networks via Web communities identification and outsourcing	22	Sidiropoulos, A., Pallas, G., Katsaros, D., Stamos, K., Vakali, A., Manolopoulos, Y.,	2008	Content distribution networks; Data dissemination techniques on the web; Internet and web-based; Web communities; Web data mining; Web prefetching
A data mining approach for location prediction in mobile environments	68	Yavas, G., Katsaros, D., Ulusoy, O., Manolopoulos, Y.,	2005	Data mining; Location prediction; Mobile computing; Mobility patterns; Mobility prediction
Broadcast program generation for Webcasting	11	Katsaros, D., Manolopoulos, Y.,	2004	Broadcast disks; Mobile computing; Push-based delivery; Scheduling; Webcasting; Wireless data dissemination
A data mining algorithm for generalized web prefetching	63	Nanopoulos, A., Katsaros, D., Manolopoulos, Y.,	2003	Association rules; Data mining; Prediction; Prefetching; Web mining
Search for gravitational waves from compact binary coalescence in LIGO and Virgo data from S5 and VSR1	41	Abadie, J., Abbott, B.P., Abbott, R., Abernathy, M., Accadia, T., Acernese, F., Adams, C., Zwegzig, J.,	2010	Physical Review D - Particles, Fields, Gravitation and Cosmology
Search for gravitational-wave bursts associated with gamma-ray bursts using data from LIGO science run 5 and Virgo science run 1	33	Abbott, B.P., Abbott, R., Acernese, F., Adhikari, R., Ajith, P., Allen, B., Allen, G., Zwegzig, A.J.,	2010	Astrophysical Journal Letters
Searching for prompt signatures of nearby core-collapse supernovae by a joint analysis of neutrino and gravitational wave data	10	Leonor, I., Cadonati, L., Cocchia, E., D'Antonio, S., Credico, A.D., Fafone, V., Frey, R., Vissani, F.,	2010	Classical and Quantum Gravity
Searches for gravitational waves from known pulsars with science run 5 ligo data	55	Abbott, B.P., Abbott, R., Acernese, F., Adhikari, R., Ajith, P., Allen, B., Allen, G., Stappers, B.,	2010	Astrophysical Journal Letters
Search for gravitational wave ringdowns from perturbed black holes in LIGO S4 data	36	Abbott, B.P., Abbott, R., Adhikari, R., Ajith, P., Allen, B., Allen, G., Amin, R., Zwegzig, J.,	2009	Physical Review D - Particles, Fields, Gravitation and Cosmology

1/3 10

4-12 Ο πίνακας με τα αποτελέσματα της αναζήτησης που πραγματοποιήσαμε πριν ως παράδειγμα.

Κεφάλαιο 5

5 Ανάλυση των οπτικοποιήσεων

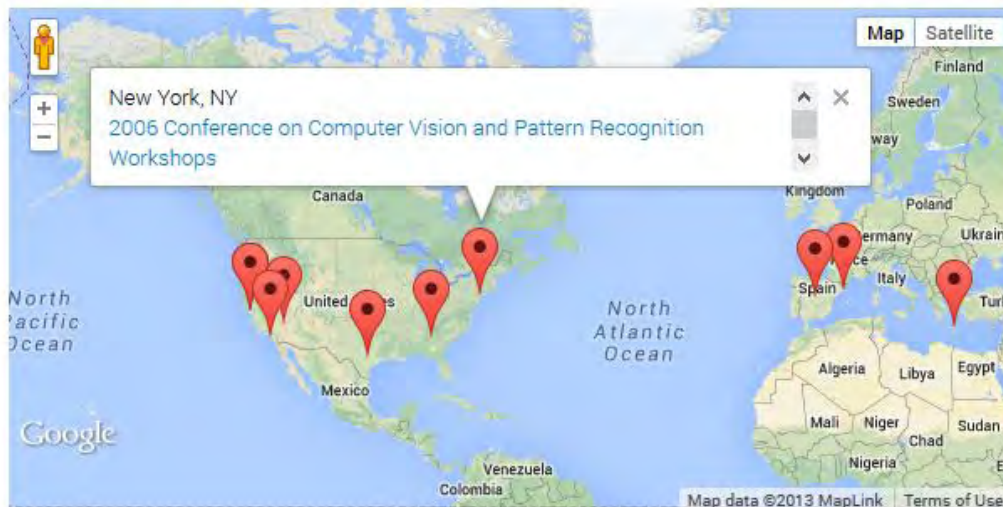
Στα προηγούμενο κεφάλαιο εξηγήσαμε τον τρόπο με τον οποίο υλοποιήσαμε το παρατηρητήριο των ερευνητικών δραστηριοτήτων των μελών ΔΕΠ του ΤΗΜΜΥ. Στην εφαρμογή χρησιμοποιήσαμε αρκετές οπτικοποιήσεις δεδομένων για να εξυπηρετήσουμε τον σκοπό μας που είναι η άντληση γνώσης για την ερευνητική δραστηριότητα του τμήματος. Παρακάτω θα αναλύσουμε χωριστά την κάθε οπτικοποίηση που υλοποιήσαμε.

