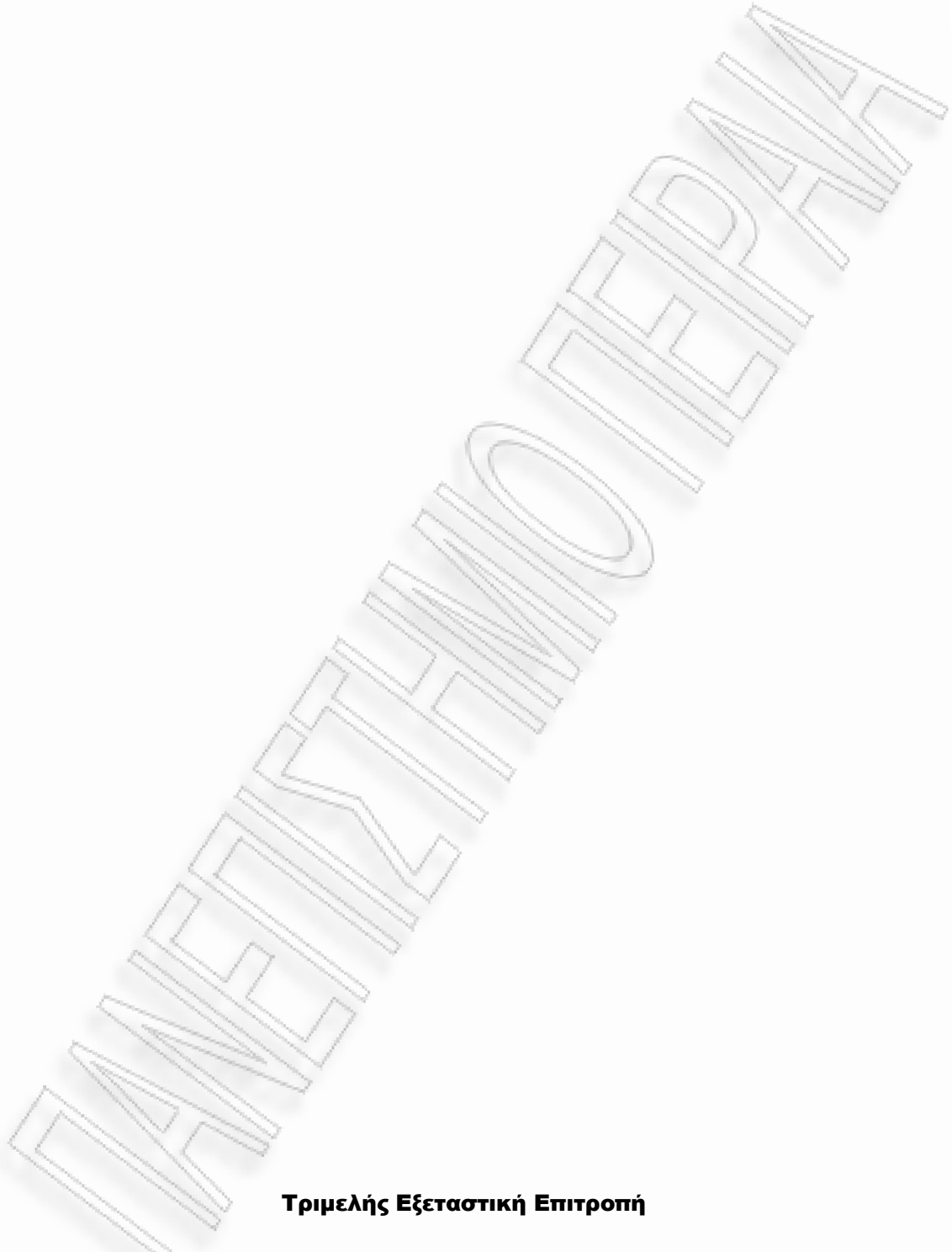


Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών
«Προηγμένα Συστήματα Πληροφορικής»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Διαδικτυακή Εφαρμογή Αντλησης Δεδομένων Εκδοτικού Οίκου Elsevier και Απεικόνισης αυτών σε Διαδραστικό Χάρτη
Όνοματεπώνυμο Φοιτητή	Ταμπάκης Παναγιώτης
Πατρώνυμο	Κωνσταντίνος
Αριθμός Μητρώου	ΜΠΣΠ/ 08023
Επιβλέπων	Δ. Δεσπότης, Καθηγητής



Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

(υπογραφή)

(υπογραφή)

Δ. Δεσπότης
Καθηγητής

Ν. Αλεξανδρής
Καθηγητής

Δ. Αποστόλου
Επικ. Καθηγητής

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

Ευχαριστίες

Θα ήθελα όμως να ευχαριστήσω την οικογένεια μου για την στήριξη τους, τον επιβλέποντα τον καθηγητή κ. Δεσπότη Δημήτριο τον υποψήφιο διδάκτορα κ. Ζυγομήτρο Αθανάσιο για την άριστη συνεργασία και την υπομονή τους.

Περίληψη

Με την ολοένα και μεγαλύτερη εξάπλωση και διάδοση του Διαδικτύου καθώς και με την συνεχόμενη ανάπτυξη των τεχνολογιών πάνω στις οποίες βασίζεται αυτό έχουν οδηγήσει στην ανάγκη ανάπτυξης Διαδικτυακών Εφαρμογών η οποίες δημιουργούν προστιθέμενη αξία. Αυτό σε συνδυασμό με την ανάγκη της ακαδημαϊκής κοινότητας για οπτικοποίηση κάποιων μεγεθών και μέτρων απόδοσης που αφορούν την έρευνα, οδήγησαν στην παρούσα Διπλωματική Εργασία. Σε αυτή δημιουργήθηκε μια Διαδικτυακή εφαρμογή με την οποία θα γίνεται άντληση δεδομένων για όλα τα papers κάποιων επιστημονικών περιοδικών του εκδοτικού οίκου, επεξεργασία των δεδομένων αυτών, μετατροπή τους σε χρήσιμη πληροφορία και απεικόνιση τους σε ένα διαδραστικό χάρτη.

Λέξεις Κλειδιά: Διαδικτυακές Εφαρμογές, Web Crawling, Web Harvesting, Οπτικοποίηση Δεδομένων, Διαδραστικοί Χάρτες, Elsevier

Abstract

With the increasing expansion and spread of the Internet and the continuing development of technologies on which it depends on, have led to the need to develop Web applications which create added value. This combined with the need of the academic community for visualization of some data and performance measures relating to research, led to this thesis. In this was created a web application with which we will be obtaining data on all papers of some scientific journals of the Publishing House, processing of these data, conversion into useful information and representation in an interactive map.

Keywords: Web Applications, Web Crawling, Web Harvesting, Data Visualization, Interactive Maps, Elsevier

Πίνακας Περιεχομένων

1. ΕΙΣΑΓΩΓΗ	1
1.1. ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ	1
2. ΕΥΦΥΗΣ ΠΡΑΚΤΟΡΕΣ	3
2.1. ΤΕΧΝΙΚΕΣ HARVESTING.....	4
2.2. ΕΡΓΑΛΕΙΑ	4
2.3. ΜΟΡΦΗ ΠΑΡΑΓΟΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ	8
3. ΕΙΣΑΓΩΓΗ ΔΕΔΟΜΕΝΩΝ ΣΤΗΝ ΒΑΣΗ	10
4. ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ	12
4.1. ΕΝΝΟΙΟΛΟΓΙΚΟΣ ΣΧΕΔΙΑΣΜΟΣ ΤΗΣ ΒΑΣΗΣ.....	12
4.2. ΛΟΓΙΚΟΣ ΣΧΕΔΙΑΣΜΟΣ ΤΗΣ ΒΑΣΗΣ	13
4.3. ΥΛΟΠΟΙΗΣΗ ΤΟΥ ΣΧΕΣΙΑΚΟΥ ΣΧΗΜΑΤΟΣ ΤΗΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ	14
4.4. ΠΥΡΟΔΟΤΗΤΕΣ (TRIGGERS).....	19
5. ΑΥΤΟΜΑΤΟΠΟΙΗΣΗ ΔΙΑΔΙΚΑΣΙΑΣ ΣΥΛΛΟΓΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΕΙΣΑΓΩΓΗΣ ΤΟΥΣ ΣΤΗΝ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ	22
6. ΑΠΕΙΚΟΝΙΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΕ ΧΑΡΤΗ	26
6.1. ΑΜΜΑΡ	26
6.2. LIVE ΣΥΝΔΕΣΗ ΑΜΜΑΡ ΜΕ ΤΗΝ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ	30
6.3. ΠΑΡΟΥΣΙΑΣΗ WEB ΕΦΑΡΜΟΓΗΣ	32
6.3.1. Αριθμός Paper ανά Χώρα	33
6.3.2. Αριθμό Paper ανά Επιστημονικό Περιοδικό και Χώρα	34
6.3.3. Κατά Κεφαλήν Αριθμός Papers – Κορυφαίες Χώρες.....	35
6.3.4. Δίκτυα Διεθνών Συνεργασιών.....	35
6.3.5. Αριθμός Paper ανά Λέξη Κλειδί και Χώρα	38
7. ΣΥΜΠΕΡΑΣΜΑΤΑ - ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ	39
8. ΒΙΒΛΙΟΓΡΑΦΙΑ	40
9. ΠΑΡΑΡΤΗΜΑ	41
9.1. WEB-HARVEST CONFIGURATION FILE.....	41
9.2. SQL SERVER DATABASE CREATION SCRIPT	44
9.3. WEB PAGE SOURCE CODE	51
9.3.1. connect2db.php	51
9.3.2. main.html.....	51
9.3.3. list.html.....	53
9.3.4. underlist1.html.....	55
9.3.5. underlist2.html.....	55
9.3.6. NumberOfArticles.php.....	56
9.3.7. form_Journals.php.....	57
9.3.8. Journals.php.....	58
9.3.9. ArticlesPerCapita.php.....	60
9.3.10. form_Cooperations.php.....	62
9.3.11. Cooperations.php	63

9.3.12.	<i>form_keywords.html</i>	67
9.3.13.	<i>keywords.php</i>	68
9.3.14.	<i>ammap_settings.xml</i>	69

Λίστα Εικόνων

Εικόνα 1: Διαγραμματική Απεικόνιση της Εφαρμογής	2
Εικόνα 2: Δομή του πίνακα της Βάσης για τον οποίο προορίζονται τα δεδομένα.....	10
Εικόνα 3: Διάγραμμα Οντοτήτων – Συσχετίσεων της Βάσης Δεδομένων	12
Εικόνα 4: Αριθμός Paper ανά Χώρα.....	33
Εικόνα 5: Αριθμός Paper Ελλάδα.....	33
Εικόνα 6: Αριθμός Paper ανά Επιστημονικό Περιοδικό και Χώρα - Επιλογή Επιστημονικού Περιοδικού.....	34
Εικόνα 7: Αριθμός Paper ανά Επιστημονικό Περιοδικό και Χώρα	35
Εικόνα 8 Κατά Κεφαλήν Αριθμός Paper – Κορυφαίες Χώρες.....	35
Εικόνα 9: Δίκτυα Διεθνών Συνεργασιών - Επιλογή Χώρας	36
Εικόνα 10: Δίκτυο Διεθνών Συνεργασιών.....	37
Εικόνα 11: Αριθμός Paper ανά Λέξη Κλειδί και Χώρα.....	38

1. Εισαγωγή

1.1. Αντικείμενο της Διπλωματικής

Αντικείμενο της παρούσας Διπλωματικής εργασίας είναι η δημιουργία μιας web εφαρμογής με την οποία θα γίνεται άντληση δεδομένων για όλα τα papers κάποιων επιστημονικών περιοδικών του εκδοτικού οίκου Elsevier, επεξεργασία των δεδομένων αυτών, μετατροπή τους σε χρήσιμη πληροφορία και απεικόνιση τους σε ένα διαδραστικό χάρτη.

Βασικό πρόβλημα κατά την δημιουργία της εφαρμογής αυτής ήταν ότι ο εκδοτικός οίκος Elsevier δεν παρέχει κάποιο API για απευθείας πρόσβαση στα δεδομένα του. Επομένως, αναγκαστήκαμε να καταφύγουμε στην λύση των ευφυών πρακτόρων και πιο συγκεκριμένα σε κάποιο web crawler ή web harvester ή web spider. Η τελική επιλογή ήταν ένα open source εργαλείο, το Web-Harvest, το οποίο θα αναλυθεί ενδελεχώς σε επόμενη ενότητα. Ένα ακόμη πρόβλημα, που προέκυψε από την επιλογή της χρήσης ενός ευφυούς πράκτορα και πιο συγκεκριμένα του Web-Harvest, ήταν ο τρόπος με τον οποίο θα γίνεται η διάτρεξη των δεδομένων. Δηλαδή, το πώς περιηγούμαστε από paper σε paper καθώς δεν υπήρχε εμφανής τρόπος ούτε μέσα από το url του κάθε άρθρου ούτε μέσα από το περιεχόμενο του.

Τη λύση σε αυτό το πρόβλημα ήρθε να δώσει το σύστημα Digital Object Identifier (DOI). Αυτό το σύστημα παρέχει κάθε ψηφιακό αντικείμενο με ένα μοναδικό κωδικό που το προσδιορίζει και με βάση το οποίο μπορείς ανατρέξεις στο συγκεκριμένο αντικείμενο. Για παράδειγμα, η url ενός paper του εκδοτικού οίκου Elsevier είναι:

<http://www.sciencedirect.com/science/article/pii/S0377221709000095>

το αντίστοιχο doi του είναι:

doi:10.1016/j.ejor.2009.01.001

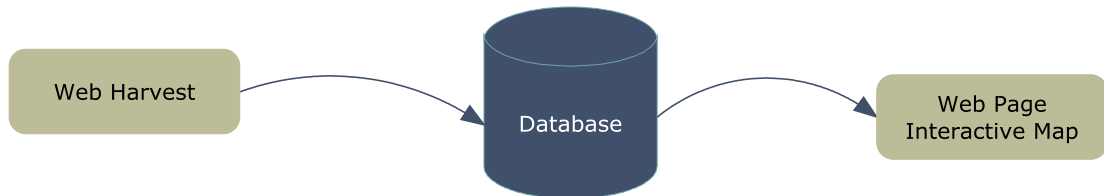
Και μπορείς να ανατρέξεις στο συγκεκριμένο paper με την url:

<http://dx.doi.org/10.1016/j.ejor.2009.01.001>

Σημασιολογικά το doi χωρίζεται σε κάποια τμήματα:

1. Το πρώτο τμήμα είναι το 10.1016 που είναι ένας κωδικός που μας παρέχει τον εκδοτικό οίκο.
2. Το δεύτερο τμήμα είναι το j.ejor. είναι ένας κωδικός που μας παρέχει το επιστημονικό περιοδικό (στη συγκεκριμένη περίπτωση είναι το European Journal of Operational Research).
3. Το τρίτο τμήμα είναι το έτος.
4. Το τέταρτο τμήμα είναι ο μήνας.
5. Και τέλος είναι ένας αύξων αριθμός από το 001 έως το 999.

Αφού ολοκληρωθεί η άντληση των δεδομένων αυτά θα πρέπει να αποθηκευθούν σε κάποια Βάση Δεδομένων όπου θα υποστούν και κάποια επεξεργασία έτσι ώστε να μετατραπούν τα πρωτογενή αυτά δεδομένα σε κάποια χρήσιμη πληροφορία. Στη συνέχεια, θα πρέπει η πληροφορία αυτή να οπτικοποιηθεί και απεικονιστεί σε έναν διαδραστικό χάρτη. Τέλος, θα πρέπει όλη αυτή η ροή εργασίας και δεδομένων να αυτοματοποιηθεί έτσι ώστε να μην χρειάζεται ανθρώπινη παρέμβαση για τη λειτουργία της. Μία χονδροειδής διαγραμματική απεικόνιση της εφαρμογής μας θα μπορούσε να είναι η εξής.



Εικόνα 1: Διαγραμματική Απεικόνιση της Εφαρμογής

2. Ευφυής Πράκτορες

Αποτελώντας ίσως το όνειρο κάθε μηχανικού μισό αιώνα νωρίτερα, το διαδίκτυο, αυτό το σύνολο από συνδεδεμένα δίκτυα υπολογιστών σε όλο τον Πλανήτη, προσφέρει την υποδομή και το πλαίσιο πάνω στο οποίο λειτουργούν οι υπηρεσίες του (ηλεκτρονικό ταχυδρομείο, απομακρυσμένη σύνδεση σε υπολογιστή, μεταφορά αρχείων, κανάλια ταυτόχρονης επικοινωνίας, παγκόσμιος ιστός) με σημαντικότερη ίσως και σίγουρα ταχύτερα εξελισσόμενη από αυτές αυτή του Παγκοσμίου Ιστού. Η ύπαρξη και η σημερινή μορφή του Παγκοσμίου Ιστού είναι κάτι που δύσκολα θα μπορούσε να φανταστεί κάποιος μερικές μόνο δεκαετίες νωρίτερα. Η επανάσταση της Πληροφορίας, στην εποχή της οποίας ζούμε και η ανθρωπότητα πέρασε με την εμφάνιση των πρώτων ηλεκτρονικών υπολογιστών, έχει επιφέρει τεράστιες αλλαγές στους περισσότερους, αν όχι σε όλους τους τομείς της ζωής μας. Αυτές οι αλλαγές δεν ήταν αρχικά εμφανείς στο ευρύ κοινό μέχρι την εποχή που πρωτοεμφανίστηκαν οι υπηρεσίες του διαδικτύου. Από τον τρόπο με τον οποίο εργαζόμαστε και ενημερωνόμαστε μέχρι και τον τρόπο με τον εκφραζόμαστε και επικοινωνούμε οι αλλαγές που υπάρχουν σε σχέση με ότι συνέβαινε στις αρχές π.χ. της δεκαετίας του 1990 είναι τεράστιες. Μέσα σε αυτή την κοινωνία της Πληροφορίας, όπως εύστοχα ονομάζεται, ο καθένας μπορεί να περιηγηθεί, περιπλανηθεί, αναζητήσει άλλα και ταυτόχρονα να προσθέσει κάτι στον ήδη τεράστιο όγκο πληροφορίας που είναι διαθέσιμος με πολύ μεγάλη ευκολία και διαμέσου μόνο του υπολογιστή μας. Ειδικότερα δε, στην εκπαιδευτική διαδικασία, ο πλούτος της προσφερόμενης γνώσης που είναι διαθέσιμη τόσο στον εκπαιδευτή όσο και στον εκπαιδευόμενο είναι ανεκτίμητος. Τόσο γιατί υπάρχει πρόσβαση σε τεράστιες ποσότητες πληροφορίας σχετικά με οποιοδήποτε θέμα όσο και γιατί υπάρχει πρόσβαση στις τελευταίες εξελίξεις και στις πιο «ενημερωμένες εκδόσεις» αυτής της πληροφορίας. Υπάρχουν όμως δυο ζητήματα που προκύπτουν: Πρώτον, πως μπορούμε να είμαστε ενημερωμένοι για το τι υπάρχει διαθέσιμο την τρέχουσα χρονική στιγμή στον τεράστιο και ταχύτατα εξελισσόμενο Παγκόσμιο Ιστό και δεύτερον, πως μπορούμε να είμαστε σίγουροι για την εγκυρότητα της διαθέσιμης πληροφορίας ; Η ανακοίνωση αυτή αναφέρεται με συντομία στο πρώτο ζήτημα.