5.1 Χάρτης συνεδριών

Για να απεικονίσουμε τα συνέδρια στα οποία έχει λάβει μέρος κάποιο μέλος ως εισηγητής χρησιμοποιήσαμε τον χάρτη της Google. Τα συνέδρια απεικονίζονται ως σημεία στον χάρτη και όταν ο χρήστης κάνει κλικ σε κάποιο από αυτά μπορεί να δει το όνομα του συνεδρίου και την χρονολογία. Μπορεί επίσης να κάνει κλικ στον τίτλο του συνεδρίου, ο οποίος είναι ταυτόχρονα ένας υπερσύνδεσμος με την αντίστοιχη σελίδα του Scopus με περισσότερες λεπτομέρειες για το συγκεκριμένο συνέδριο.

Για να αντιστοιχίσουμε τα δεδομένα μας που είναι ονόματα των πόλεων στις οποίες διεξήχθησαν τα συνέδρια με τα γεωγραφικά μήκη και πλάτη τους χρησιμοποιήσαμε την υπηρεσία Geolocation του Google Map API v3.

Conferences



Εικόνα 5-1 Δισδιάστατη οπτικοποίηση για την απεικόνιση των συνεδριών του Αναπληρωτή Καθηγητή του Τ.Η.Μ.Μ.Υ. κ. Μπέλλα. Παρατηρούμε ότι τα περισσότερα συνέδρια στα οποία συμμετείχε έχουν διεξαχθεί στις Η.Π.Α.

Το πλεονέκτημα της οπτικοποίησης είναι ότι κάποιος μπορεί να δει με μια γρήγορη ματιά τις περιοχές στις οποίες έχει δραστηριοποιηθεί το κάθε μέλος, όσον αφορά την διεξαγωγή συνεδρίων.

5.2 Σύννεφο ετικετών (Tag Cloud) για τις λέξεις κλειδιά

Rank aggregation Rank fusion **Ranking** Reducing length representation **Replacement policy**
Replication Requirements Research area Research field SEM SER SaaS Scheduling Scientist
Scientometrics **Search Engine** Search engines Semistructured data **Sensor networks**
Services Simulation Skewed access SoC Social network analysis **Software** Software cost Software
development TCO Temporal locality **Topology Control** Trace-driven simulation Tree mining Tuning
time Utility computing **Web** Web communities Web data mining Web mining Web prefetching
Webcasting Wireless ad-hoc network. Wireless data dissemination **Wireless network** Wireless
networks **Wireless sensor network** World Wide Web backpressure
distributed algorithms internet software as a service software cost metrics throughput total cost of
ownership **wireless wireless sensor**
networks Architecture Association rules Attractive Backbone formation Betweenness
centrality Blogosphere Borda Count Broadcast disks Cache coherency **Caching** Centrality
metrics Clustering Content distribution networks Content-based similarity Cooperative caching Data
coverage Data mining Energy conservation Evaluation ad hoc networks **cloud computing**

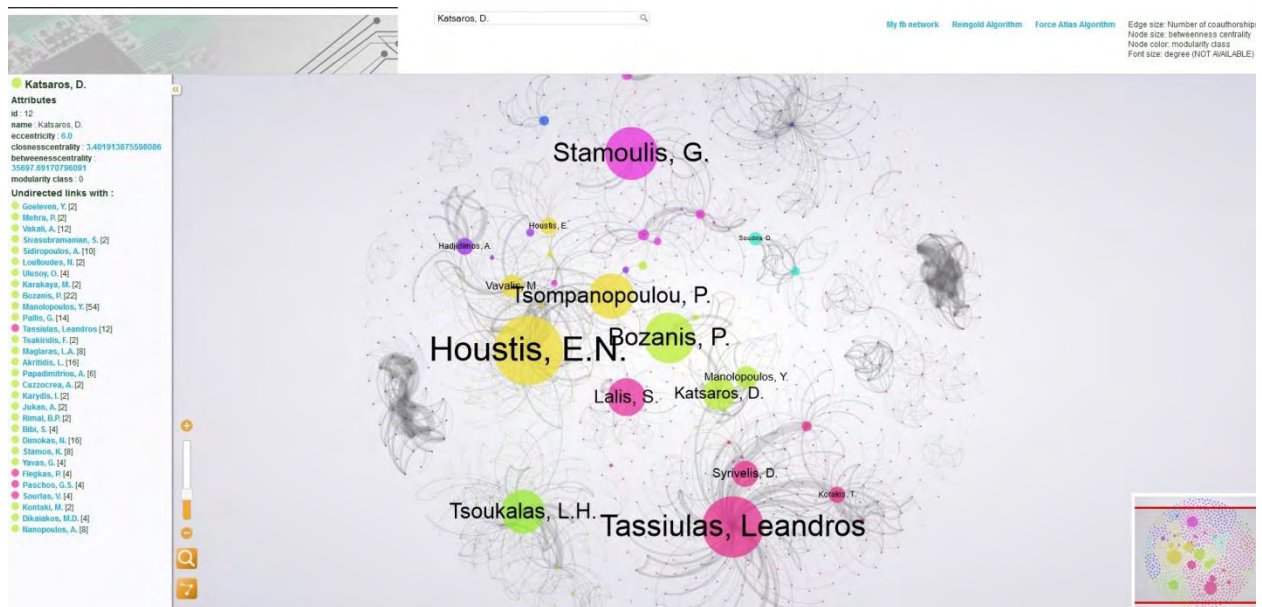
Εικόνα 5-2 Παρατηρώντας το σύννεφο ετικετών του κ. Κατσαρού βλέπουμε ότι οι περιοχές που δραστηριοποιείται περισσότερο είναι, μεταξύ άλλων, τα ασύρματα δίκτυα και το Web.

Μία καλή μέθοδος για να κατανοήσουμε τα θεματικά αντικείμενα τα οποία πραγματεύεται το κάθε μέλος είναι να δούμε ποιες λέξεις κλειδιά χρησιμοποιεί συχνότερα. Για να απεικονίσουμε τις λέξεις κλειδιά χρησιμοποιήσαμε ένα σύννεφο ετικετών, το οποίο αποτελεί μία μονοδιάστατη οπτικοποίηση με μία μεταβλητή, το μέγεθος της γραμματοσειράς της κάθε λέξης, το οποίο είναι ανάλογο με την συχνότητα εμφάνισης της λέξεως-κλειδί. Χρησιμοποιήσαμε 10 επίπεδα γραμματοσειρών. Το εύρος του κάθε επιπέδου προσαρμόζεται σε κάθε ερευνητή ανάλογα με τον αριθμό εμφανίσεων της πιο συχνά χρησιμοποιούμενης λέξης, και κωδικοποιείται ανάλογα με κάποιο μέγεθος γραμματοσειράς.

Το σύννεφο ετικετών υλοποιήθηκε με JSP.

5.3 Δίκτυο συνεργασιών

Με χρήση οπτικοποιήσεων δεδομένων δικτύου προσπαθήσαμε να απεικονίσουμε, το δίκτυο των συνεργασιών των μελών ΔΕΠ του τμήματος με άλλους ερευνητές. Η οπτικοποίηση του δικτύου συνεργασιών μας δίνει μια ξεκάθαρη εικόνα για το ποιες συνεργασίες αναπτύσσονται μεταξύ των ερευνητών του T.H.M.M.Y. με άλλους επιστήμονες που εργάζονται σε άλλα ιδρύματα.