Κατά τη δεκαετία του '90, με την αυξανόμενη σημασία του Internet, ανάπτυξη γνώρισαν οι ευφυείς πράκτορες, αυτόνομο λογισμικό TN τοποθετημένο σε κάποιο περιβάλλον με το οποίο αλληλεπιδρά, οι οποίοι βρήκαν μεγάλο πεδίο εφαρμογών λόγω της εξάπλωσης του Διαδικτύου. Οι πράκτορες στοχεύουν συνήθως στην παροχή βοήθειας στους χρήστες τους, στη συλλογή ή ανάλυση γιγάντιων συνόλων δεδομένων ή στην αυτοματοποίηση επαναλαμβανόμενων εργασιών (π.χ. βλέπε διαδικτυακό ρομπότ), ενώ στους τρόπους κατασκευής και λειτουργίας τους συνοψίζουν όλες τις γνωστές μεθοδολογίες TN που αναπτύχθηκαν με το πέρασμα του χρόνου.

Οι ευφυείς Πράκτορες Λογισμικού (ΠΛ) μπορούν να θεωρηθούν ως οντότητες λογισμικού, οι οποίες επιδεικνύουν ένα σύνολο από χαρακτηριστικά, όπως αυτονομία, επικοινωνία, κινητικότητα και προνοητικότητα. Οι ΠΛ είναι επίσης σε ικανοί να επιδεικνύουν ευφυείς συμπεριφορές, μέσα στο πλαίσιο κοινωνιών πρακτόρων. Τα συστήματα πολλαπλών πρακτόρων (ΣΠΠ - multi-agent systems) είναι επομένως το περιβάλλον στο οποίο ευφυείς, αυτόνομοι ΠΛ αλληλεπιδρούν για την επίτευξη ατομικών (ανταγωνιστικά ΣΠΠ) ή συλλογικών (συνεργατικά ΣΠΠ) στόχων.

Μια καλύτερη λύση, ειδικά όταν μας ενδιαφέρει να εκμεταλλευτούμε μια ευρεία γκάμα στοιχείων, εναπόκειται στα screen/Web scraping ή Web extraction - mining Web ή Web harvesting εργαλεία. Το scraping είναι η διαδικασία της χρησιμοποίησης των εργαλείων λογισμικού για την ανάλυση περιεχόμενου που γράφτηκε αρχικά για κατανάλωση από ανθρώπους με σκοπό την εξαγωγή σημασιολογικών δομών δεδομένων αντιπροσωπευτικών αυτών των πληροφοριών που μπορούμε να χρησιμοποιήσουμε και να διαχειριστούμε προγραμματιστικά.

Το λογισμικό αυτό αυτόματα εξάγει τις πληροφορίες από τον Ιστό και συνεχίζει από εκεί που σταματούν οι μηχανές αναζήτησης, κάνοντας την εργασία που αυτές δεν μπορούν. Κάθε Web site και κάθε ιστοσελίδα είναι δομημένη χρησιμοποιώντας κάποια λογική. Απαιτείται επομένως να περιγραφεί η αντίστροφη διαδικασία - πώς να προσκομιστούν τα επιθυμητά στοιχεία από το

μικτό περιεχόμενο. Το Web harvesting αναφέρεται στην ανάκτηση και επεξεργασία των δεδομένων που έχουν συλλέξει από πόρους Ιστού. Τα εργαλεία λογισμικού για Web harvesting βρίσκουν, ανακτούν, και έξυπνα επεξεργάζονται μη δομημένα δεδομένα που είναι διαθέσιμα μέσω του web.

Σε μερικές περιπτώσεις ο στόχος είναι ακόμα πιο σύνθετος από αυτόν: Τα στοιχεία μπορεί να είναι διασκορπισμένα σε περισσότερες σελίδες, να χρειάζεται πυροδότηση ενός GET/POST request για τη λήψη της σελίδας εισόδου ή να απαιτείται η έγκριση για πλοήγηση στη σελίδα ενδιαφέροντος.

2.1. Τεχνικές Harvesting

Υπάρχουν τρεις τρόποι που μπορούμε να εξαγάγουμε χρήσιμες πληροφορίες από τον Ιστό. Η πρώτη τεχνική, το **harvesting web περιεχομένου**, ενδιαφέρεται άμεσα για το συγκεκριμένο περιεχόμενο των εγγράφων ή των περιγραφών τους, όπως HTML αρχεία, εικόνες ή e-mail μηνύματα. Δεδομένου ότι τα περισσότερα έγγραφα κειμένων είναι σχετικά μη δομημένα (τουλάχιστον όσον αφορά στην ερμηνεία από μηχανές), μια κοινή προσέγγιση είναι να χρησιμοποιηθεί ότι είναι ήδη γνωστό για τη γενική δομή των εγγράφων και να αντιστοιχιστεί σε κάποιο μοντέλο δεδομένων. Μια άλλη προσέγγιση περιλαμβάνει την προσπάθεια βελτίωσης στις αναζητήσεις περιεχομένου που εκτελούν εργαλεία όπως οι μηχανές αναζήτησης. Αυτός ο τύπος content harvesting υπερβαίνει την εξαγωγή λέξεων κλειδιών και την παραγωγή απλών στατιστικών σχετικά με λέξεις και φράσεις στα έγγραφα.

Μια άλλη τεχνική, το **harvesting Web δομής**, εκμεταλλεύεται το γεγονός ότι οι ιστοσελίδες μπορούν να αποκαλύψουν περισσότερες πληροφορίες από ότι μόνο το προφανές περιεχόμενό τους. Οι σύνδεσμοι από άλλες πηγές που δείχνουν στην ιστοσελίδα δείχνουν τη δημοτικότητα της σελίδας αυτής, ενώ σύνδεσμοι μέσα στη σελίδα που δείχνουν σε άλλους πόρους μπορεί να υποδεικνύουν την αφθονία ή την ποικιλία των θεμάτων που καλύπτονται σε εκείνη την σελίδα.

Η τρίτη τεχνική, το **harvesting Web χρήσης**, χρησιμοποιεί δεδομένα που καταγράφονται από Web servers για τις αλληλεπιδράσεις χρηστών για να βοηθήσει στην κατανόηση της συμπεριφοράς χρηστών και την εκτίμηση της αποτελεσματικότητας της Web δομής. Η καταγραφή μοτίβων γενικής πρόσβασης αναλύει τα Web logs για την κατανόηση μοτίβων πρόσβασης και τάσεις προκειμένου να αναγνωριστούν δομικά ζητήματα και ομαδοποίηση πόρων. Η καταγραφή προσαρμοσμένης χρήσης αναλύει μεμονωμένες τάσεις έτσι ώστε οι ιστοχώροι να μπορούν να προσωποποιηθούν σε συγκεκριμένους χρήστες. Κατά τη διάρκεια του χρόνου, βασισμένου στα μοτίβα πρόσβασης, ένας ιστότοπος μπορεί να προσαρμόζεται δυναμικά για έναν χρήστη από την άποψη των πληροφοριών που επιδεικνύονται, του βάθους της δομής του site και της μορφής των πόρων που παρουσιάζονται.

2.2. Εργαλεία

Θα παρουσιάσουμε ένα εργαλείο που θα μπορούσε να χρησιμοποιηθεί για να κάνουμε scraping στο προκειμένου να εξάγουμε τα απαιτούμενα στοιχεία για τα εργαλεία κάθε κατηγορίας όπως τα έχουμε καθορίσει.

Το **Web-Harvest** είναι ένα ανοιχτού κώδικα εξαγωγής δεδομένων Ιστού εργαλείο γραμμένο σε Java. Προσφέρει έναν τρόπο να συλλέξουμε επιθυμητές ιστοσελίδες και να εξάγουμε χρήσιμα δεδομένα από αυτές. Προκειμένου να γίνει αυτό, χρησιμοποιεί καθιερωμένες τεχνικές και τεχνολογίες για διαχείριση κειμένου/xml όπως *XSLT*, *XQuery* και *Regular Expressions*. Το Web-Harvest κυρίως εστιάζει σε **HTML/XML βασισμένα web sites**, τα οποία αποτελούν ακόμα τη μεγάλη πλειοψηφία του περιεχομένου Ιστού. Αφ' ετέρου, θα μπορούσε να συμπληρωθεί εύκολα από συνθήκες Java βιβλιοθήκες προκειμένου να αυξηθούν οι ικανότητες εξαγωγής του. Ο κύριος στόχος πίσω από το εργαλείο δεν είναι να προτείνει μια νέα μέθοδο, αλλά να παρέχει ένα τρόπο για εύκολη χρήση και συνδυασμό των διαθέσιμων τεχνολογιών εξαγωγής.

Κάθε διαδικασία εξαγωγής στο Web-Harvest καθορίζεται από το χρήστη μέσω XML-based *configuration αρχείων*. Κάθε αρχείο διαμόρφωσης περιγράφει μια ακολουθία επεξεργασιών

εκτελώντας κάποιους κοινές εργασίες προκειμένου να ολοκληρωθεί ο τελικός στόχος. Οι επεξεργαστές εκτελούνται υπό μορφή *pipeline*. Κατά συνέπεια, η έξοδος μιας εκτέλεσης επεξεργαστή εισάγεται σε έναν άλλο. Αυτό μπορεί να εξηγηθεί καλύτερα χρησιμοποιώντας το απλό κομμάτι διαμόρφωσης:

```
<xpath expression="//a[@shape='rect']/@href"> <html-to-xml>
<http url="http://www.somesite.com/">
</html-to-xml> </xpath>
```

Όταν το Web-Harvest εκτελεί αυτό το κομμάτι, γίνονται τα ακόλουθα βήματα:

1. Ο *http* επεξεργαστής κατεβάζει περιεχόμενο από το καθορισμένο URL.
2. Ο *html-to-xml* επεξεργαστής «καθαρίζει» αυτήν την HTML παράγοντας XHTML περιεχόμενο.
3. Ο *xpath* επεξεργαστής αναζητά συγκεκριμένους συνδέσμους στο XHTML από το προηγούμενο βήμα δίνοντας τη URL ακολουθία σαν αποτέλεσμα.

Το Web-Harvest υποστηρίζει ένα σύνολο χρήσιμων επεξεργασιών για το χειρισμό μεταβλητών, υπό όρους διακλαδώσεις, συναρτήσεις, HTML και XML επεξεργασία κ.α. Το Web-Harvest λειτουργεί με τη διαμόρφωση κάθε ιστοτόπου σε ένα XML αρχείο που περιγράφει μια ακολουθία επεξεργασιών για τη μετατροπή της HTML σε χρήσιμη XML. Το αποτέλεσμα της εξαγωγής θα μπορούσε να είναι διαθέσιμο σε αρχεία που δημιουργούνται κατά την εξαγωγή ή από από το πλαίσιο μεταβλητών αν το Web-Harvest χρησιμοποιείται προγραμματιστικά.

Όλα τα στοιχεία που παράγονται και που καταναλώνονται κατά τη διάρκεια της διαδικασίας εξαγωγής στο Web-Harvest έχουν τρεις αναπαραστάσεις: κείμενο, δυαδικό και λίστες. Ποια μορφή χρησιμοποιείται - εξαρτάται από τον επεξεργαστή που καταναλώνει τα δεδομένα. Ο *html-to-xml* επεξεργαστής χρησιμοποιεί το downloaded περιεχόμενο σαν κείμενο προκειμένου να το μετασχηματίσει σε HTML, ο *loop* επεξεργαστής χρησιμοποιεί τη μεταβλητή *urlList* σαν λίστα για τις επαναλήψεις κ.α.. Στις περισσότερες περιπτώσεις η κατάλληλη αναπαράσταση επιλέγεται από το Web-Harvest. Εκτός από το σύνολο ισχυρών επεξεργασιών κειμένων και χειρισμού XML, το Web-Harvest υποστηρίζει πραγματικές scripting γλώσσες των οποίων ο κώδικας μπορεί να ενσωματωθεί στις scraper configurations.

Στο Παράρτημα παρατίθεται το configuration αρχείο το οποίο θα χρησιμοποιείται στην εφαρμογή μας έτσι ώστε να συλλέγουμε μια φορά το μήνα τα δεδομένα που μας ενδιαφέρουν για ένα συγκεκριμένο επιστημονικό περιοδικό. Στην παρούσα φάση θα ήταν καλύτερο να βλέπαμε κομμάτια του κώδικα του configuration αρχείου μαζί με την επεξήγηση της λειτουργίας τους. Ως παράδειγμα το επιστημονικό περιοδικό European Journal of Operational Research :

```
<var-def name="startUrl">http://dx.doi.org/10.1016/</var-def>
<var-def name="midUrl">j.ejor.</var-def>
<var-def name="YYYYUrl">
<script return=" thisyear">
    thisyear = Integer.parseInt(sys.datetime("yyyy"));
</script>
</var-def>
<var-def name="MM.Url"><template>${sys.datetime(".MM.")}</template></var-def>
<var-def name="i">
<var-def name="j">
<var-def name="k">
    <file action="write" path="${midUrl}.xml" charset="UTF-8">
        <![CDATA[ <ROOT> ]]>
```

```

</file>
<while condition="true" index="i" maxloops="9">
<while condition="true" index="j" maxloops="9">
<while condition="true" index="k" maxloops="9">

```

Στο κομμάτι αυτό ορίζουμε τις διάφορες μεταβλητές που θα μας βοηθήσουν αργότερα στην συλλογή των url που μας ενδιαφέρουν καθώς και στην πρόσβαση και ανάκτηση των δεδομένων. Πιο συγκεκριμένα,

- η μεταβλητή startUrl μας προμηθεύει με το πρώτο κομμάτι της url που θα είναι ίδιο για όλα τα επιστημονικά περιοδικά για τον συγκεκριμένο εκδοτικό οίκο.
- η μεταβλητή midUrl μας προμηθεύει με το μεσαίο κομμάτι της url που μας ορίζει ένα συγκεκριμένο επιστημονικό περιοδικό
- η μεταβλητή YYYYUrl μας προμηθεύει με τα 4 ψηφία του έτους που διανύουμε
- η μεταβλητή MM.Url προμηθεύει με τα 2 ψηφία του μήνα που διανύουμε
- οι 3 μεταβλητές i, j και k που μας προμηθεύουν με τον αύξων αριθμό του paper για το συγκεκριμένο επιστημονικό περιοδικό, έτος και μήνα
- στη συνέχεια, με το <file action="write" path="{midUrl}.xml" charset="UTF-8"> <![CDATA[<ROOT>]]> </file>, βλέπουμε την δημιουργία και το άνοιγμα του αρχείου που θα φιλοξενήσει τα δεδομένα που μας ενδιαφέρουν
- τέλος, με τα 3 loops που δημιουργούνται θα αναζητήσουμε όλα τα δυνατά url για το συγκεκριμένο επιστημονικό περιοδικό, έτος και μήνα

```

file action="append" path="{midUrl}.xml" charset="UTF-8">
<loop item="articleUrl" index="l" maxloops="1">

<list>
<xpath expression="//div[@class='articleText']">
<html-to-xml>
<http url="{startUrl}{midUrl}{YYYYUrl}{MM.Url}{i}{j}{k}"/>
</html-to-xml>
</xpath>
<xpath expression="//div[@class='articleTitle']">
<html-to-xml>
<http url="{startUrl}{midUrl}{YYYYUrl}{MM.Url}{i}{j}{k}"/>
</html-to-xml>
</xpath>
<xpath expression="//div[@class='articleHeaderInner']">
<html-to-xml>
<http url="{startUrl}{midUrl}{YYYYUrl}{MM.Url}{i}{j}{k}"/>
</html-to-xml>
</xpath>
</list>

```

Στο κομμάτι αυτό του κώδικα συλλέγουμε όλες τις url και τις συγκεκριμένες παραγράφους που μας ενδιαφέρουν

```

<xquery>
  <xq-param name="doc">
    <html-to-xml>
      <http url="{startUrl}{midUrl}{YYYYUrl}{MM.Url}{i}{j}{k}"/>
    </html-to-xml>
  </xq-param>
  <xq-expression><![CDATA[
    declare variable $doc as node() external;
    let $aff := data($doc//div[@class='articleHeaderInner'])
    let $aff7 := data($doc//div[@class='articleHeaderInner'])
    let $aff1 := data($doc//div[@class='articleTitle'])
    let $aff2 := data($doc//div[@class='articleText authorsNoEnt'])

    return
      <article>
        <id>{data(substring-before((substring-after($aff,
          " |"))}</id>
          <scientific_field>{data(substring-before((substring-after($aff7,
"Reprints")), $aff1))}</scientific_field>
          <title>{data($aff1)}</title>
          <affiliation>{data($aff2)}</affiliation>
          <received>{data(substring-before((substring-after($aff,
"Received
")), ";"))}</received>
          <accepted>{data(substring-before((substring-after($aff,
"accepted
")), "."))}</accepted>
          <abstract>{data(substring-before((substring-after($aff,
"Abstract")),
"Keywords"))}</abstract>
          <keywords>{data(replace((substring-before((substring-after($aff,
"Keywords:")), "Article Outline")), ";", ","))}</keywords>
        </article>
      ]]></xq-expression>
</xquery>
</body>

</loop>
</file>

</while>
</while>
</while>
<file action="append" path="{midUrl}.xml" charset="UTF-8">
  <![CDATA[ </ROOT> ]]>

```

Τέλος, στο κομμάτι αυτό του κώδικα συλλέγουμε τα δεδομένα που μας ενδιαφέρουν και κλείνουμε το αρχείο με τα δεδομένα.

Γενικότερα το scraping θεωρείται συχνά άκομψη λύση, και αυτό γιατί έχει ένα βασικό μειονεκτήματα. Αντίθετα από τα APIs με interfaces, το scraping δεν έχει καμία συγκεκριμένη

προγραμματική σύμβαση μεταξύ του προμηθευτή και του καταναλωτή του περιεχομένου. Οι scrapers πρέπει να σχεδιάσουν τα εργαλεία τους γύρω από ένα πρότυπο του περιεχομένου πηγής και να ελπίζουν ότι ο πάροχος εμμένει σε αυτό το πρότυπο παρουσίασης με συνέπεια. Οι ιστοτόποι τείνουν να εξετάζουν και να τροποποιούν το look-and-feel τους περιοδικά για να παραμένουν ανανεωμένοι και κομψοί, κάτι που δημιουργεί προβλήματα όσον αφορά την συντήρηση από την πλευρά των scrapers. Με τον ίδιο τρόπο οποιαδήποτε αλλαγή στους ιστοτόπους που χρησιμοποιούμε θα αλλάξει και όλο το σχεδιασμό μας για την εξαγωγή πληροφορίας. Ταυτόχρονα διαφορετικές πηγές σημαίνουν και ποικιλία στη δομή και παρουσίαση των δεδομένων. Έτσι θα πρέπει να προσαρμόζουμε κάθε φορά το scraping σύμφωνα με τα δεδομένα του συγκεκριμένου ιστοτόπου, καθώς δεν υπάρχει ενιαίος τρόπος αντιμετώπισης όλων των περιπτώσεων.