5-3 Το δίκτυο των συνεργασιών των μελών ΔΕΠ. Οι κόμβοι εδώ έχουν διαταχθεί με τον αλγόριθμο Reingold.

Για να υλοποιήσουμε την οπτικοποίηση βασιστήκαμε στην υπόθεση ότι μια συνεργασία μεταξύ δύο ερευνητών προκύπτει αν αυτοί έχουν συντάξει από κοινού κάποιο επιστημονικό άρθρο. Στον γράφο μας, κάθε κόμβος συμβολίζει έναν ερευνητή και κάθε ακμή μια σύνδεση συνεργασίας.

Χρησιμοποιούμε τέσσερις μεταβλητές: το μέγεθος του κόμβου κωδικοποιεί την διαμεσότητα (betweenness centrality), το μέγεθος της ακμής τον αριθμό των συνεργασιών, το μέγεθος της γραμματοσειράς τον βαθμό απλής συνδεσιμότητας (Degree) του κόμβου και το χρώμα του κόμβου την κοινωνική ομάδα στην οποία ανήκει το κάθε μέλος. Άτομα που ανήκουν στην ίδια ομάδα εκτός από το ότι έχουν το ίδιο χρώμα βρίσκονται και σε κοντινή απόσταση μεταξύ τους.

Ο χρήστης μπορεί να επιλέξει με το ποντίκι ή πληκτρολογώντας στο πεδίο αναζήτησης (το οποίο είναι αυτόματης συμπλήρωσης), κάποιο μέλος. Όταν το κάνει αυτό τονίζεται περισσότερο το δικό του δίκτυο συνεργατών ενώ δίνεται και η δυνατότητα να εξαφανιστούν τελείως όσοι κόμβοι δεν συνδέονται με το επιλεγθέν μέλος. Επίσης σε μία στήλη δίπλα στον γράφο εμφανίζονται κάποιες πληροφορίες προφίλ του επιλεγθέντος μέλους. Τέλος, ο χρήστης μπορεί να κάνει ζουμ σε οποιαδήποτε περιοχή του γράφου θέλει και να κάνει εναλλαγή αν ο γράφος του θέλει να είναι σχεδιασμένος με τον αλγόριθμο των Fruchterman - Reingold (Fruchterman, T. M. J., & Reingold, E. M., 1991) ή με τον αλγόριθμο Force Atlas.

Με το Gephi υπολογίζουμε το μέσο "clustering coefficient" του δικτύου και βλέπουμε ότι ισούται με 0,881. Από αυτό αλλά και παρατηρώντας τον γράφο συμπεραίνουμε ότι το δίκτυο μας είναι δίκτυο μικρού κόσμου και έχει δομή όμοια με αυτή που θα είχε ένα κοινωνικό δίκτυο, δεν είναι τυχαία, αλλά συσταδοποιημένη (clustered). Αυτό φαίνεται πρακτικά από το ότι οι συνεργάτες κάποιου ερευνητή τείνουν να έχουν τους ίδιους συνεργάτες. Από την θεωρία γνωρίζουμε ότι δίκτυα με το μεγαλύτερο δυνατό μέσο «clustering coefficient» εμφανίζουν δομή ενότητων (modular). Αυτό το διαπιστώνουμε και πειραματικά με την εφαρμογή του αλγορίθμου "Modularity" του Gephi. Στο overview του Gephi οι συστάδες είναι ορατές και κωδικοποιούνται με ανάλογο χρωματισμό των κόμβων που ανήκουν στην ίδια συστάδα. Στο δίκτυο μας οι συστάδες κωδικοποιούν τις ομάδες συνεργασιών που αναπτύσσονται κυρίως εντός του Τ.Η.Μ.Μ.Υ.

Μελετώντας την οπτικοποίηση, εκτός από το δίκτυο συνεργατών του κάθε μέλους ΔΕΠ, μπορούμε να δούμε ελέγχοντας το μέγεθος του κόμβου ποιοι ερευνητές εμφανίζουν μεγάλη διαμεσότητα. Όπως είδαμε και στο κεφάλαιο 2 άτομα με υψηλή διαμεσότητα είναι κρίσιμα για συνεργασίες μεταξύ διαφορετικών κοινωνικών υποομάδων λειτουργώντας ως διαμεσολαβητές πληροφορίας και συνεργασίας.