2.3. Μορφή Παραγόμενων Δεδομένων

Έχοντας ολοκληρώσει την διαδικασία συλλογής δεδομένων, προχωράμε στην αποθήκευση των δεδομένων αυτών. Αυτά είναι xml αρχεία της μορφής:

```
<article>
  <id>10.1016/j.ejor.2008.05.001</id>
  <scientific_field>Discrete Optimization</scientific_field>
  <title>Knapsack problems with setups</title>
  <affiliation>aISEL-LMAH, Universit  du Havre, Quai Frissard, BP 1137, 76063 Le Havre Cedex, FranceOrange Labs, recherche and d veloppement, 38β 40 rue G n ral Leclerc, 92794 Issy-Les-Moulineaux Cedex 9, FranceInstitut de Math matiques de Bordeaux, Universit  Bordeaux 1, 351 Cours de la Liberation, F-33405 Telence Cedex, France</affiliation>
  <received>30 April 2007</received>
  <accepted>1 May 2008</accepted>
  <abstract>Knapsack problems with setups find their application in many concrete industrial and financial problems. Moreover, they also arise as subproblems in a Dantzig-Wolfe decomposition approach to more complex combinatorial optimization problems, where they need to be solved repeatedly and therefore efficiently. Here, we consider the multiple-class integer knapsack problem with setups. Items are partitioned into classes whose use implies a setup cost and associated capacity consumption. Item weights are assumed to be a multiple of their class weight. The total weight of selected items and setups is bounded. The objective is to maximize the difference between the profits of selected items and the fixed costs incurred for setting-up classes. A special case is the bounded integer knapsack problem with setups where each class holds a single item and its continuous version where a fraction of an item can be selected while incurring a full setup. The paper shows the extent to which classical results for the knapsack problem can be generalized to these variants with setups. In particular, an extension of the branch-and-bound algorithm of Horowitz and Sahni is developed for problems with positive setup costs. Our direct approach is compared experimentally with the approach proposed in the literature consisting in converting the problem into a multiple choice knapsack with pseudo-polynomial size.</abstract>
  <keywords>Knapsack problem, Fixed cost, Setup, Variable upper bound, Branch-and-bound</keywords>
</article>
```

Πιο συγκεκριμένα, έχουμε:

- Το tag <article>, το οποίο ορίζει κάθε paper. Ένα τέτοιο xml αρχείο μπορεί να έχει από κανένα έως 999 papers.
- Μέσα στο κάθε tag <article> έχουμε 8 attributes:

- Το <id>, οποίο προσδιορίζει μονοσήμαντα το κάθε paper και είναι ο κωδικός doi
- Το <scientific_field>, το οποίο είναι το επιστημονικό πεδίο στο οποίο ανήκει το συγκεκριμένο paper. Εδώ θα πρέπει να αναφέρουμε, πάντως, ότι στα περισσότερα από τα επιστημονικά περιοδικά με τα οποία θα ασχοληθούμε δεν μας δίνεται το συγκεκριμένο πεδίο.
 - Το <title>, το οποίο είναι ο τίτλος του συγκεκριμένου paper
 - Το <affiliation>, το οποίο είναι η επαγγελματική διεύθυνση του συγγραφέα ή των συγγραφέων. Το πεδίο αυτό είναι πολύ σημαντικό καθώς θα μας προμηθεύσει με την χώρα ή τις χώρες προέλευσης των papers
 - Το <received>, το οποίο είναι η ημερομηνία παραλαβής του paper από το επιστημονικό περιοδικό
 - Το <accepted>, το οποίο είναι η ημερομηνία αποδοχής του paper από το επιστημονικό περιοδικό
 - Το <abstract>, το οποίο είναι μια σύντομη περιγραφή του paper
 - Το <keywords>, το οποίο είναι οι λέξεις κλειδιά του συγκεκριμένου paper.

3. Εισαγωγή δεδομένων στην Βάση

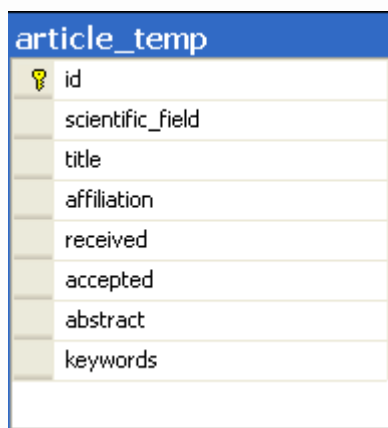
Όπως προαναφέραμε, τα δεδομένα τα οποία παίρνουμε από την διαδικασία του crawling είναι xml αρχεία και δεν μας δίνεται η δυνατότητα από το συγκεκριμένο λογισμικό να εισάγουμε τα δεδομένα μας απευθείας στην Βάση μας μέσω κάποιου JDBC Driver. Επομένως, θα πρέπει να βρούμε κάποιο τρόπο έτσι ώστε να τα εισάγουμε με «έμμεσο» τρόπο.

Τη λύση έρχεται να μας δώσει ένα αυτόνομο COM αντικείμενο το οποίο μας επιτρέπει την εισαγωγή xml δεδομένων σε πινάκες του SQL Server. Το αντικείμενο αυτό είναι το XML Bulk Load. Θα μπορούσαμε να εισάγουμε xml δεδομένα σε μία Βάση Δεδομένων του SQL Server χρησιμοποιώντας ένα INSERT statement και την συνάρτηση OPENXML. Παρόλαυτα, το αντικείμενο XML Bulk Load μαζικής εισαγωγής xml δεδομένων μας παρέχει καλύτερη απόδοση στην εισαγωγή μεγάλου όγκου δεδομένων.

Για να εκτελεστεί η λειτουργία του XML Bulk Load χρειάζεται,

1. όπως είναι προφανές, ένα αρχείο με τα xml δεδομένα ή ένα stream δεδομένων
2. και ένα αρχείο, της μορφής XSD (XML Schema Definition) ή της μορφής (XDR) XML-Data Reduced schema, το οποίο θα κάνει ουσιαστικά την αντιστοίχιση των attributes των xml δεδομένων με τα attributes του πίνακα της Βάσης Δεδομένων στον οποίο θέλουμε να εισάγουμε τα δεδομένα μας.

Έχουμε ήδη δει την δομή του αρχείου με τα xml δεδομένα, επομένως ας δούμε και την δομή του πίνακα της Βάσης για τον οποίο προορίζονται τα δεδομένα αυτά.



article_temp	
id	Primary Key
scientific_field	
title	
affiliation	
received	
accepted	
abstract	
keywords	

Εικόνα 2: Δομή του πίνακα της Βάσης για τον οποίο προορίζονται τα δεδομένα

Λεπτομέρειες, πάντως, για την υλοποίηση της Βάσης θα δούμε αργότερα. Εφόσον, έχουμε δει και την δομή του αρχείου με τα xml δεδομένα και την δομή του πίνακα της Βάσης για τον οποίο προορίζονται τα δεδομένα αυτά, μπορούμε να προχωρήσουμε στο XSD αρχείο το οποίο θα κάνει ουσιαστικά την αντιστοίχιση των attributes των xml δεδομένων με τα attributes του πίνακα της Βάσης Δεδομένων στον οποίο θέλουμε να εισάγουμε τα δεδομένα μας. Έτσι έχουμε το αρχείο xmlmapping.xml:

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:sql="urn:schemas-microsoft-com:mapping-schema">

  <xsd:element name="article" sql:relation="article" >
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="id" type="xsd:string" />
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

```

<xsd:element name="scientific_field" type="xsd:string" />
<xsd:element name="title" type="xsd:string" />
<xsd:element name="affiliation" type="xsd:string" />
<xsd:element name="received" type="xsd:string" />
<xsd:element name="accepted" type="xsd:string" />
<xsd:element name="abstract" type="xsd:string" />
<xsd:element name="keywords" type="xsd:string" />
</xsd:sequence>
</xsd:complexType>
</xsd:element>
</xsd:schema>

```

Τέλος, μπορούμε να δούμε τον κώδικα του αντικειμένου XML Bulk Load:

```

set objBL = CreateObject("SQLXMLBulkLoad.SQLXMLBulkload.4.0")
objBL.ConnectionString = "Provider=SQLNCLI.1;Data Source=PANOSBP;Integrated
Security=SSPI;Initial Catalog=webharvest"
objBL.ErrorLogFile = "C:\Documents and Settings\Admin\Desktop\Master Thesis\error.log"
objBL.FireTriggers = True
objBL.Execute "C:\webharvest1-project\xmlmapping.xml", "C:\webharvest1-
project\j.cie\2000.01..xml"

```

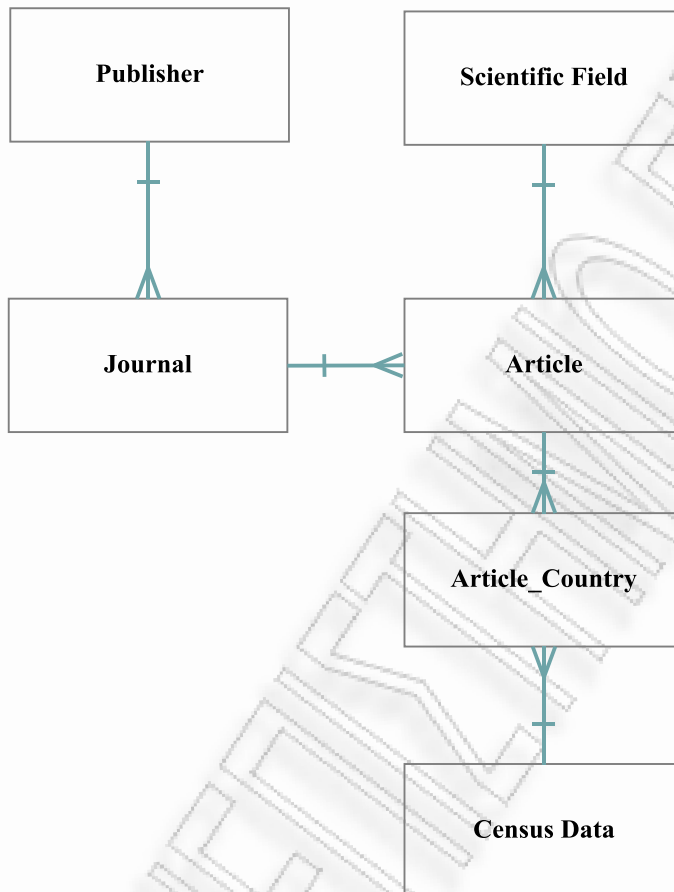
Όπου,

- στην αρχή δημιουργούμε το αντικείμενο XML Bulk Load
- έπειτα ορίζουμε το Connection String προς την Βάση μας
- δημιουργούμε ένα error file για να μας βοηθήσει σε περίπτωση σφάλματος
- θέτουμε το FireTriggers = True έτσι ώστε να μπορούμε να χρησιμοποιήσουμε on insert triggers στην βάση μας. Λεπτομέρειες γι'αυτό θα δώσουμε αργότερα.
- Και τέλος, ορίζουμε την θέση του αρχείου με τα xml δεδομένα και την θέση του XSD αρχείου το οποίο θα κάνει ουσιαστικά την αντιστοίχιση των attributes των xml δεδομένων με τα attributes του πίνακα της Βάσης Δεδομένων.

4. Βάση Δεδομένων

4.1. Εννοιολογικός Σχεδιασμός της βάσης

Παρακάτω φαίνεται το διάγραμμα Οντοτήτων – Συσχετίσεων που προέκυψε



Εικόνα 3: Διάγραμμα Οντοτήτων – Συσχετίσεων της Βάσης Δεδομένων

4.2. Λογικός Σχεδιασμός της βάσης

Το Σχεσιακό Μοντέλο που προκύπτει από το παραπάνω E-R διάγραμμα έχει τους παρακάτω πίνακες:

- **Πίνακας Article**

(id, scientific_field, title, affiliation, received, accepted, abstract, keywords, JournalID)

- Πρωτεύων κλειδί στον πίνακα αυτόν είναι το id
- Το πεδίο JournalID είναι foreign Key στο πίνακα Journal
- Το πεδίο scientific_field είναι foreign Key στο πίνακα Scientific_Field

- **Πίνακας Journal**

(JournalID, JournalName, PublisherID)

- Πρωτεύων κλειδί στον πίνακα αυτόν είναι το JournalID
- Το πεδίο PublisherID είναι foreign Key στο πίνακα Publisher

- **Πίνακας Publisher**

(PublisherID, PublisherName)

- Πρωτεύων κλειδί στον πίνακα αυτόν είναι το PublisherID

- **Πίνακας Scientific_Field**

(Scientific_Field_Code, Scientific_Field)

- Πρωτεύων κλειδί στον πίνακα αυτόν είναι το Scientific_Field_Code

- **Πίνακας CensusData**

(Country / Territory , ISO Alpha-2, ISO Alpha-3, Population, Latitude, Longitude, ISO 3166 Country Name, Alternative name 0, Alternative name 1, Alternative name 2, Alternative name 3, Alternative name 4, Alternative name 5, Alternative name 6, Alternative name 7, Alternative name 8, Alternative name 9, Alternative name 10, Alternative name 11, Alternative name 12, IncomePerCapita)

- Πρωτεύων κλειδί στον πίνακα αυτόν είναι το ISO Alpha-2

- **Πίνακας ArticleCountry**

(CountryName, CountryID, ArticleID)

- Πρωτεύων κλειδί στον πίνακα αυτόν είναι το CountryID, ArticleID
- Το πεδίο CountryID είναι foreign Key στο πίνακα CensusData
- Το πεδίο ArticleID είναι foreign Key στο πίνακα Article

4.3. Υλοποίηση του Σχεσιακού Σχήματος της Βάσης Δεδομένων

Η υλοποίηση Σχεσιακού Σχήματος της Β.Δ πραγματοποιήθηκε στον SQL Server 2005 με το παρακάτω script:

```

SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[article]')
AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[article](
[id] [varchar](50) NOT NULL,
[scientific_field] [varchar](max) NULL,
[title] [varchar](max) NULL,
[affiliation] [varchar](max) NULL,
[received] [varchar](max) NULL,
[accepted] [varchar](max) NULL,
[abstract] [varchar](max) NULL,
[keywords] [varchar](max) NULL,
[JournalID] [nvarchar](50) NULL,
CONSTRAINT [PK_article] PRIMARY KEY CLUSTERED
(
[id] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO

SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO

IF NOT EXISTS (SELECT * FROM sys.triggers WHERE object_id =
OBJECT_ID(N'[dbo].[DeletefromArticleCountry]'))
EXEC dbo.sp_executesql @statement = N'CREATE TRIGGER [dbo].[DeletefromArticleCountry]
ON [dbo].[article]
AFTER DELETE
AS
DELETE FROM ArticleCountry
FROM DELETED INNER JOIN
ArticleCountry ON DELETED.id = ArticleCountry.ArticleID

```

```

GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

IF NOT EXISTS (SELECT * FROM sys.triggers WHERE object_id =
OBJECT_ID(N'[dbo].[InsertIntoArticleCountry]'))
EXEC dbo.sp_executesql @statement = N'

-- =====
-- Author:          <Author,,Name>
-- Create date:    <Create Date,,>
-- Description:    <Description,,>
-- =====

CREATE TRIGGER [dbo].[InsertIntoArticleCountry]
ON [dbo].[article]
AFTER INSERT
AS
INSERT INTO ArticleCountry
SELECT CensusData.[Country / Territory ], CensusData.[ISO Alpha-2], INSERTED.id
FROM     INSERTED INNER JOIN CensusData
ON (INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Country /
Territory]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[ISO 3166
Country Name]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 0]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 1]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 2]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 3]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 4]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 5]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 6]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 7]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 8]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative
name 9]) }, "%") } + "%"

```

```
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative name 10]) }, "%") } + "%"
```

```
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative name 11]) }, "%") } + "%"
```

```
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Alternative name 12]) }, "%") } + "%"
```

```
INSERT INTO ArticleCountry
(CountryName, CountryID, ArticleID)
VALUES ( "", "", "" )
```

```
DELETE FROM ArticleCountry
WHERE (CountryName = "") AND (CountryID = "") AND (ArticleID = "")
```

```
,
```

```
GO
```

```
SET ANSI_NULLS ON
GO
```

```
SET QUOTED_IDENTIFIER ON
GO
```

```
IF NOT EXISTS (SELECT * FROM sys.triggers WHERE object_id =
OBJECT_ID(N'[dbo].[UpdateJournalID]'))
EXEC dbo.sp_executesql @statement = N'
```

```
-- =====
-- Author: <Author,,Name>
-- Create date: <Create Date,,>
-- Description: <Description,,>
-- =====
```

```
CREATE TRIGGER [dbo].[UpdateJournalID]
ON [dbo].[article]
AFTER INSERT
AS
UPDATE article
SET JournalID = substring(article.id, PATINDEX( "%/%", article.id)+1, len(article.id)-
len(right(article.id, PATINDEX( "%/%", article.id)))-12)
```

```
,
```

```
GO
```

```
SET ANSI_NULLS ON
GO
```

```
SET QUOTED_IDENTIFIER ON
GO
```



```

IF NOT EXISTS (SELECT * FROM sys.triggers WHERE object_id =
OBJECT_ID(N'[dbo].[UpdateScientificField]'))
EXEC dbo.sp_executesql @statement = N'CREATE TRIGGER [dbo].[UpdateScientificField]
on [dbo].[article]
AFTER INSERT
AS
UPDATE article
SET scientific_field = Scientific_Field.Scientific_Field
FROM article INNER JOIN Scientific_Field
ON (Scientific_Field.Scientific_Field LIKE "%" + { fn CONCAT({ fn CONCAT("%",
article.scientific_field) }, "%") } + "%")'
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[Publisher]')
AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[Publisher](
[PublisherID] [nvarchar](50) NOT NULL,
[PublisherName] [nvarchar](50) NOT NULL,
CONSTRAINT [PK_Table_1] PRIMARY KEY CLUSTERED
(
[PublisherID] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[CensusData]') AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[CensusData](
[Country / Territory ] [varchar](50) NULL,
[ISO Alpha-2] [varchar](50) NOT NULL,
[ISO Alpha-3] [varchar](50) NOT NULL,
[Population ] [numeric](18, 4) NULL,
[Latitude] [float] NULL,
[Longitude] [float] NULL,
[ISO 3166 Country Name] [nvarchar](255) NULL,
[Alternative name 0] [nvarchar](255) NULL,

```

```

[Alternative name 1] [nvarchar](255) NULL,
[Alternative name 2] [nvarchar](255) NULL,
[Alternative name 3] [nvarchar](255) NULL,
[Alternative name 4] [nvarchar](255) NULL,
[Alternative name 5] [nvarchar](255) NULL,
[Alternative name 6] [nvarchar](255) NULL,
[Alternative name 7] [nvarchar](255) NULL,
[Alternative name 8] [nvarchar](255) NULL,
[Alternative name 9] [nvarchar](255) NULL,
[Alternative name 10] [nvarchar](255) NULL,
[Alternative name 11] [nvarchar](255) NULL,
[Alternative name 12] [nvarchar](255) NULL,
[IncomePerCapita] [money] NULL,
CONSTRAINT [PK_CensusData_1] PRIMARY KEY CLUSTERED
(
    [ISO Alpha-2] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[Scientific_Field]') AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[Scientific_Field](
    [Scientific_Field_Code] [int] IDENTITY(1,1) NOT NULL,
    [Scientific_Field] [varchar](max) NULL,
    CONSTRAINT [PK_Scientific_Field] PRIMARY KEY CLUSTERED
(
    [Scientific_Field_Code] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[ArticleCountry]') AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[ArticleCountry](

```

```

[CountryName] [nvarchar](50) NOT NULL,
[CountryID] [nchar](10) NOT NULL,
[ArticleID] [nchar](50) NOT NULL,
CONSTRAINT [PK_ArticleCountry] PRIMARY KEY CLUSTERED
(
    [CountryID] ASC,
    [ArticleID] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[Journal]')
AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[Journal](
    [JournalID] [nvarchar](50) NOT NULL,
    [JournalName] [nvarchar](50) NOT NULL,
    [PublisherID] [nvarchar](50) NOT NULL,
    CONSTRAINT [PK_Journal] PRIMARY KEY CLUSTERED
(
    [JournalID] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO
IF NOT EXISTS (SELECT * FROM sys.foreign_keys WHERE object_id =
OBJECT_ID(N'[dbo].[FK_Journal_Journal]') AND parent_object_id = OBJECT_ID(N'[dbo].[Journal]'))
ALTER TABLE [dbo].[Journal] WITH CHECK ADD CONSTRAINT [FK_Journal_Journal] FOREIGN
KEY([PublisherID])
REFERENCES [dbo].[Publisher] ([PublisherID])
GO
ALTER TABLE [dbo].[Journal] CHECK CONSTRAINT [FK_Journal_Journal]

```

4.4. Πυροδοτητές (Triggers)

Όπως μπορούμε να παρατηρήσουμε και από το script δημιουργίας της Βάσης, έχουν δημιουργηθεί κάποιοι triggers. Πιο συγκεκριμένα, έχουν δημιουργηθεί 6 triggers έτσι ώστε να αυτοματοποιηθούν κάποιες λειτουργίες στην υπό μελέτη Βάση Δεδομένων. Ας δούμε ενδεικτικά τους 2 σημαντικότερους.

- InsertIntoArticleCountry

```

CREATE TRIGGER InsertIntoArticleCountry
ON article

```

```

AFTER INSERT
AS
INSERT INTO ArticleCountry
SELECT    CensusData.[Country / Territory ], CensusData.[ISO Alpha-2], article.id
FROM      article INNER JOIN CensusData
ON (article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Country / Territory]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[ISO 3166 Country
Name]) }, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 0]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 1]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 2]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 3]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 4]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 5]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 6]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 7]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 8]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 9]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 10]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 11]
}, '%') } + '%'
OR article.affiliation LIKE '%' + { fn CONCAT({ fn CONCAT('%', CensusData.[Alternative name 12]
}, '%') } + '%'

```

```

INSERT INTO ArticleCountry
(CountryName, CountryID, ArticleID)
VALUES ('', '', '')

```

```

DELETE FROM ArticleCountry
WHERE (CountryName = '') AND (CountryID = '') AND (ArticleID = '')

```

Αυτός ίσως να είναι και ο σημαντικότερος trigger, καθώς είναι αυτός ο οποίος «γемίξει» τον πίνακα ArticleCountry ψάχνοντας μέσα στο πεδίο affiliation του πίνακα Article και συσχετίζει το κάθε paper ξεχωριστά με το CountryID. Παράλληλα προβλέπει και την χρήση εναλλακτικών ονομάτων μέσα στο πεδίο affiliation για την ονομασία της κάθε χώρας.

- DeletefromArticleCountry

```

CREATE TRIGGER DeletefromArticleCountry
ON article

```

```
AFTER DELETE  
AS  
DELETE FROM ArticleCountry  
FROM DELETED INNER JOIN  
ArticleCountry ON DELETED.id = ArticleCountry.ArticleID
```

Αυτός ο trigger ακολουθεί την αντίστροφη διαδικασία, δηλαδή αν κάποιο paper διαγραφεί για οποιονδήποτε λόγο από τον πίνακα Article τότε διαγράφει και τις αντίστοιχες εγγραφές από τον πίνακα ArticleCountry.

5. Αυτοματοποίηση Διαδικασίας Συλλογής των Δεδομένων και Εισαγωγής τους στην Βάση Δεδομένων

Έχουμε ήδη δει την λειτουργία του συστήματος μας αναλυτικά και βήμα προς βήμα. Αυτό που μας ενδιαφέρει τώρα είναι να αυτοματοποιήσουμε την όλη αυτή διαδικασία έτσι ώστε να ενημερώνεται η βάση μας χωρίς να παρεμβαίνει «ανθρώπινο χέρι».

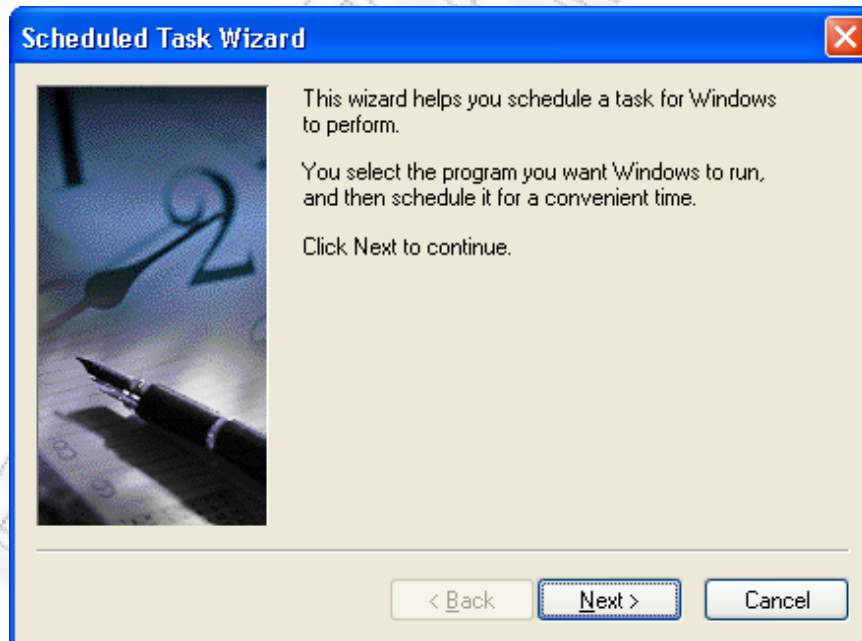
Αυτό θα το καταφέρουμε με την βοήθεια ενός batch file το οποίο θα εκτελεί το Web Harvest παίρνοντας σαν παράμετρο το configuration file που επιθυμούμε και στη συνέχεια, αφού έχει εκτελεσθεί και έχει παραχθεί το xml αρχείο με τα δεδομένα μας, εκτελείται το vbs script με το οποίο κάνουμε μαζική είσοδο των παραγόμενων δεδομένων. Τέλος, όσον αφορά την βάση, εκεί αναλαμβάνουν οι triggers οι οποίοι πυροδοτούνται με την εισαγωγή των δεδομένων σε αυτή.

Πιο συγκεκριμένα ας δούμε τον κώδικα του batch file

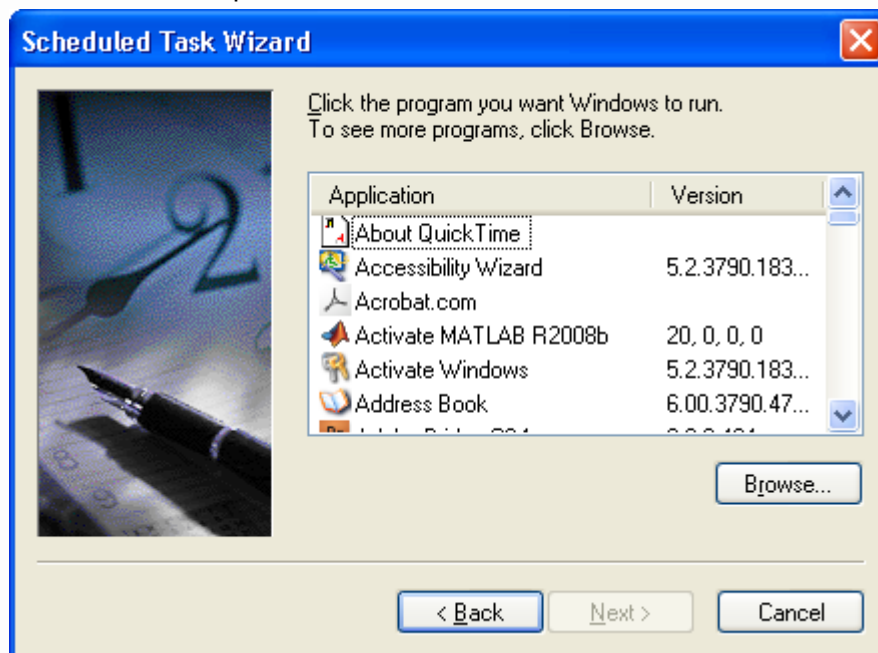
```
@echo off
cd C:\webharvest1-project
java -jar webharvest_all_1.jar config=C:\webharvest1-project\test.xml workdir=C:\webharvest1-project
cd C:\webharvest1-project
cscript Insertarticles.vbs
```

Τέλος για να αυτοματοποιηθεί πλήρως η όλη αυτή διαδικασία θα δημιουργήσουμε μια προγραμματισμένη εργασία, η οποία θα τρέχει ανά τακτά χρονικά διαστήματα. Αυτό μπορεί να γίνει με τα εξής βήματα,

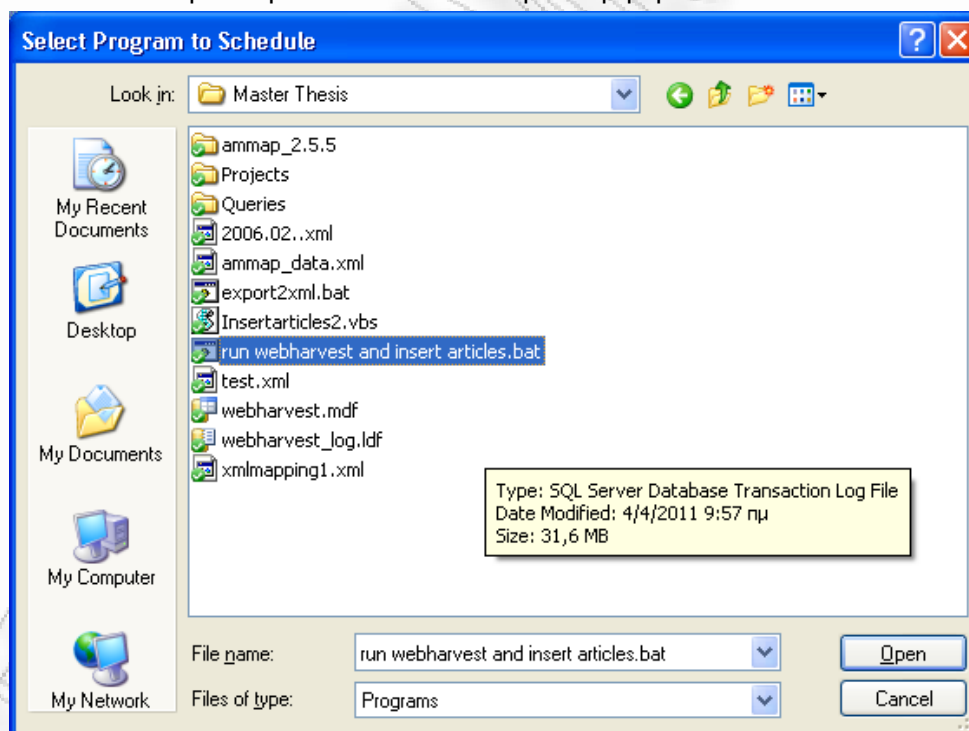
1. Start → Settings → Control Panel
2. Scheduled Tasks
3. Add Scheduled Task
4. Πατάμε Next



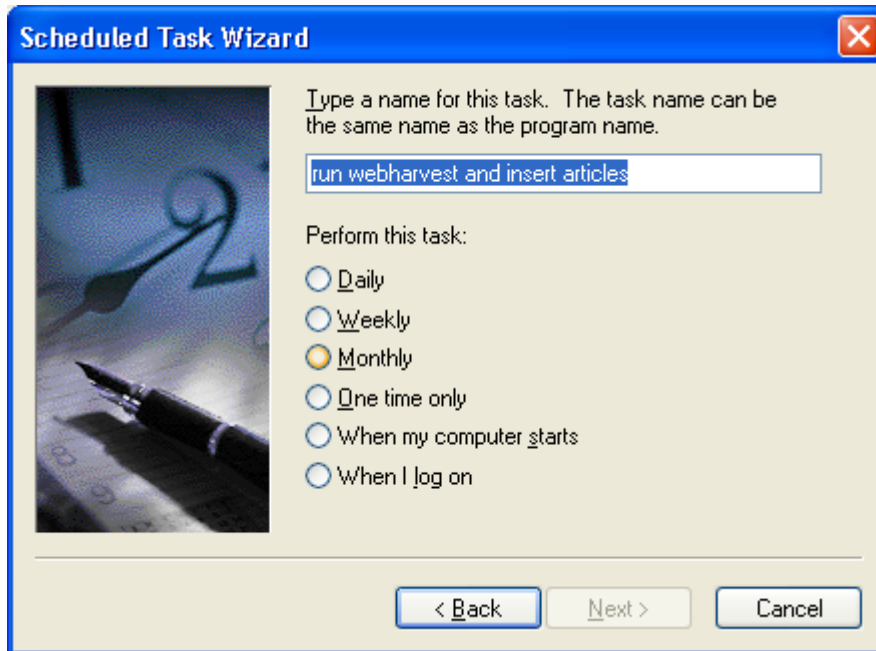
5. Πατάμε Browse



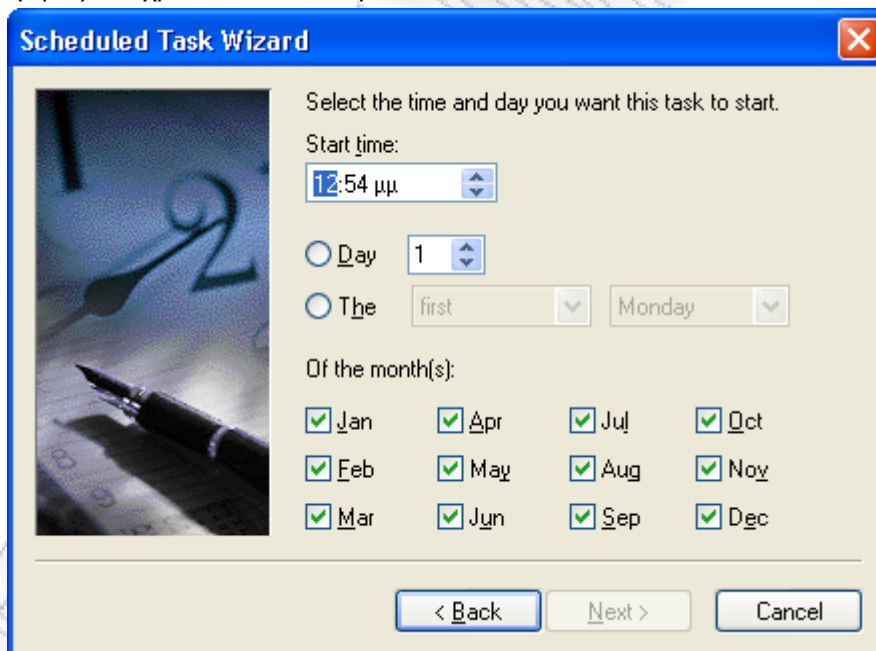
6. Βρίσκουμε το batch file που προαναφέραμε



7. Δίνουμε ένα όνομα στην προγραμματισμένη εργασία και επιλέγουμε κάθε πότε θέλουμε να εκτελείτε



8. Επιλέγουμε την ώρα που θέλουμε να αρχίσει η εκτέλεση, την μέρα και τους μήνες του χρόνου που θέλουμε να εκτελεστεί



9. Εισάγουμε Username και Password του χρήστη που θα τρέχει την προγραμματισμένη εργασία



Scheduled Task Wizard

Enter the name and password of a user. The task will run as if it were started by that user.

Enter the user name: PANOSBP\Admin

Enter the password: ●●●●●●●●●●●●

Confirm password: ●●●●●●●●●●●●

If a password is not entered, scheduled tasks might not run.

< Back Next > Cancel

10. Και τέλος πατάμε Finish



Scheduled Task Wizard

You have successfully scheduled the following task:
run webharvest and insert articles

Windows will perform this task:
At 12:54 μμ on day 1 of every month, starting
1/6/2011

Open advanced properties for this task when I click Finish.

Click Finish to add this task to your Windows schedule.

< Back Finish Cancel

6. Απεικόνιση των Δεδομένων σε Χάρτη

Έχοντας ήδη αναλύσει τον τρόπο συλλογής των δεδομένων μας, την αποθήκευση και επεξεργασία τους σε μια Βάση και την αυτοματοποίηση της όλης διαδικασίας θα πρέπει να περάσουμε και στο τελευταίο στάδιο της εφαρμογής που έχουμε αναπτύξει το οποίο είναι η απεικόνιση των δεδομένων μας σε ένα χάρτη και η διεπαφή με τον χρήστη. Το κομμάτι αυτό είναι πολύ σημαντικό κυρίως για δύο λόγους:

1. είναι αυτό που δίνει προστιθέμενη αξία στην εφαρμογή μας, καθώς όχι μόνο έχουν μετατραπεί τα δεδομένα μας σε πληροφορία αλλά έχουμε και μια οπτική απεικόνιση αυτών
2. είναι το μόνο σημείο της εφαρμογής που αντιλαμβάνεται ο χρήστης, επομένως, όσον αφορά τον χρήστη, όλη εφαρμογή είναι αυτό το κομμάτι

Επομένως, έπρεπε να δοθεί ιδιαίτερη προσοχή όσον αφορά την λειτουργικότητα, την εμφάνιση και την διεπαφή με τον χρήστη.