Για παράδειγμα, στην υλοποίηση μας βλέπουμε ότι οι κύριοι Χούστης, Τσουκαλάς, Σταμούλης και Τασιούλας εμφανίζουν την μεγαλύτερη διαμεσότητα στον γράφο μας. Αν διαγραφεί κάποιος από αυτούς τους κόμβους από το δίκτυο μας, τότε θα έχουμε πρόβλημα στην ανάπτυξη των συνεργασιών, μιας και οι κόμβοι αυτοί στην ουσία είναι οι περισσότερο υπεύθυνοι για τις συνεργασίες που αναπτύσσονται μεταξύ διαφορετικών κοινωνικών υποομάδων που βρίσκονται μέσα στο δίκτυό μας.

Ένα άλλο παράδειγμα είναι αυτό που εμφανίζεται στην εικόνα 5-4. Εδώ έχουμε απομονώσει το δίκτυο συνεργατών του κύριου Μποζάνη. Βλέπουμε ότι εκτός από τους πράσινους κόμβους που αντιπροσωπεύουν ερευνητές που ανήκουν στην ίδια ομάδα συνεργασιών, ο κύριος Μποζάνης έχει αναπτύξει και συνεργασία με ερευνητές που ανήκουν σε μία άλλη ομάδα συνεργασιών και μάλιστα χαρακτηρίζονται από υψηλή διαμεσότητα. Αυτό προσδίδει και στον κύριο Μποζάνη υψηλή διαμεσότητα, αφού στην ουσία λειτουργεί ως απαραίτητος συνδετικός κρίκος μεταξύ των δύο υποομάδων.



Εικόνα 5-4 Το δίκτυο συνεργασιών του κυρίου Μποζάνη

Ο γράφος μας σχεδιάστηκε με το εργαλείο Gephi στο οποίο εισαγάγαμε τα δεδομένα από την βάση μας. Κατόπιν με χρήση Javascript κάναμε εξαγωγή του γράφου ώστε να είναι προσβάσιμος μέσω Ιστού και προσθέσαμε τα διαδραστικά στοιχεία.

Στο μέλλον, φιλοδοξούμε να εντοπίσουμε χωρικά δεδομένα για τον τόπο που δραστηριοποιούνται οι ερευνητές και να αποτυπώσουμε το δίκτυο μας πάνω στον χάρτη. Στην περίπτωση αυτή οι κόμβοι δεν θα διαταχθούν από κάποιον αλγόριθμο κατευθυνόμενο με δυνάμεις, αλλά θα τοποθετηθούν στο

γεωγραφικό μήκος και πλάτος που τους αντιστοιχεί. Με τον τρόπο αυτό θα μπορέσουμε να εξερευνήσουμε πόσο εξωστρεφή είναι τα μέλη ΔΕΠ του τμήματός μας, αλλά και κατά πόσο εμφανίζουν διασυνδέσεις με ερευνητές από διάφορα μέρη του πλανήτη. Επίσης σκοπός μας για το μέλλον είναι να εισάγουμε μία επιπλέον διάσταση που θα αντιπροσωπεύει το επιστημονικό αντικείμενο πάνω στο οποίο αναπτύχθηκαν οι συνεργασίες. Για να το κάνουμε αυτό μπορούμε για παράδειγμα να υποθέσουμε την κατηγοριοποίηση που είδαμε στην ενότητα 4-1-4. Η διάσταση αυτή μπορεί να κωδικοποιηθεί με χρωματισμό των ακμών του γράφου ανάλογα με το επιστημονικό πεδίο. Επειδή όμως μια ακμή μπορεί να αντιστοιχεί σε περισσότερα του ενός είδη συνεργασιών, μπορούμε να λάβουμε υπ' όψιν μόνο το είδος με την μεγαλύτερη συχνότητα εμφανίσεων.