6.1. amMap

Μία από τις σημαντικότερες αποφάσεις ήταν η επιλογή της τεχνολογίας και του εργαλείου που θα χρησιμοποιούσαμε. Όσον αφορά την τεχνολογία επιλέχθηκε η χρήση ενός Flash map. Οι κύριοι λόγοι που οδήγησαν σε αυτή την επιλογή είναι:

1. το εντυπωσιακό αισθητικό αποτέλεσμα που δίνουν οι Flash Maps
2. και η πολύ καλή διεπαφή με τον χρήστη

Αναφορικά την επιλογή του εργαλείου που θα χρησιμοποιούσαμε, καταλήξαμε στον amMap. Το amMap είναι ένα εργαλείο για την δημιουργία διαδραστικών Flash maps το οποίο σου δίνει την δυνατότητα να χρησιμοποιήσεις πολλές λειτουργίες, όπως:

1. Drill down
2. Zoomifyer
3. Timeline map
4. Charts
5. Areas (countries, regions)
6. Movies (cities, icons, images)
7. Labels (text)
8. Lines
9. Snap points

Το μεγάλο μειονέκτημα των Flash maps και παρεμπιπτόντως και του amMap είναι η αδυναμία που υπάρχει στην απευθείας σύνδεση τους με κάποια Βάση Δεδομένων. Αυτό που συνήθως συμβαίνει είναι η παροχή των δεδομένων από κάποιο εξωτερικό αρχείο. Αυτό συμβαίνει και στην περίπτωση του amMap. Τα δεδομένα προέρχονται από κάποιο xml αρχείο που έχει μια συγκεκριμένη μορφή. Στην πιο απλή περίπτωση αυτό το xml αρχείο είναι της παρακάτω μορφής

```
<map map_file="maps/world.swf" zoom="100">
<areas>
<area title="Argentina" mc_name="AR" value="64"/>
<area title="Austria" mc_name="AT" value="184"/>
<area title="Barbados" mc_name="BB" value="1"/>
<area title="Brunei" mc_name="BN" value="4"/>
<area title="Bahrain" mc_name="BH" value="3"/>
<area title="Algeria" mc_name="DZ" value="19"/>
<area title="Cameroon" mc_name="CM" value="4"/>
<area title="Belarus" mc_name="BY" value="38"/>
```

```
<area title="Australia" mc_name="AU" value="491"/>
<area title="Chad" mc_name="TD" value="1"/>
<area title="Croatia" mc_name="HR" value="14"/>
<area title="Cyprus" mc_name="CY" value="20"/>
<area title="Belize" mc_name="BZ" value="1"/>
<area title="Armenia" mc_name="AM" value="1"/>
<area title="Czech Republic" mc_name="CZ" value="127"/>
<area title="Denmark" mc_name="DK" value="128"/>
<area title="Finland" mc_name="FI" value="161"/>
<area title="Benin" mc_name="BJ" value="2"/>
<area title="Bosnia and Herzegovina" mc_name="BA" value="1"/>
<area title="Botswana" mc_name="BW" value="2"/>
<area title="China" mc_name="CN" value="2297"/>
<area title="Colombia" mc_name="CO" value="16"/>
<area title="Costa Rica" mc_name="CR" value="3"/>
<area title="France" mc_name="FR" value="955"/>
<area title="Guinea-Bissau" mc_name="GW" value="1"/>
<area title="Jordan" mc_name="JO" value="34"/>
<area title="Luxembourg" mc_name="LU" value="7"/>
<area title="Macau" mc_name="MO" value="15"/>
<area title="Mali" mc_name="ML" value="1"/>
<area title="New Zealand" mc_name="NZ" value="94"/>
<area title="Norway" mc_name="NO" value="119"/>
<area title="Oman" mc_name="OM" value="15"/>
<area title="Portugal" mc_name="PT" value="185"/>
<area title="Puerto Rico" mc_name="PR" value="6"/>
<area title="Qatar" mc_name="QA" value="13"/>
<area title="Singapore" mc_name="SG" value="281"/>
<area title="Spain" mc_name="ES" value="1217"/>
<area title="Sweden" mc_name="SE" value="159"/>
<area title="Cuba" mc_name="CU" value="9"/>
<area title="Syria" mc_name="SY" value="1"/>
<area title="Turkey" mc_name="TR" value="831"/>
<area title="Uruguay" mc_name="UY" value="10"/>
<area title="Egypt" mc_name="EG" value="108"/>
<area title="Guam" mc_name="GU" value="2"/>
<area title="Zimbabwe" mc_name="ZW" value="4"/>
<area title="Iceland" mc_name="IS" value="10"/>
<area title="Brazil" mc_name="BR" value="328"/>
<area title="Bulgaria" mc_name="BG" value="13"/>
<area title="India" mc_name="IN" value="584"/>
<area title="Italy" mc_name="IT" value="811"/>
<area title="Jamaica" mc_name="JM" value="4"/>
<area title="Kuwait" mc_name="KW" value="32"/>
<area title="Canada" mc_name="CA" value="1159"/>
<area title="Germany" mc_name="DE" value="738"/>
```

```
<area title="Lithuania" mc_name="LT" value="11"/>
<area title="Guinea" mc_name="GN" value="1"/>
<area title="Malaysia" mc_name="MY" value="72"/>
<area title="Hong Kong" mc_name="HK" value="698"/>
<area title="Indonesia" mc_name="ID" value="17"/>
<area title="Iran" mc_name="IR" value="361"/>
<area title="Ireland" mc_name="IE" value="68"/>
<area title="Latvia" mc_name="LV" value="9"/>
<area title="Montenegro" mc_name="ME" value="33"/>
<area title="Lebanon" mc_name="LB" value="20"/>
<area title="Nigeria" mc_name="NG" value="6"/>
<area title="Lesotho" mc_name="LS" value="1"/>
<area title="Pakistan" mc_name="PK" value="25"/>
<area title="Papua New Guinea" mc_name="PG" value="1"/>
<area title="Mexico" mc_name="MX" value="170"/>
<area title="Poland" mc_name="PL" value="256"/>
<area title="Romania" mc_name="RO" value="76"/>
<area title="Moldova" mc_name="MD" value="1"/>
<area title="Russia" mc_name="RU" value="86"/>
<area title="Morocco" mc_name="MA" value="15"/>
<area title="Saudi Arabia" mc_name="SA" value="67"/>
<area title="Slovakia" mc_name="SK" value="57"/>
<area title="Netherlands" mc_name="NL" value="514"/>
<area title="Slovenia" mc_name="SI" value="28"/>
<area title="Serbia" mc_name="RS" value="68"/>
<area title="South Korea" mc_name="KR" value="306"/>
<area title="Switzerland" mc_name="CH" value="160"/>
<area title="Trinidad and Tobago" mc_name="TT" value="1"/>
<area title="Tunisia" mc_name="TN" value="56"/>
<area title="United Kingdom" mc_name="GB" value="269"/>
<area title="Taiwan" mc_name="TW" value="2344"/>
<area title="Tanzania" mc_name="TZ" value="1"/>
<area title="Vietnam" mc_name="VN" value="56"/>
<area title="Thailand" mc_name="TH" value="62"/>
<area title="Ukraine" mc_name="UA" value="15"/>
<area title="Azerbaijan" mc_name="AZ" value="6"/>
<area title="Bangladesh" mc_name="BD" value="2"/>
<area title="Belgium" mc_name="BE" value="451"/>
<area title="Chile" mc_name="CL" value="97"/>
<area title="Ecuador" mc_name="EC" value="4"/>
<area title="Estonia" mc_name="EE" value="3"/>
<area title="Fiji" mc_name="FJ" value="1"/>
<area title="Georgia" mc_name="GE" value="187"/>
<area title="Greece" mc_name="GR" value="355"/>
<area title="Hungary" mc_name="HU" value="94"/>
<area title="Israel" mc_name="IL" value="232"/>
```

```
<area title="Japan" mc_name="JP" value="535"/>
<area title="Jersey" mc_name="JE" value="65"/>
<area title="Kazakhstan" mc_name="KZ" value="4"/>
<area title="Malawi" mc_name="MW" value="1"/>
<area title="Malta" mc_name="MT" value="2"/>
<area title="Mongolia" mc_name="MN" value="3"/>
<area title="Niger" mc_name="NE" value="6"/>
<area title="Palestine" mc_name="PS" value="1"/>
<area title="Peru" mc_name="PE" value="19"/>
<area title="Philippines" mc_name="PH" value="3"/>
<area title="Senegal" mc_name="SN" value="1"/>
<area title="South Africa" mc_name="ZA" value="82"/>
<area title="United Arab Emirates" mc_name="AE" value="32"/>
<area title="United States" mc_name="US" value="4421"/>
<area title="Uzbekistan" mc_name="UZ" value="1"/>
<area title="Venezuela" mc_name="VE" value="13"/>
<area title="Yemen" mc_name="YE" value="1"/>
<area title="Zambia" mc_name="ZM" value="1"/>
</areas>
</map>
```

Αυτός ο τρόπος παροχής των δεδομένων έχει πολλά προβλήματα, καθώς κάθε φορά που θα ενημερώνεται η βάση θα πρέπει να παράγεται ένα τέτοιο αρχείο. Επίσης στην περίπτωση που θα θέλαμε να κάνουμε μία αναζήτηση πάνω στην Βάση με κάποιο open text πεδίο αυτό θα ήταν αδύνατο. Από τα παραπάνω μπορούμε να συμπεράνουμε πως θα ήταν πολύ σημαντικό αν μπορούσαμε με κάποιο τρόπο να κάνουμε το amMap να συνδεθεί live με την Βάση μας.

6.2. Live Σύνδεση amMap με την Βάση Δεδομένων

Ο τρόπος με τον οποίο έγινε εφικτή η σύνδεση του amMap με την Βάση Δεδομένων είναι με την βοήθεια της php. Χρησιμοποιώντας php έγινε εφικτή η δυναμική δημιουργία, από την Βάση μας, δεδομένων της μορφής που δέχεται το amMap. Πιο συγκεκριμένα, ας το δούμε αυτό στην απλούστερη μορφή του, με τον Συνολικό Αριθμό Papers ανά Χώρα.

Τα δεδομένα θα πρέπει να έχουν την μορφή που έχουμε ήδη δει μόνο που στο πεδίο value θα έχουμε τον Συνολικό Αριθμό Papers ανά Χώρα. Επομένως, εμείς θέλουμε να δημιουργήσουμε δεδομένα της μορφής αυτής δυναμικά από την Βάση μας. Έτσι έχουμε,

```
<?php
include('connect2db.php');
if (!$conn)
    {exit("Connection Failed: " . $conn);}
$sql="SELECT CountryName, CountryID, COUNT(ArticleID) AS NumberOfArticles
FROM ArticleCountry
GROUP BY CountryName, CountryID";
$rs=odbc_exec($conn,$sql);
if (!$rs)
    {exit("Error in SQL");}
$myString1 = '<map map_file="maps/world.swf" zoom="100"> <areas>';
echo $myString1;
while (odbc_fetch_row($rs))
    {
    $myString2 = '<area title="';
    $myString3 = 'mc_name="';
    $myString4 = 'value="';
    $myString5 = "' ";
    $myString6 = "'/>';
    $CountryName=odbc_result($rs,"CountryName");
    $CountryID=odbc_result($rs,"CountryID");
    $NumberOfArticles=odbc_result($rs,"NumberOfArticles");

    echo $myString2.trim($CountryName).$myString5;
    echo $myString3.trim($CountryID).$myString5;
    echo $myString4.$NumberOfArticles.$myString6;
    }
$myString7='</areas>
</map>';
echo $myString7;
odbc_close($conn);
?>
```

Όπου,

1. Συνδεόμαστε με την Βάση

2. Δημιουργούμε, εκτελούμε και αποθηκεύουμε τα αποτελέσματα του sql ερωτήματος που θέλουμε
3. Στη συνέχεια δημιουργούμε το κομμάτι που εμφανίζεται μόνο μία φορά στο αρχείο δεδομένων, το `<map map_file="maps/world.swf" zoom="100"> <areas>`
4. Με ένα while loop δημιουργούμε τα δεδομένα που θέλουμε για κάθε εγγραφή των αποτελεσμάτων του sql ερωτήματος.
5. Αφού τελειώσει το loop δημιουργούμε και τα closing tags `</areas> </map>`
6. Τέλος, τερματίζουμε τη σύνδεση με τη Βάση

Ο κώδικας αυτός θα μας δώσει το παρακάτω αποτέλεσμα, που είναι και το ζητούμενο μας:

```
<map map_file="maps/world.swf" zoom="100"> <areas><area title="Malaysia" mc_name="MY" value="79"/><area title="Malta" mc_name="MT" value="2"/><area title="Norway" mc_name="NO" value="132"/><area title="Barbados" mc_name="BB" value="1"/><area title="Nigeria" mc_name="NG" value="6"/><area title="Puerto Rico" mc_name="PR" value="6"/><area title="Australia" mc_name="AU" value="560"/><area title="Moldova" mc_name="MD" value="1"/><area title="Hungary" mc_name="HU" value="107"/><area title="Oman" mc_name="OM" value="15"/><area title="Italy" mc_name="IT" value="899"/><area title="Ireland" mc_name="IE" value="76"/><area title="Chile" mc_name="CL" value="106"/><area title="Egypt" mc_name="EG" value="116"/><area title="Pakistan" mc_name="PK" value="27"/><area title="Singapore" mc_name="SG" value="318"/><area title="United Arab Emirates" mc_name="AE" value="37"/><area title="New Zealand" mc_name="NZ" value="106"/><area title="Guam" mc_name="GU" value="2"/><area title="Lebanon" mc_name="LB" value="24"/><area title="Vietnam" mc_name="VN" value="65"/><area title="Jordan" mc_name="JO" value="35"/><area title="Austria" mc_name="AT" value="207"/><area title="Ukraine" mc_name="UA" value="15"/><area title="Switzerland" mc_name="CH" value="183"/><area title="Iran" mc_name="IR" value="417"/><area title="Mauritius" mc_name="MU" value="1"/><area title="Bahrain" mc_name="BH" value="5"/><area title="Tanzania" mc_name="TZ" value="1"/><area title="Georgia" mc_name="GE" value="225"/><area title="Guinea" mc_name="GN" value="1"/><area title="Thailand" mc_name="TH" value="68"/><area title="Germany" mc_name="DE" value="829"/><area title="Russia" mc_name="RU" value="92"/><area title="United Kingdom" mc_name="GB" value="291"/><area title="Serbia" mc_name="RS" value="72"/><area title="Zambia" mc_name="ZM" value="1"/><area title="Luxembourg" mc_name="LU" value="10"/><area title="Tunisia" mc_name="TN" value="68"/><area title="Belgium" mc_name="BE" value="502"/><area title="Netherlands" mc_name="NL" value="599"/><area title="Turkey" mc_name="TR" value="910"/><area title="Senegal" mc_name="SN" value="1"/><area title="Japan" mc_name="JP" value="574"/><area title="Greece" mc_name="GR" value="379"/><area title="Sweden" mc_name="SE" value="173"/><area title="Romania" mc_name="RO" value="83"/><area title="Botswana" mc_name="BW" value="2"/><area title="Uzbekistan" mc_name="UZ" value="1"/><area title="Guinea-Bissau" mc_name="GW" value="1"/><area title="Malawi" mc_name="MW" value="1"/><area title="Portugal" mc_name="PT" value="211"/><area title="Cyprus" mc_name="CY" value="25"/><area title="Indonesia" mc_name="ID" value="19"/><area title="Trinidad and Tobago" mc_name="TT" value="1"/><area title="Zimbabwe" mc_name="ZW" value="4"/><area title="France" mc_name="FR" value="1055"/><area title="Bosnia and Herzegovina" mc_name="BA" value="1"/><area title="Brunei" mc_name="BN" value="5"/><area title="Fiji" mc_name="FJ" value="2"/><area title="Czech Republic" mc_name="CZ" value="135"/><area title="Peru" mc_name="PE" value="21"/><area title="Kuwait" mc_name="KW" value="37"/><area title="Mexico" mc_name="MX" value="194"/><area title="Saudi Arabia" mc_name="SA" value="75"/><area title="South Korea" mc_name="KR" value="333"/><area title="Costa Rica" mc_name="CR" value="3"/><area title="Slovakia" mc_name="SK" value="62"/><area title="Denmark" mc_name="DK" value="142"/><area title="Lithuania" mc_name="LT" value="14"/><area title="Brazil" mc_name="BR" value="376"/><area title="Yemen" mc_name="YE" value="1"/><area title="United States" mc_name="US" value="5015"/><area title="Belarus" mc_name="BY" value="44"/><area title="India" mc_name="IN" value="639"/><area title="Jamaica" mc_name="JM" value="4"/><area title="Estonia" mc_name="EE" value="4"/><area
```

```

title="Canada" mc_name="CA" value="1275"/><area title="Niger" mc_name="NE"
value="6"/><area title="Sri Lanka" mc_name="LK" value="1"/><area title="Montenegro"
mc_name="ME" value="33"/><area title="Finland" mc_name="FI" value="183"/><area title="Hong
Kong" mc_name="HK" value="793"/><area title="Qatar" mc_name="QA" value="13"/><area
title="Croatia" mc_name="HR" value="16"/><area title="Latvia" mc_name="LV" value="10"/><area
title="Papua New Guinea" mc_name="PG" value="1"/><area title="Morocco" mc_name="MA"
value="15"/><area title="Mongolia" mc_name="MN" value="3"/><area title="Benin" mc_name="BJ"
value="2"/><area title="Iceland" mc_name="IS" value="12"/><area title="Kazakhstan"
mc_name="KZ" value="4"/><area title="Israel" mc_name="IL" value="252"/><area title="Spain"
mc_name="ES" value="1379"/><area title="Uruguay" mc_name="UY" value="12"/><area
title="Syria" mc_name="SY" value="1"/><area title="Ecuador" mc_name="EC" value="5"/><area
title="Philippines" mc_name="PH" value="3"/><area title="Cuba" mc_name="CU"
value="13"/><area title="Belize" mc_name="BZ" value="1"/><area title="Armenia" mc_name="AM"
value="1"/><area title="Colombia" mc_name="CO" value="18"/><area title="Palestine"
mc_name="PS" value="1"/><area title="Jersey" mc_name="JE" value="74"/><area title="Bulgaria"
mc_name="BG" value="14"/><area title="Poland" mc_name="PL" value="282"/><area title="Mali"
mc_name="ML" value="1"/><area title="Macau" mc_name="MO" value="16"/><area title="South
Africa" mc_name="ZA" value="87"/><area title="Chad" mc_name="TD" value="1"/><area
title="Algeria" mc_name="DZ" value="24"/><area title="China" mc_name="CN"
value="2506"/><area title="Venezuela" mc_name="VE" value="14"/><area title="Slovenia"
mc_name="SI" value="34"/><area title="Lesotho" mc_name="LS" value="1"/><area
title="Bangladesh" mc_name="BD" value="4"/><area title="Argentina" mc_name="AR"
value="71"/><area title="Cameroon" mc_name="CM" value="4"/><area title="Taiwan"
mc_name="TW" value="2579"/><area title="Azerbaijan" mc_name="AZ"
value="6"/></areas></map>

```

6.3. Παρουσίαση Web Εφαρμογής

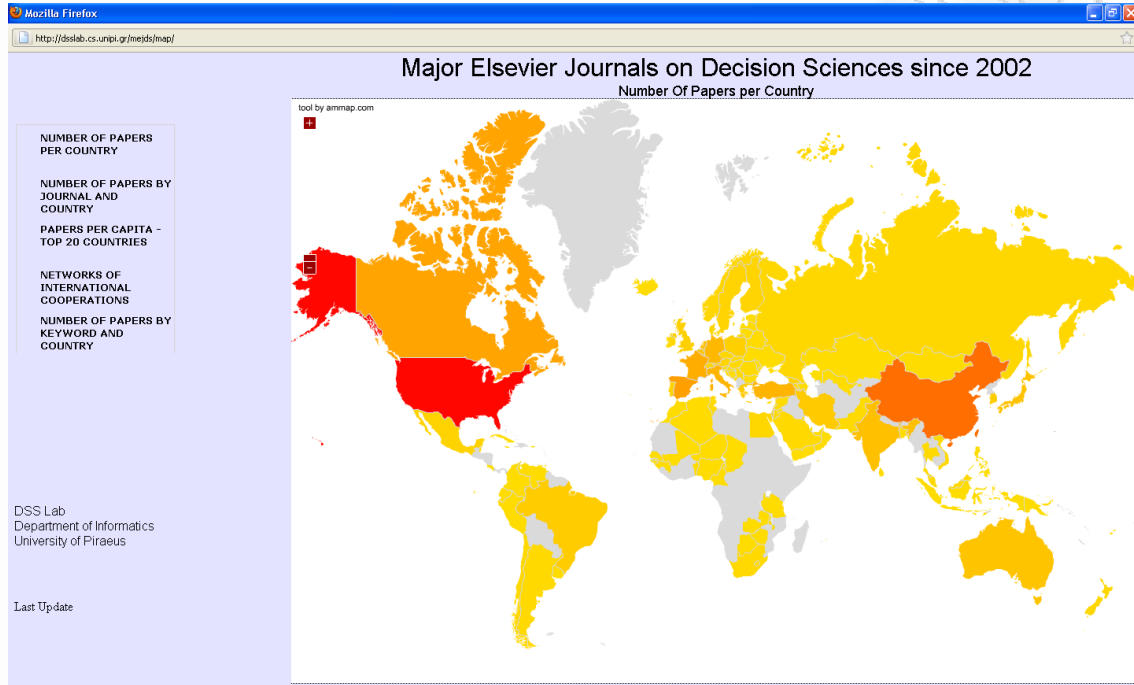
Έχοντας αναλύσει τα επιμέρους τμήματα της εφαρμογής μπορούμε να προχωρήσουμε στην παρουσίαση της τελικής web εφαρμογής. Σε αυτή μας δίνεται η δυνατότητα για την οπτικοποίηση 5 μεγεθών:

1. Αριθμό Paper ανά Χώρα
2. Αριθμό Paper ανά Επιστημονικό Περιοδικό και Χώρα
3. ατά Κεφαλήν Αριθμός Papers – Κορυφαίες 20 Χώρες
4. Δίκτυα Διεθνών Συνεργασιών
5. Αριθμό Paper ανά Λέξη Κλειδί και Χώρα

Παρακάτω θα αναλύσουμε το καθένα ξεχωριστά.

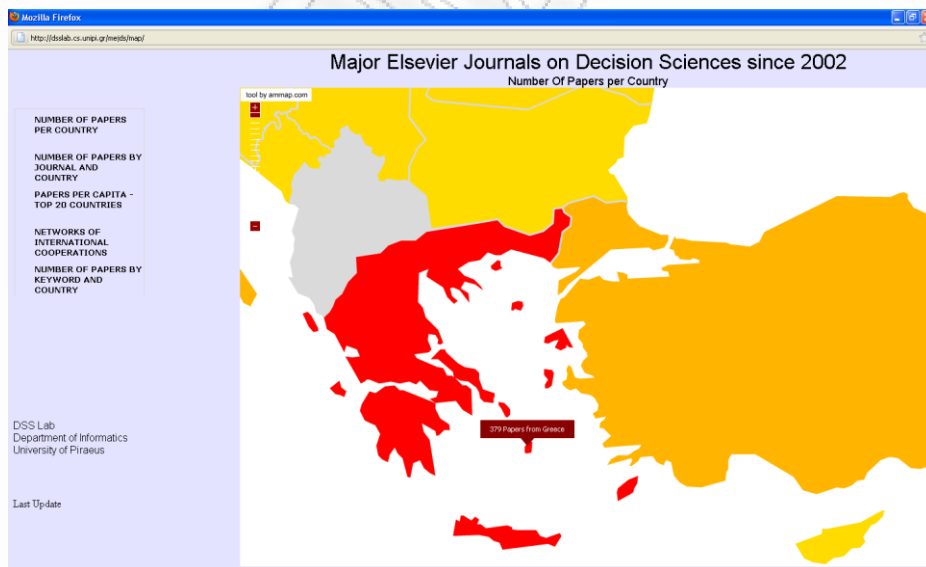
6.3.1. Αριθμός Paper ανά Χώρα

Πρόκειται για το βασικό μέγεθος γι' αυτό και εμφανίζεται στην αρχική σελίδα. Ας ρίξουμε όμως μια ματιά.



Εικόνα 4: Αριθμός Paper ανά Χώρα

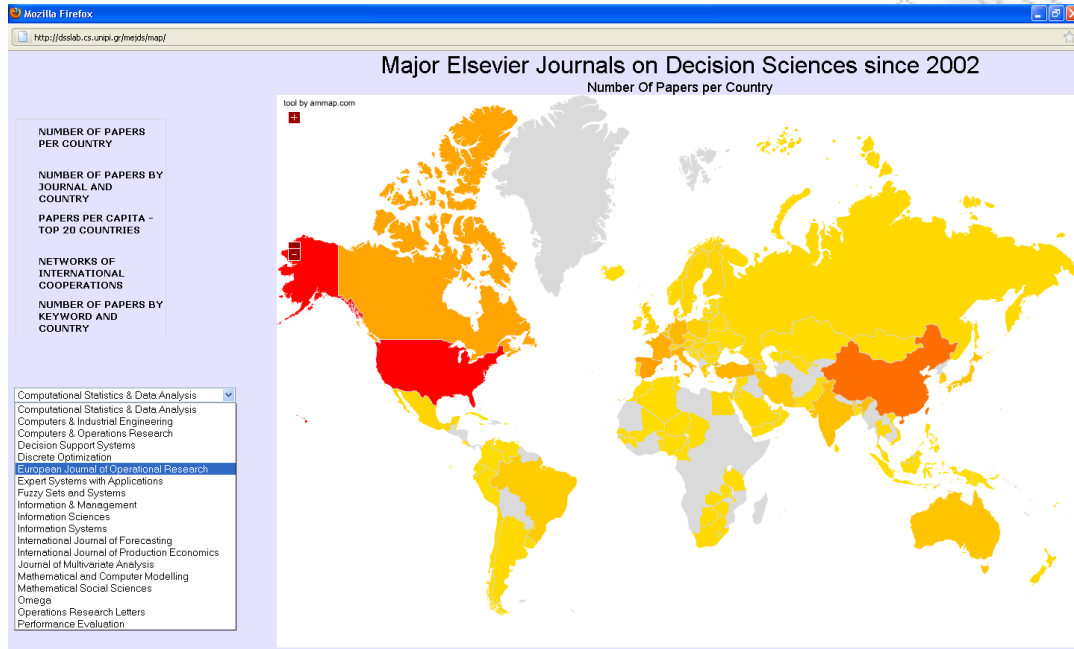
Παρατηρούμε την χρωματική διαρρύθμιση, ανάλογα με τον αριθμό των papers (όσο περισσότερα τόσο πιο σκούρο το χρώμα). Επίσης έχουμε την δυνατότητα zoom και περνώντας με τον κέρσορα πάνω από κάποια χώρα μας εμφανίζεται μία ετικέτα με τον αριθμό των paper από την χώρα αυτή.



Εικόνα 5: Αριθμός Paper Ελλάδας

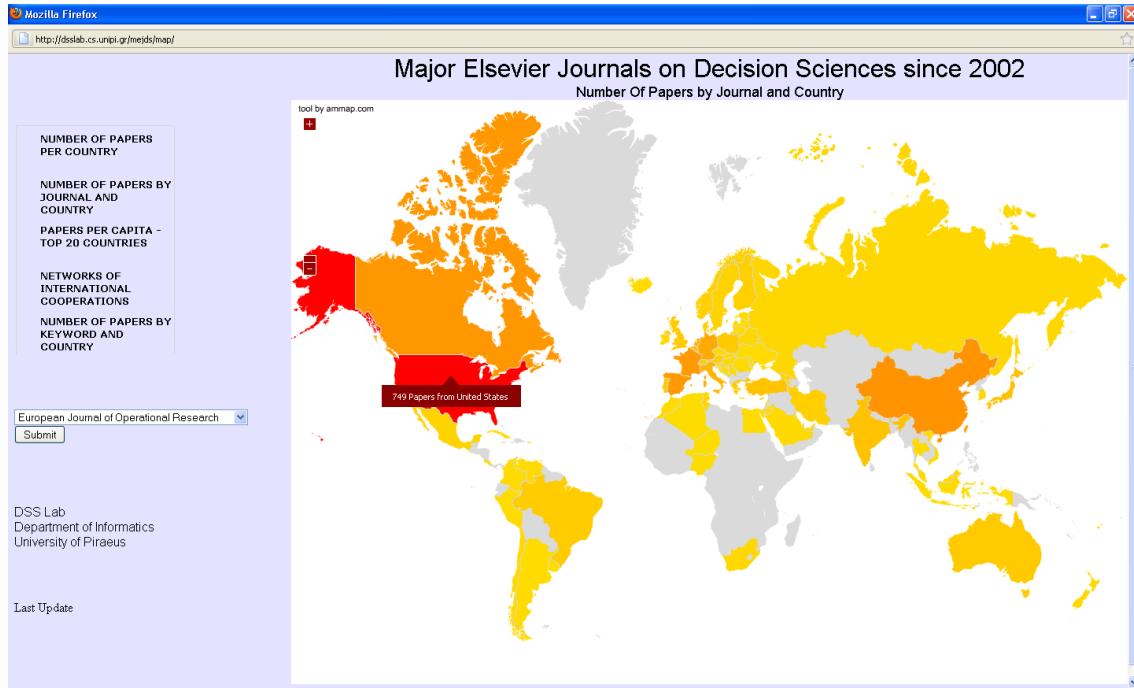
6.3.2. Αριθμός Paper ανά Επιστημονικό Περιοδικό και Χώρα

Πρόκειται για ένα πολύ ενδιαφέρον μέγεθος το οποίο έχει οπτικοποιηθεί. Εδώ επιλέγουμε το επιστημονικό περιοδικό που μας ενδιαφέρει από μία drop down list. Η λίστα αυτή δημιουργείται δυναμικά από την Βάση Δεδομένων αλλά δεν θα παραθέσουμε σε αυτό το σημείο τον κώδικα. Αν ενδιαφέρει πάντως οποιονδήποτε ο πλήρης κώδικας της εφαρμογής βρίσκεται στο Παράρτημα.



Εικόνα 6: Αριθμός Paper ανά Επιστημονικό Περιοδικό και Χώρα - Επιλογή Επιστημονικού Περιοδικού

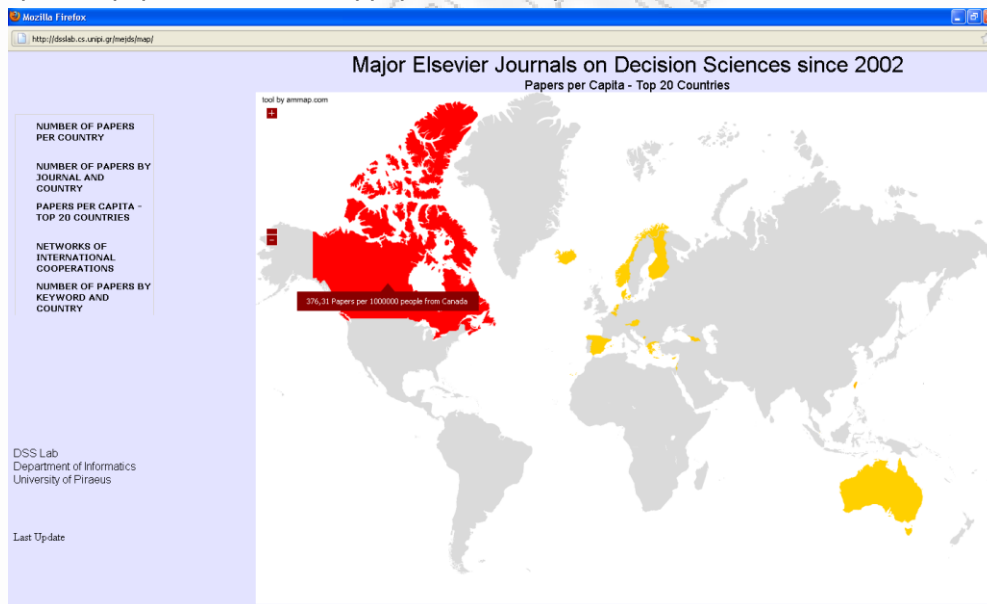
Αφού επιλέξουμε το περιοδικό που μας ενδιαφέρει και πατήσουμε submit έχουμε



Εικόνα 7: Αριθμός Paper ανά Επιστημονικό Περιοδικό και Χώρα

6.3.3. Κατά Κεφαλήν Αριθμός Paper – Κορυφαίες Χώρες

Πρόκειται για ένα αρκετά ενδιαφέρον μέγεθος το οποίο έχει οπτικοποιηθεί και το οποίο μας δίνει τον αριθμό των paper ανά ένα εκατομμύριο κατοίκους.

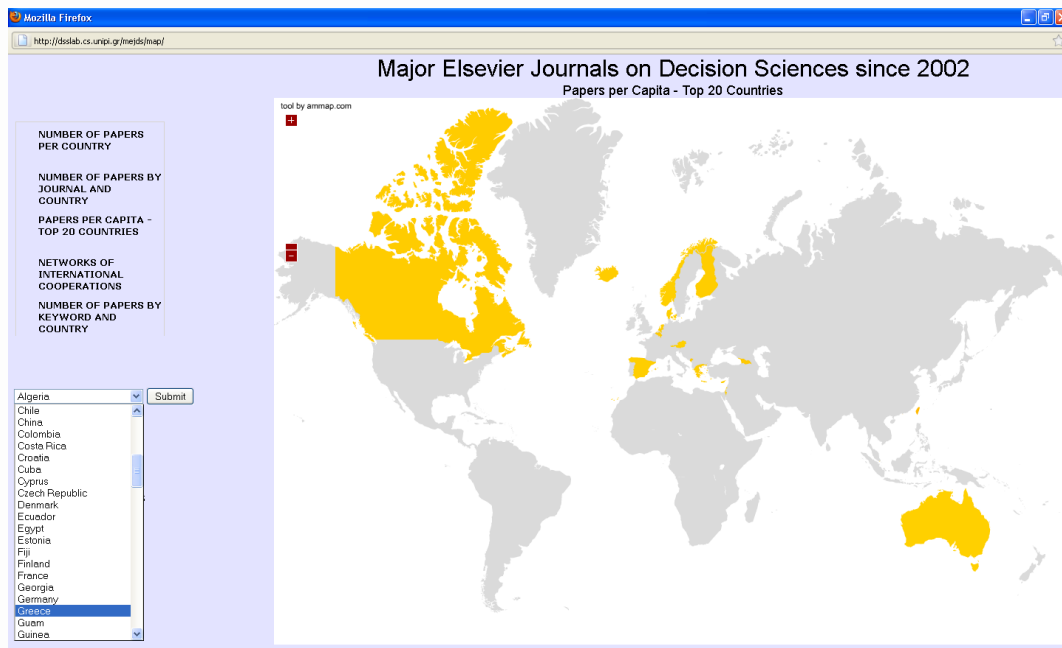


Εικόνα 8 Κατά Κεφαλήν Αριθμός Paper – Κορυφαίες Χώρες

6.3.4. Δίκτυα Διεθνών Συνεργασιών

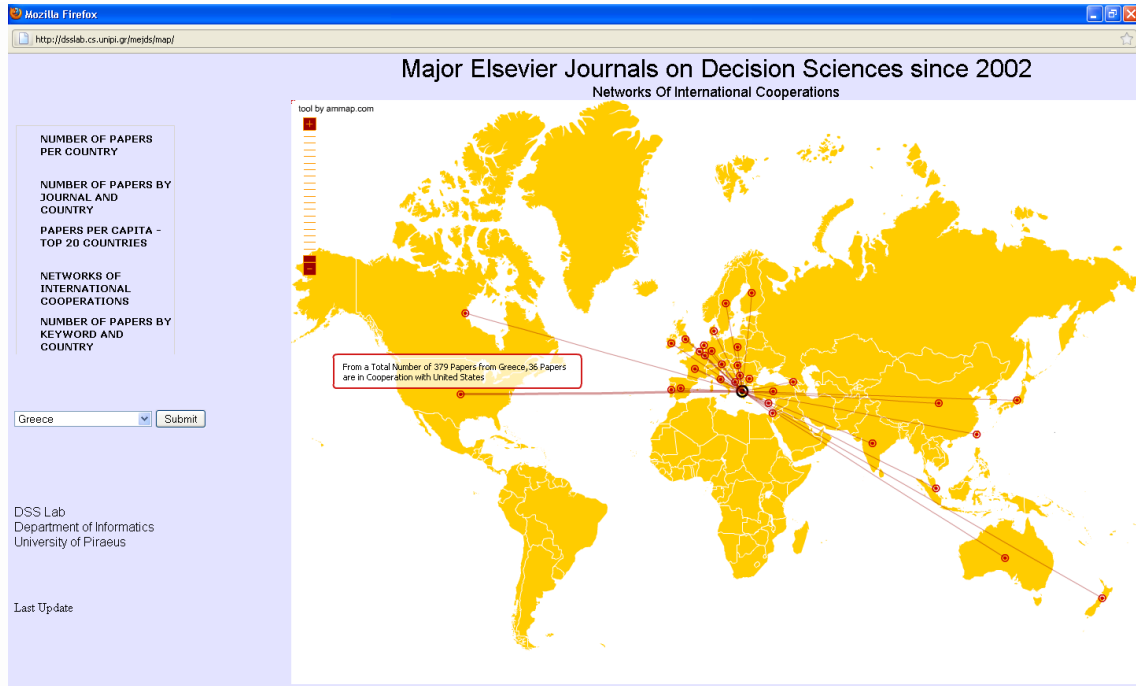
Ίσως το πιο ενδιαφέρον μέγεθος και από την πλευρά της πληροφορίας που μας δίνει αλλά και από την πλευρά της οπτικοποίησης αυτής καθεαυτής. Αυτό το μέγεθος μας παρέχει την πληροφορία του αριθμού των papers, μιας χώρας που θα επιλέξουμε, που έχουν γίνει σε συνεργασία με άλλες χώρες.