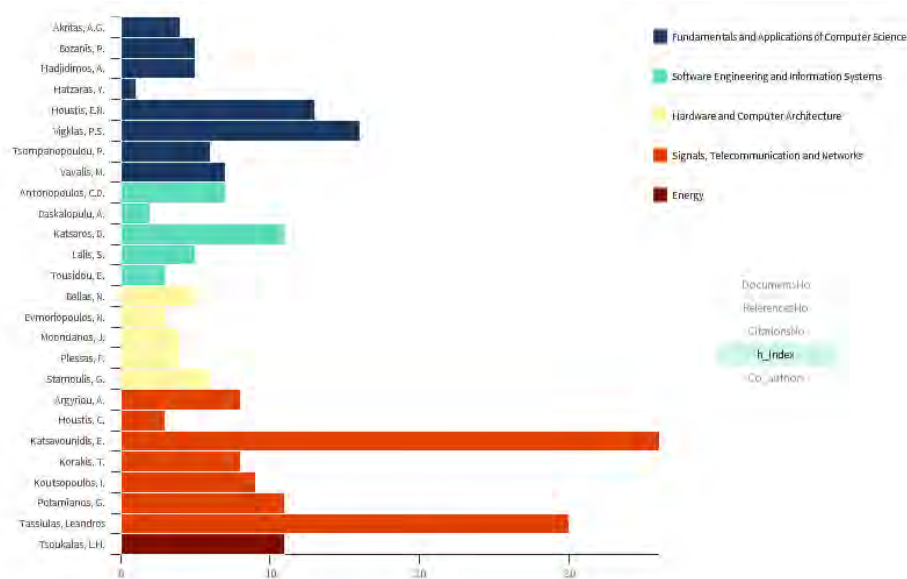
5.4 Διάγραμμα ράβδων για αναπαράσταση των βιβλιομετρικών δεικτών

Στο διάγραμμα αυτό αναπαριστούμε κάθε φορά τρεις μεταβλητές, μια για την κάθε οντότητα ερευνητή μία ποσοτική που κάθε φορά μπορεί να είναι η τιμή κάποιου βιβλιομετρικού δείκτη και μια κατηγορική μεταβλητή που κωδικοποιεί σε ποιο τμήμα ανήκει ο κάθε ερευνητής. Η πρώτη μεταβλητή αναπαρίσταται στον κάθετο άξονα η δεύτερη στον οριζόντιο και η κατηγορική με το χρώμα της κάθε ράβδου.

Το γράφημα είναι διαδραστικό με τον χρήστη όσον αφορά την επιλογή του βιβλιομετρικού δείκτη. Για παράδειγμα στην παραπάνω εικόνα βλέπουμε την ένταση του h index για τον κάθε ερευνητή. Επιπλέον, μπορεί ο χρήστης να επιλέξει αναπαράσταση για τον αριθμό των δημοσιεύσεων, των αριθμό των αναφορών κ.ο.κ. Αυτομάτως, κάθε φορά που αλλάζει ο δείκτης έχουμε προσαρμογή του οριζοντίου άξονα σε μια κλίμακα ορίζεται από το μηδέν και την μέγιστη τιμή για τον αντίστοιχο δείκτη.

Μελετώντας το γράφημα, εκτός του ότι βλέπουμε άμεσα και ξεκάθαρα ποιοι ερευνητές, όσον αφορά τους παραπάνω δείκτες, συγκεντρώνουν μεγαλύτερες επιδόσεις, μπορούμε να βγάλουμε και πιο γενικά συμπεράσματα για τις επιδόσεις των Τμημάτων. Παρατηρώντας το διάγραμμα διαπιστώνουμε για παράδειγμα ότι οι ερευνητές του Τμήματος "Σήματα, Τηλεπικοινωνίες και Δίκτυα" λαμβάνουν υψηλότερες επιδόσεις.

Το διάγραμμα υλοποιήθηκε με το Quadrigram, μέσω του οποίου και συγκεκριμένα ενός "database connector" που δημιουργήσαμε κάνουμε επερώτηση στην βάση δεδομένων μας για να πάρουμε έναν πίνακα με πεδία το ονόμα, το id του Τμήματος και τις τιμές του κάθε δείκτη. Με βάση τον πίνακα αυτό και χρήση των κατάλληλων modules του Quadrigram, υλοποιήσαμε την οπτικοποίησή μας.