Και εδώ επιλέγουμε την χώρα που μας ενδιαφέρει από μία drop down list η οποία δημιουργείται δυναμικά από την Βάση Δεδομένων.



Εικόνα 9: Δίκτυα Διεθνών Συνεργασιών - Επιλογή Χώρας

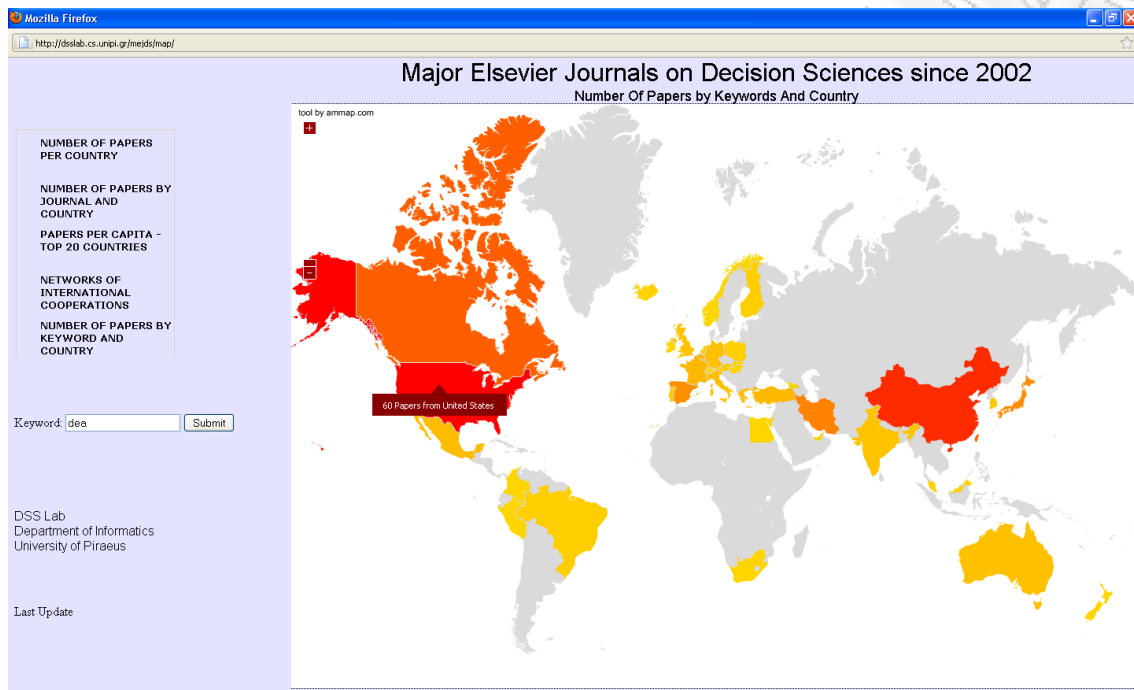
Αφού επιλέξουμε την χώρα που μας ενδιαφέρει και πατήσουμε submit έχουμε



Εικόνα 10: Δίκτυο Διεθνών Συνεργασιών

6.3.5. Αριθμός Paper ανά Λέξη Κλειδί και Χώρα

Πρόκειται για ένα πολύ ενδιαφέρον μέγεθος το οποίο έχει οπτικοποιηθεί. Εδώ έχουμε το μοναδικό open text πεδίο στο οποίο μπορείς να πληκτρολογήσεις κάποια λέξη κλειδί και πατώντας submit να οπτικοποιήσεις στον χάρτη τον αριθμό των paper ανά χώρα που περιέχουν αυτή την λέξη κλειδί. Για παράδειγμα ας πληκτρολογήσουμε "dea" και πατήσουμε submit.



Εικόνα 11: Αριθμός Paper ανά Λέξη Κλειδί και Χώρα

7. Συμπεράσματα - Μελλοντική εργασία

Λαμβάνοντας υπ' όψιν όλα τα παραπάνω μπορούμε να καταλήξουμε στο ότι πρόκειται για ένα πλήρως αυτοματοποιημένο σύστημα άντλησης δεδομένων από το Διαδίκτυο και πιο συγκεκριμένα του εκδοτικού οίκου Elsevier, επεξεργασίας των δεδομένων αυτών και μετατροπής τους σε πληροφορία και οπτικοποίησης της πληροφορίας αυτής. Μελλοντικά, σε περίπτωση που δοθεί κάποιο API από τον εκδοτικό οίκο Elsevier που θα δίνει πρόσβαση στα δεδομένα του θα μπορούσε να μετασχηματιστεί η εφαρμογή έτσι ώστε να μην χρειάζεται να γίνεται harvesting στα δεδομένα τους. Τέλος, μια πολύ σημαντική επέκταση της εφαρμογής θα μπορούσε να είναι η πρόσβαση στα πρωτογενή δεδομένα μέσω της web εφαρμογής. Για παράδειγμα, θα μπορούσαμε στην οπτικοποίηση του μεγέθους Αριθμός Paper ανά Χώρα, να κάναμε κλικ πάνω σε μια χώρα στον χάρτη και να άνοιγε ένα data grid με τα paper (id, scientific_field, title, affiliation, received, accepted, abstract, keywords, JournalID) από την συγκεκριμένη χώρα.

8. Βιβλιογραφία

- [1] <http://web-harvest.sourceforge.net/>
- [2] <http://www.doi.org/>
- [3] <http://www.w3schools.com/xpath/default.asp>
- [4] <http://www.w3schools.com/xquery/default.asp>
- [5] <http://www.ammmap.com/>
- [6] Master Thesis Μπέσσα Σωτηρία, «Ημιαυτόματες διαδικασίες συλλογής και σύνθεσης πληροφορίας από ιστοσελίδες και Web 2.0 εκπαιδευτικές συνιστώσες. Μελέτη και εφαρμογή τους στα πλαίσια υλοποίησης της πύλης WEP», Πάτρα 2008

9. Παράρτημα

9.1. Web-Harvest Configuration File

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<config charset="UTF-8">
```

```
<var-def name="startUrl">http://dx.doi.org/10.1016/</var-def>
<var-def name="midUrl">j.ejor.</var-def>
<var-def name="YYYYUrl">
  <script return="lastyear">
    thisyear = Integer.parseInt(sys.datetime("yyyy"));
    lastyear = thisyear - 1;
  </script>
</var-def>
<var-def name="MM.Url"><template>${sys.datetime(".MM.")}</template></var-def>
<var-def name="i">
<var-def name="j">
<var-def name="k">
  <file action="write" path="${midUrl}.xml" charset="UTF-8">
    <![CDATA[ <ROOT> ]]>
  </file>
<while condition="true" index="i" maxloops="9">
<while condition="true" index="j" maxloops="9">
<while condition="true" index="k" maxloops="9">

<file action="append" path="${midUrl}.xml" charset="UTF-8">
  <loop item="articleUrl" index="l" maxloops="1">
    <!-- collects URLs of all articles from the front page -->
    <list>
      <xpath expression="//div[@class='articleText']">
        <html-to-xml>
          <http url="${startUrl}${midUrl}${YYYYUrl}${MM.Url}${i}${j}${k}"/>
        </html-to-xml>
      </xpath>
      <xpath expression="//div[@class='articleTitle']">
        <html-to-xml>
          <http url="${startUrl}${midUrl}${YYYYUrl}${MM.Url}${i}${j}${k}"/>
```

```

        </html-to-xml>
    </xpath>
    <xpath expression="//div[@class='articleHeaderInner']">
        <html-to-xml>
            <http url="{startUrl}{midUrl}{YYYYUrl}{MM.Url}{i}{j}{k}"/>
        </html-to-xml>
    </xpath>
</list>

<!-- downloads each article and extract data from it -->
<body>
    <xquery>
        <xq-param name="doc">
            <html-to-xml>
                <http url="{startUrl}{midUrl}{YYYYUrl}{MM.Url}{i}{j}{k}"/>
            </html-to-xml>
        </xq-param>
        <xq-expression><![CDATA[
            declare variable $doc as node() external;
            let $aff := data($doc//div[@class='articleHeaderInner'])
            let $aff7 := data($doc//div[@class='articleHeaderInner'])
            let $aff1 := data($doc//div[@class='articleTitle'])
            let $aff2 := data($doc//div[@class='articleText authorsNoEnt'])

            return
                <article>
                    <id>{data(substring-before((substring-after($aff,
                        "doi:")),
                    " |"))}</id>
                    <scientific_field>{data(substring-before((substring-after($aff7,
                    "Reprints")), $aff1))}</scientific_field>
                    <title>{data($aff1)}</title>
                    <affiliation>{data($aff2)}</affiliation>
                    <received>{data(substring-before((substring-after($aff,
                    "Received"), ";"))}</received>
                    <accepted>{data(substring-before((substring-after($aff,
                    "accepted"), "."))}</accepted>
                    <abstract>{data(substring-before((substring-after($aff,
                    "Abstract")),
                    "Keywords"))}</abstract>
                    <keywords>{data(replace((substring-before((substring-after($aff,
                    "Keywords:")), "Article Outline")), ";", ","))}</keywords>
        ]]>
    </xq-expression>
</body>
</xquery>
</xsl:template>
</xsl:stylesheet>

```

```
        </article>
    ]]></xq-expression>
</xquery>
</body>

</loop>
</file>

</while>
</while>
</while>
<file action="append" path="{midUrl}.xml" charset="UTF-8">
    <![CDATA[ </ROOT> ]]>
</file>
</var-def>
</var-def>
</var-def>
</var-def>
</config>
```

9.2. SQL Server Database Creation Script

```

SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[article]') AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[article](
[id] [varchar](50) NOT NULL,
[scientific_field] [varchar](max) NULL,
[title] [varchar](max) NULL,
[affiliation] [varchar](max) NULL,
[received] [varchar](max) NULL,
[accepted] [varchar](max) NULL,
[abstract] [varchar](max) NULL,
[keywords] [varchar](max) NULL,
[JournalID] [nvarchar](50) NULL,
CONSTRAINT [PK_article] PRIMARY KEY CLUSTERED
(
[id] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO

SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO

IF NOT EXISTS (SELECT * FROM sys.triggers WHERE object_id =
OBJECT_ID(N'[dbo].[DeletefromArticleCountry]'))
EXEC dbo.sp_executesql @statement = N'CREATE TRIGGER
[dbo].[DeletefromArticleCountry]

```

```

ON [dbo].[article]
AFTER DELETE
AS
DELETE FROM ArticleCountry
FROM DELETED INNER JOIN
      ArticleCountry ON DELETED.id = ArticleCountry.ArticleID
,
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

IF NOT EXISTS (SELECT * FROM sys.triggers WHERE object_id =
OBJECT_ID(N'[dbo].[InsertIntoArticleCountry]'))
EXEC dbo.sp_executesql @statement = N'

-- =====
-- Author:      <Author,,Name>
-- Create date: <Create Date,,>
-- Description:  <Description,,>
-- =====

CREATE TRIGGER [dbo].[InsertIntoArticleCountry]
ON [dbo].[article]
AFTER INSERT
AS
INSERT INTO ArticleCountry
SELECT CensusData.[Country / Territory ], CensusData.[ISO Alpha-2], INSERTED.id
FROM INSERTED INNER JOIN CensusData
ON (INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[Country
/ Territory]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%", CensusData.[ISO
3166 Country Name]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 0]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 1]) }, "%") } + "%"

```

```

OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 2]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 3]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 4]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 5]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 6]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 7]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 8]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 9]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 10]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 11]) }, "%") } + "%"
OR INSERTED.affiliation LIKE "%" + { fn CONCAT({ fn CONCAT("%",
CensusData.[Alternative name 12]) }, "%") } + "%"

```

```

INSERT INTO ArticleCountry
(CountryName, CountryID, ArticleID)
VALUES ('','','')

```

```

DELETE FROM ArticleCountry
WHERE (CountryName = '') AND (CountryID = '') AND (ArticleID = '')

```

```

,
GO

```

```

SET ANSI_NULLS ON
GO

```

```

SET QUOTED_IDENTIFIER ON
GO

```

```

IF NOT EXISTS (SELECT * FROM sys.triggers WHERE object_id =
OBJECT_ID(N'[dbo].[UpdateJournaID]'))

```

```

EXEC dbo.sp_executesql @statement = N'
-----
-- Author:          <Author,,Name>
-- Create date: <Create Date,,>
-- Description:     <Description,,>
-----

CREATE TRIGGER [dbo].[UpdateJournalID]
  ON [dbo].[article]
  AFTER INSERT
AS
UPDATE article
SET      JournalID = substring(article.id, PATINDEX( "%/%", article.id)+1,
len(article.id)-len(right(article.id, PATINDEX( "%/%", article.id)))-12)

'

GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

IF NOT EXISTS (SELECT * FROM sys.triggers WHERE object_id =
OBJECT_ID(N'[dbo].[UpdateScientificField]'))
EXEC dbo.sp_executesql @statement = N'CREATE TRIGGER [dbo].[UpdateScientificField]
on [dbo].[article]
AFTER INSERT
AS
UPDATE article
SET      scientific_field = Scientific_Field.Scientific_Field
FROM      article INNER JOIN Scientific_Field
ON (Scientific_Field.Scientific_Field LIKE "%" + { fn CONCAT({ fn CONCAT("%",
article.scientific_field) }, "%") } + "%")'
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO

```

```

IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[Publisher]') AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[Publisher](
    [PublisherID] [nvarchar](50) NOT NULL,
    [PublisherName] [nvarchar](50) NOT NULL,
    CONSTRAINT [PK_Table_1] PRIMARY KEY CLUSTERED
(
    [PublisherID] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[CensusData]') AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[CensusData](
    [Country / Territory ] [varchar](50) NULL,
    [ISO Alpha-2] [varchar](50) NOT NULL,
    [ISO Alpha-3] [varchar](50) NOT NULL,
    [Population ] [numeric](18, 4) NULL,
    [Latitude] [float] NULL,
    [Longitude] [float] NULL,
    [ISO 3166 Country Name] [nvarchar](255) NULL,
    [Alternative name 0] [nvarchar](255) NULL,
    [Alternative name 1] [nvarchar](255) NULL,
    [Alternative name 2] [nvarchar](255) NULL,
    [Alternative name 3] [nvarchar](255) NULL,
    [Alternative name 4] [nvarchar](255) NULL,
    [Alternative name 5] [nvarchar](255) NULL,
    [Alternative name 6] [nvarchar](255) NULL,
    [Alternative name 7] [nvarchar](255) NULL,
    [Alternative name 8] [nvarchar](255) NULL,
    [Alternative name 9] [nvarchar](255) NULL,

```



```

[Alternative name 10] [nvarchar](255) NULL,
[Alternative name 11] [nvarchar](255) NULL,
[Alternative name 12] [nvarchar](255) NULL,
[IncomePerCapita] [money] NULL,
CONSTRAINT [PK_CensusData_1] PRIMARY KEY CLUSTERED
(
    [ISO Alpha-2] ASC
) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[Scientific_Field]') AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[Scientific_Field](
    [Scientific_Field_Code] [int] IDENTITY(1,1) NOT NULL,
    [Scientific_Field] [varchar](max) NULL,
    CONSTRAINT [PK_Scientific_Field] PRIMARY KEY CLUSTERED
(
    [Scientific_Field_Code] ASC
) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
END
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[ArticleCountry]') AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[ArticleCountry](
    [CountryName] [nvarchar](50) NOT NULL,

```

```

        [CountryID] [nchar](10) NOT NULL,
        [ArticleID] [nchar](50) NOT NULL,
        CONSTRAINT [PK_ArticleCountry] PRIMARY KEY CLUSTERED
    (
        [CountryID] ASC,
        [ArticleID] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
END
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[Journal]') AND type in (N'U'))
BEGIN
CREATE TABLE [dbo].[Journal](
    [JournalID] [nvarchar](50) NOT NULL,
    [JournalName] [nvarchar](50) NOT NULL,
    [PublisherID] [nvarchar](50) NOT NULL,
    CONSTRAINT [PK_Journal] PRIMARY KEY CLUSTERED
    (
        [JournalID] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
END
GO
IF NOT EXISTS (SELECT * FROM sys.foreign_keys WHERE object_id =
OBJECT_ID(N'[dbo].[FK_Journal_Journal]') AND parent_object_id =
OBJECT_ID(N'[dbo].[Journal]'))
ALTER TABLE [dbo].[Journal] WITH CHECK ADD CONSTRAINT [FK_Journal_Journal]
FOREIGN KEY([PublisherID])
REFERENCES [dbo].[Publisher] ([PublisherID])
GO
ALTER TABLE [dbo].[Journal] CHECK CONSTRAINT [FK_Journal_Journal]

```

9.3. Web Page Source Code

9.3.1. connect2db.php

```
<?php
$DSN="webharvest";
$user="";
$pass="";
$conn=odbc_connect( $DSN, $user, $pass);
?>
```

9.3.2. main.html

```
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<title>ammap</title>
<style>
  body{
    margin: 0px;
  }
</style>
</head>

<body onLoad="ammap.focus()" bgcolor="#E3E3FF" SCROLL=NO>
<div align="center">
<font face="Arial" size="6">
Major Elsevier Journals on Decision Sciences since 2002
</font>
</div>
<div align="center">
<font face="Arial" size="4">
Number Of Papers per Country
</font>
</div>
<!-- saved from url=(0013)about:internet -->
<!-- ammap script-->
  <script type="text/javascript" src="../ammap/swfobject.js"></script>
  <div id="flashcontent">
```

```
        <strong>You need to upgrade your Flash Player</strong>
    </div>
    <script type="text/javascript">
        // 
        var so = new SWFObject("../ammap/ammap.swf", "ammap", "100%", "92%",
"8", "#FFFFFF");
        so.addVariable("path", "../ammap/");
        so.addVariable("data_file", escape("NumberOfArticles.php"));
        so.addVariable("settings_file", escape("ammap_settings.xml"));
        so.addVariable("preloader_color", "#999999");
        so.write("flashcontent");
        // ]]&gt;
    &lt;/script&gt;
&lt;!-- end of ammap script --&gt;
&lt;div&gt;
&lt;script type="text/javascript"&gt;

var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-23652679-1']);
_gaq.push(['_trackPageview']);

(function() {
var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-
analytics.com/ga.js';
var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
})();

&lt;/script&gt;
&lt;/div&gt;
&lt;/body&gt;
&lt;/html&gt;</pre></div><div data-bbox="138 856 689 872" data-label="Page-Footer">Άντληση Δεδομένων Εκδοτικού Οίκου Elsevier και Απεικόνιση αυτών σε Διαδραστικό Χάρτη</div><div data-bbox="835 856 862 870" data-label="Page-Footer">52</div>
```

9.3.3. list.html