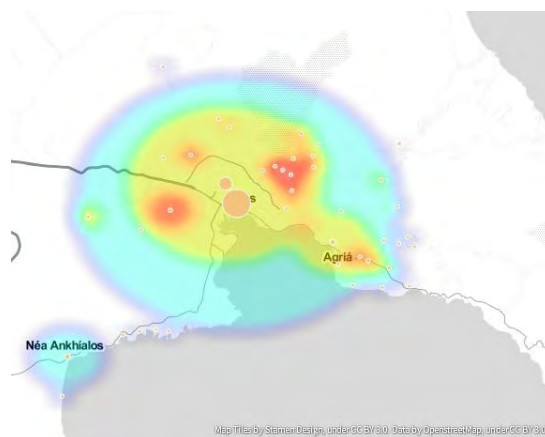


Εικόνα 5-5 Κατάταξη με βάση το H-Index

5.5 Heat Map για τον πληθυσμό του Δήμου Βόλου

Στο κεφάλαιο δύο είδαμε τι είναι ένας Heat Map. Χρησιμοποιώντας το Quadrigram υλοποιήσαμε μία εναλλακτική τεχνική αποτύπωσης της απογραφής πληθυσμού με χρήση Heat Map. Βασιζόμενοι στους πληθυσμούς των οικισμών και κωδικοποιώντας τους ως αναπτύσσουμε μία διερευνητική οπτικοποίηση.

Στον χάρτη βλέπουμε τις περιοχές οι οποίες περιέχουν οικισμούς με υψηλό πληθυσμό να εμφανίζονται με θερμότερα χρώματα, ενώ τις περιοχές που δεν περιέχουν οικισμούς με μεγάλο πληθυσμό να εμφανίζονται με ψυχρότερα χρώματα. Ανάλογα αν θέλουμε να επικεντρωθούμε σε συγκεκριμένα χωρικά πεδία ή να αποκτήσουμε γενικότερη αντίληψη για τον πληθυσμό στον Δήμο Βόλου μπορούμε να επιλέξουμε το επιθυμητό επίπεδο εστίασης.



5-6Χάρτης θερμότητας για τον πληθυσμό του Δήμου Βόλου.

Κεφάλαιο 6

6 Σύνοψη

Αντικείμενο της παρούσας διπλωματικής ήταν η οπτικοποίηση των δεδομένων, η οποία κρίνεται απαραίτητη ώστε να μπορούμε να αντλούμε νόημα και γνώση μέσα από τον καταιγισμό των δεδομένων που βιώνουμε στις μέρες μας. Και επειδή όπως αναφέραμε, τα δεδομένα είναι καταγραφές στιγμιότυπων του «έξω» κόσμου, η γνώση που αποκτούμε θα μας δώσει την δυνατότητα να εξάγουμε χρήσιμα συμπεράσματα, να επιδιώξουμε έναν καλύτερο κόσμο, να βελτιώσουμε το επίπεδο ζωής και να λαμβάνουμε τις σωστές αποφάσεις.

Με την χρήση των κατάλληλων κάθε φορά εργαλείων προσπαθήσαμε να εφαρμόσουμε στην πράξη τα όσα είδαμε στην θεωρία, υλοποιώντας μία απεικόνιση απογραφής πληθυσμού και ένα παρατηρητήριο για τις ερευνητικές δραστηριότητες των μελών ΔΕΠ του Τ.Η.Μ.Μ.Υ. Τα αποτελέσματα ήταν ικανοποιητικά καθώς κάθε οπτικοποίηση που αναπτύξαμε μας οδήγησε σε πολύτιμα συμπεράσματα.

Όσον αφορά την εφαρμογή του παρατηρητηρίου, εφόσον βασιζόμαστε σε μία βάση δεδομένων η οποία ανανεώνεται ανά τακτά χρονικά διαστήματα, και επειδή οι τεχνολογίες που χρησιμοποιούμε υποστηρίζουν υλοποίηση συστημάτων πραγματικού χρόνου, με την έννοια ότι οι επερωτήσεις προς την βάση δεδομένων γίνονται επαναλαμβανόμενα ανά κάποιο χρονικό διάστημα (λίγο μεγαλύτερο της περιόδου ανανέωσης της βάσης δεδομένων), οι οπτικοποιήσεις μας θα είναι δυναμικές και θα ενημερώνονται κάθε φορά που συμβαίνει κάποιο γεγονός, δηλαδή κάποια αλλαγή στα δεδομένα μας.

Στόχος μας για το μέλλον είναι να μπορούμε να ενημερώνουμε τον χρήστη σε πραγματικό χρόνο, εμφανίζοντας του ειδοποιήσεις κάθε φορά που συμβαίνει κάποια αλλαγή σε κάτι που τον ενδιαφέρει. Στην περίπτωση αυτή η οπτικοποίησή μας θα είναι άμεσα διαδραστική και θα εξελίσσεται δυναμικά ως ροή γεγονότων που συμβαίνουν επί των δεδομένων. Για να είμαστε σε θέση να υλοποιήσουμε ένα τέτοιο σύστημα πραγματικού χρόνου θα πρέπει μέσω διεπαφών, να έχουμε συνεχή πρόσβαση σε πολλαπλές μη ομοιογενείς πηγές δεδομένων και να εισάγουμε τα δεδομένα αυτά σε πραγματικό χρόνο στο σύστημα μας.

Κεφάλαιο 7

7 Βιβλιογραφία

7.1 Βιβλία

Joel Laumans. *An introduction to visualizing data* .

Galitz, W. O. (2002). *The essential guide to user interface Design – An Introduction to GUI Design Principles and Techniques* .

Stephen Few. *Table and graph design for enlightening communication. Perceptual Edge*.

Chun-houh Chen, Wolfgang Hardle, Antony Unwin. *Handbook of data visualization*

Noab Iliinski & Julie Steele. *Designing data visualizations*

Keith Andrews. *Information visualization - course notes Ao.Univ*.

7.2 Ιστοσελίδες

1. Εθνικό Κέντρο Τεκμηρίωσης www.ekt.gr

2. Εθνικό Κέντρο Τεκμηρίωσης - Δείκτες Έρευνας και Καινοτομίας <http://metrics.ekt.gr/>

3. Webscience Trust - Ορισμός Web Observatory <http://webscience.org/a-definition-of-the-web-observatory/>

4. Κέντρο ΠΛΗ.ΝΕ.Τ. Ν. Φλώρινας - Εισαγωγή στην JSP <http://dide.flo.sch.gr/Plinet/Tutorials/Tutorials-JSP-1-Introductiion.html>

5. Webapptester - Αρχιτεκτονική MVC <http://webapptester.com/mvc-framework-first-impression/>

6. Wikipedia: data computing

7. Wikipedia: Force-directed graph drawing

8. Wikipedia: User Experience

9. Βικιπαίδεια: Δεδομένα

10. Βικιπαίδεια: [Γράφος](#)

11. D3 Data Driven Documents <http://d3js.org/>

12. Dev Opera, SVG or Canvas? Choosing between the two <http://dev.opera.com/articles/view/svg-or-canvas-choosing-between-the-two/>

13. Who is central to a social network? It depends on your centrality measure.
<http://www.activatenetworks.net/blog/who-is-central-to-a-social-network-it-depends-on-your-centrality-measure/>
14. Google Maps Javascript V3 <https://developers.google.com/maps/documentation/javascript/>
15. Wikipedia: Clustering Coefficient http://en.wikipedia.org/wiki/Clustering_coefficient
16. Okeanos: <https://okeanos.grnet.gr/home/>

7.3 Αναφορές

Anne-Wil Harzing. (2010). *The Publish or Perish Book*.

Ben Shneiderman-University of Maryland, C. P.-U. (2009). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*.

Christopher G. Healey, James. T. Enns. (2011). Attention and Visual Memory in Visualization and Computer Graphics.

Freeman, Linton. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph Drawing by Force-Directed Placement. *Software - Practice and Experience* .

Jennifer Tidwell. (2010). *Designing Interfaces*. O Reilly.

Leland Wilkinson. (2005). *The Grammar of Graphics*.

Tufte, Edward R. (2001). *The Visual Display of Quantitative Information*.

Wikipedia. (n.d.). *Wikipedia*. Retrieved from <http://en.wikipedia.org/wiki/PageRank>