```
<html>
<head>
  <style type="text/css">

body {
  font-family: Verdana, Arial, Helvetica, sans-serif;
  margin: 0;
  font-size: 80%;
  font-weight: bold;
  background: #E3E3FF;
}

h2 {
  font: bold 14px Verdana, Arial, Helvetica, sans-serif;
  color: #000;
  margin: 0px;
  padding: 0px 0px 0px 15px;
}

ul {
  list-style: none;
  margin: 0;
  padding: 0;
}

img {
  border: none;
}

/*- Menu 2----- */

#menu2 {
  width: 200px;
  margin: 10px;
  border-style: solid solid none solid;
```

```
border-color: #D8D5D1;
border-size: 1px;
border-width: 1px;
}

#menu2 li a {
height: 32px;
voice-family: "\"}\"";
voice-family: inherit;
height: 50px;
text-decoration: none;
}

#menu2 li a:link, #menu2 li a:visited {
color: #000000;
display: block;
background: url(menu2.gif);
padding: 8px 0 0 30px;
}

#menu2 li a:hover {
color: #3688BA;
background: url(menu2.gif) 15 -32px;
padding: 8px 0 0 32px;
}

</style>
</head>

<body>

<div id="menu2">
  <ul>
    <li><a
                                href="main.html"
                                onclick="parent.banner.location='banner.html'" target="main">NUMBER OF PAPERS PER
                                COUNTRY</a></li>
```

```

        </li><head><base                                target="banner"></head><a
href="form_Journals.php"                            onclick="parent.titlos.location='titlejournals.html'"
target="banner">NUMBER OF PAPERS BY JOURNAL AND COUNTRY</a></li>
        </li><a                                        href="ArticlesPerCapita.php"
onclick="parent.banner.location='banner.html'" target="main">PAPERS PER CAPITA - TOP 20
COUNTRIES</a></li>
        </li><head><base                                target="banner"></head><a
href="../flight_routes/form_Cooperations.php"
onclick="parent.titlos.location='titleCooperations.html'" target="banner">NETWORKS OF
INTERNATIONAL COOPERATIONS</a></li>
        </li><head><base                                target="banner"></head><a
href="form_keywords.html"                          onclick="parent.titlos.location='titleKeywords.html'"
target="banner">NUMBER OF PAPERS BY KEYWORD AND COUNTRY</a></li>
    </ul>
</div>
<br />

</body>
</html>

```

9.3.4. underlist1.html

```

<head>
<BODY bgcolor="#E3E3FF" SCROLL=NO>
<div align="left">
<font face="Arial">
DSS Lab
<div align="left">
Department of Informatics
<div align="left">
University of Piraeus
</font>
</BODY>
</head>

```

9.3.5. underlist2.html

```

<html>
<BODY bgcolor="#E3E3FF">
<?php
include('connect2db.php');

```

```

if (!$conn)
    {exit("Connection Failed: " . $conn);}
$sql="SELECT distinct last_user_update as LastUserUpdate
FROM sys.dm_db_index_usage_stats
WHERE      database_id      =      DB_ID(      'webharvest')      AND
OBJECT_ID=OBJECT_ID('ArticleCountry');
$rs=odbc_exec($conn,$sql);
if (!$rs)
    {exit("Error in SQL");}
$myString='Last Update';
$LastUserUpdate=odbc_result($rs,"LastUserUpdate");
echo $myString.' '.$LastUserUpdate;
odbc_close($conn);
?>
</BODY></html>

```

9.3.6. NumberOfArticles.php

```

<?php
include('connect2db.php');
if (!$conn)
    {exit("Connection Failed: " . $conn);}
$sql="SELECT CountryName, CountryID, COUNT(ArticleID) AS NumberOfArticles
FROM      ArticleCountry
GROUP BY CountryName, CountryID";
$rs=odbc_exec($conn,$sql);
if (!$rs)
    {exit("Error in SQL");}
$myString1 = '<map map_file="maps/world.swf" zoom="100"> <areas>';
echo $myString1;
while (odbc_fetch_row($rs))
    {
    $myString2 = '<area title="';
    $myString3 = 'mc_name="';
    $myString4 = 'value="';
    $myString5 = "' ";
    $myString6 = "'/>';
    $CountryName=odbc_result($rs,"CountryName");
    $CountryID=odbc_result($rs,"CountryID");

```



```

$NumberOfArticles=odbc_result($rs,"NumberOfArticles");

echo $myString2.trim($CountryName).$myString5;
echo $myString3.trim($CountryID).$myString5;
echo $myString4.$NumberOfArticles.$myString6;
}
odbc_close($conn);
$myString7='</areas>
          </map>';
echo $myString7;
?>

```

9.3.7. form_Journals.php

```

<?php
include('connect2db.php');
$JournalList = "SELECT    JournalID, JournalName
FROM      Journal
order by JournalName";
$result=odbc_exec($conn,$JournalList);
?>
<HTML>
<head>
<base target="main">
</head>
<body bgcolor="#E3E3FF">
<form method="POST" action="Journals.php" align=top>
<select name="option">
<?php
while ($row = odbc_fetch_row($result)) {
$myString1='<option value=';
$myString2='>';
$myString3='</option\>';
$JournalID=odbc_result($result,"JournalID");
$JournalName=odbc_result($result,"JournalName");
echo $myString1.$JournalID.$myString2.$JournalName.$myString3;
}
?>
</select>

```

```
<input type="submit" value="Submit" name="Submit"/></form>
</body>
</HTML>
```

9.3.8. Journals.php

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<title>ammap</title>
<style>
  body{
    margin: 0px;
  }
</style>
</head>

<body onLoad="ammap.focus()" bgcolor="#E3E3FF" SCROLL=NO>
<div align="center">
<font face="Arial" size="6">
Major Elsevier Journals on Decision Sciences since 2002
</font>
<div align="center">
<font face="Arial" size="4">
Number Of Papers by Journal and Country
</font>
<!-- saved from url=(0013)about:internet -->
<!-- ammap script-->
  <script type="text/javascript" src="../ammap/swfobject.js"></script>
  <div id="flashcontent">
    <strong>You need to upgrade your Flash Player</strong>
  </div>
  <!-- saved from url=(0013)about:internet -->
  <!-- ammap script-->
  <script type="text/javascript" src="../ammap/swfobject.js"></script>
  <div id="flashcontent">
    <strong>You need to upgrade your Flash Player</strong>
  </div>
```

```

<script type="text/javascript">
    // 
        var so = new SWFObject("../ammap/ammap.swf", "ammap", "100%", "92%",
"8", "#FFFFFF");
        so.addVariable("path", "../ammap/");
        so.addVariable("map_data", escape('&lt;?php include ('connect2db.php'); if (!$conn)
{exit("Connection Failed: " . $conn);} $sql="SELECT ArticleCountry.CountryName,
ArticleCountry.CountryID, COUNT(ArticleCountry.ArticleID) AS NumberOfJournals FROM article
INNER JOIN ArticleCountry ON article.id = ArticleCountry.ArticleID INNER JOIN Journal ON
article.JournalID = Journal.JournalID WHERE Journal.JournalID = '{$_POST['option']}' group by
ArticleCountry.CountryName, ArticleCountry.CountryID order by ArticleCountry.CountryName";
$rs=odbc_exec($conn,$sql); if (!$rs) {exit("Error in SQL");} $myString1 = '&lt;map
map_file="maps/world.swf" zoom="100"&gt; &lt;areas&gt;'; echo $myString1; while
(odbc_fetch_row($rs)) {$myString2 = '&lt;area title="'; $myString3 = 'mc_name="'; $myString4
= 'value="'; $myString5 = ' '; $myString6 = '"/&gt;';
$CountryName=odbc_result($rs,"CountryName"); $CountryID=odbc_result($rs,"CountryID");
$NumberOfJournals=odbc_result($rs,"NumberOfJournals"); echo
$myString2.$CountryName.$myString5; echo $myString3.trim($CountryID).$myString5; echo
$myString4.$NumberOfJournals.$myString6;} odbc_close($conn);
$myString7='&lt;/areas&gt;&lt;/map&gt;'; echo $myString7; ?&gt;));
        so.addVariable("settings_file", escape("ammap_settings.xml"));
        so.addVariable("preloader_color", "#999999");
        so.write("flashcontent");
    // ]]&gt;
&lt;/script&gt;
&lt;!-- end of ammap script --&gt;
&lt;div&gt;
&lt;script type="text/javascript"&gt;

var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-23652679-1']);
_gaq.push(['_trackPageview']);

(function() {
var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-
analytics.com/ga.js';
var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
})();

&lt;/script&gt;
&lt;/div&gt;
</pre>
</div>
<div data-bbox="138 856 689 871" data-label="Page-Footer">Αντίληψη Δεδομένων Εκδοτικού Οίκου Elsevier και Απεικόνιση αυτών σε Διαδραστικό Χάρτη</div>
<div data-bbox="832 856 859 870" data-label="Page-Footer">59</div>
```

```
</body>
</html>
```

9.3.9. ArticlesPerCapita.php

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<title>ammap</title>
<style>
  body{
    margin: 0px;
  }
</style>
</head>
<body onLoad="ammap.focus()" bgcolor="#E3E3FF" SCROLL=NO>
<div align="center">
<font face="Arial" size="6">
Major Elsevier Journals on Decision Sciences since 2002
</font>
<div align="center">
<font face="Arial" size="4">
Papers per Capita - Top 20 Countries
</font>
<!-- saved from url=(0013)about:internet -->
<!-- ammap script-->
  <script type="text/javascript" src="../ammap/swfobject.js"></script>
    <div id="flashcontent">
      <strong>You need to upgrade your Flash Player</strong>
    </div>

  <script type="text/javascript">
    // 
      var so = new SWFObject("../ammap/ammap.swf", "ammap", "100%", "92%",
"8", "#FFFFFF");
      so.addVariable("path", "../ammap/");
      so.addVariable("map_data", escape('&lt;?php
include('connect2db.php');
if (!$conn)</pre>
</div>
<div data-bbox="138 856 689 871" data-label="Page-Footer">Αντίληψη Δεδομένων Εκδοτικού Οίκου Elsevier και Απεικόνιση αυτών σε Διαδραστικό Χάρτη</div>
<div data-bbox="832 856 859 870" data-label="Page-Footer">60</div>
```

```

    {exit("Connection Failed: " . $conn);}
    $sql="SELECT TOP 20      ArticleCountry.CountryName, ArticleCountry.CountryID,
CAST((COUNT(ArticleCountry.ArticleID)/CensusData.Population)*10000000 AS NUMERIC(10,2))
as ratio
FROM      ArticleCountry INNER JOIN
          CensusData ON ArticleCountry.CountryID = CensusData.[ISO Alpha-2]
GROUP BY ArticleCountry.CountryID, CensusData.Population, ArticleCountry.CountryName
ORDER BY ratio DESC";
$rs=odbc_exec($conn,$sql);
if (!$rs)
    {exit("Error in SQL");}
$myString1 = '<map map_file="maps/world.swf" zoom="100"> <areas>';
echo $myString1;
while (odbc_fetch_row($rs))
    {
    $myString2 = '<area title="';
    $myString3 = 'mc_name="';
    $myString4 = 'value="';
    $myString5 = "' ";
    $myString6 = "'/>';
    $CountryName=odbc_result($rs,"CountryName");
    $CountryID=odbc_result($rs,"CountryID");
    $NumberOfArticles=odbc_result($rs,"ratio");

    echo $myString2.trim($CountryName).$myString5;
    echo $myString3.trim($CountryID).$myString5;
    echo $myString4.$NumberOfArticles.$myString6;
    }
odbc_close($conn);
$myString7='</areas></map>';
echo $myString7;
?>));
    so.addVariable("settings_file", escape("ammap_settings1.xml"));
    so.addVariable("preloader_color", "#999999");
    so.write("flashcontent");
    // ]]>
</script>
<!-- end of ammap script -->
<div>

```

```

<script type="text/javascript">

var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-23652679-1']);
_gaq.push(['_trackPageview']);

(function() {
  var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
  ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-
analytics.com/ga.js';
  var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
})();

</script>
</div>
</body>
</html>

```

9.3.10. form_Cooperations.php

```

<?php
//Select info from MySQL
include('connect2db.php');

$CountryList = "SELECT DISTINCT CountryID, CountryName
FROM ArticleCountry
ORDER BY CountryName";
$result=odbc_exec($conn,$CountryList);
?>
<HTML>
<head>
<base target="main">
</head>
<body bgcolor="#E3E3FF">
<form method="POST" action="Cooperations.php">
<select name="option">
<?php
while ($row = odbc_fetch_row($result)) {
$myString1='<option value=';

```

```

$myString2='\>';
$myString3='</option\>';
$CountryID=odbc_result($result,"CountryID");
$CountryName=odbc_result($result,"CountryName");
echo $myString1.$CountryID.$myString2.$CountryName.$myString3;
}
?>
</select>
<input type="submit" value="Submit" name="Submit"/>
</form>
</body>
</HTML>

```

9.3.11. Cooperations.php

```

<html>
<base target="main">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<title>ammap</title>
<style>
  body{
    margin: 0px;
  }
</style>
</head>

<body onLoad="ammap.focus()" bgcolor="#E3E3FF" SCROLL=NO>
<div align="center">
<font face="Arial" size="6">
Major Elsevier Journals on Decision Sciences since 2002
</font>
<div align="center">
<font face="Arial" size="4">
Networks Of International Cooperations
</font>
<!-- saved from url=(0013)about:internet -->
<!-- ammap script-->
  <script type="text/javascript" src="../ammap/swfobject.js"></script>

```

```

<div id="flashcontent">
    <strong>You need to upgrade your Flash Player</strong>
</div>

<script type="text/javascript">
    // 
        var so = new SWFObject("../ammap/ammap.swf", "ammap", "100%", "92%",
"8", "#FFFFFF");
        so.addVariable("path", "../ammap/");
        so.addVariable("map_data", escape('&lt;?php include ('connect2db.php');
        if (!$conn) {exit("Connection Failed: " . $conn);}
        $sql0="SELECT ArticleCountry.CountryName FROM ArticleCountry inner join
ArticleCountry AS ArticleCountry_1 on ArticleCountry.ArticleID=ArticleCountry_1.ArticleID inner
join CensusData on ArticleCountry.CountryID=[CensusData].[ISO Alpha-2] inner join
CensusData AS CensusData_1 on ArticleCountry_1.CountryID=[CensusData_1].[ISO Alpha-2]
where ArticleCountry.CountryID&lt;&gt;ArticleCountry_1.CountryID AND
ArticleCountry.CountryID='{$_POST['option']}' group by ArticleCountry.CountryName,
ArticleCountry.CountryName, ArticleCountry.CountryID, ArticleCountry_1.CountryID,
CensusData.Latitude, CensusData_1.Latitude, CensusData.Longitude, CensusData_1.Longitude,
ArticleCountry_1.CountryName";
        $rs0=odbc_exec($conn,$sql0);
        if (!$rs0) {exit("Error in SQL");}
        $myString0 ='&lt;map map_file="maps/world.swf" tl_long="-168.49" tl_lat="83.63"
br_long="190.3" br_lat="-55.58" url="#movie_';
        $CountryName=odbc_result($rs0,"CountryName");
        $myString1 = ""&gt;&lt;areas&gt;&lt;area title="borders" mc_name="borders" color="#FFFFFF"
balloon="false"&gt;&lt;/area&gt;&lt;/areas&gt;&lt;movies&gt;';
        echo $myString0.$CountryName.$myString1;
        $sql1="SELECT ArticleCountry.CountryName, CAST(CensusData.Latitude AS
numeric(10,4)) AS lat, CAST(CensusData.Longitude AS numeric(10,4)) AS long,
ArticleCountry_1.CountryName as CountryNameTarget, CAST(CensusData_1.Latitude AS
numeric(10,4)) AS latTarget, CAST(CensusData_1.Longitude AS numeric(10,4)) AS longTarget,
count(ArticleCountry.ArticleID) as NumberofArticles, log(cast(count(ArticleCountry.ArticleID) AS
numeric(10,4))) as width FROM ArticleCountry inner join ArticleCountry AS ArticleCountry_1 on
ArticleCountry.ArticleID=ArticleCountry_1.ArticleID inner join CensusData on
ArticleCountry.CountryID=[CensusData].[ISO Alpha-2] inner join CensusData AS CensusData_1
on ArticleCountry_1.CountryID=[CensusData_1].[ISO Alpha-2] where
ArticleCountry.CountryID&lt;&gt;ArticleCountry_1.CountryID AND
ArticleCountry.CountryID='{$_POST['option']}' group by ArticleCountry.CountryName,
ArticleCountry.CountryName, ArticleCountry.CountryID, ArticleCountry_1.CountryID,
CensusData.Latitude, CensusData_1.Latitude, CensusData.Longitude, CensusData_1.Longitude,
ArticleCountry_1.CountryName order by NumberofArticles";
        $rs1=odbc_exec($conn,$sql1);
</pre>
</div>
<div data-bbox="138 856 689 871" data-label="Page-Footer">Άντληση Δεδομένων Εκδοτικού Οίκου Elsevier και Απεικόνιση αυτών σε Διαδραστικό Χάρτη</div>
<div data-bbox="832 856 859 870" data-label="Page-Footer">64</div>
```



```

if (!$rs1) {exit("Error in SQL");}
while (odbc_fetch_row($rs1))
{
$myString2 = '<movie color="CC0000" file="target" title="';
$myString3 = ' lat="';
$myString4 = ' long="';
$myString5 = ' width="10" height="10" fixed_size="true"></movie>';
$CountryNameTarget=odbc_result($rs1,"CountryNameTarget");
$lat=odbc_result($rs1,"latTarget");
$long=odbc_result($rs1,"longTarget");
$NumberOfArticles=odbc_result($rs1,"NumberofArticles");
$sql11="SELECT ArticleCountry.CountryName, count(ArticleCountry.CountryName) as
NumberofPapers FROM ArticleCountry where ArticleCountry.CountryName='{ $CountryName}'
group by ArticleCountry.CountryName";
$rs11=odbc_exec($conn,$sql11);
$NumberOfPapers=odbc_result($rs11,"NumberofPapers");
echo $myString2."From a Total Number of ".$NumberOfPapers." Papers from
".$CountryName.",".$NumberOfArticles." Papers are in Cooperation with
".$CountryNameTarget.$myString3.$lat.$myString4.$long.$myString5;}
$myString6 = '<movie oid="movie_';
$CountryName=odbc_result($rs1,"CountryName");
$myString7 = ' file="target" title="';
$myString8 = ' lat="';
$lat=odbc_result($rs1,"lat");
$myString9 = ' long="';
$long=odbc_result($rs1,"long");
$myString10 = ' width="16" height="16" fixed_size="true">';
echo
$myString6.$CountryName.$myString7.$CountryName.$myString8.$lat.$myString9.$long.$myString10;
$sql2="SELECT ArticleCountry.CountryName, (CAST(CensusData.Latitude AS
varchar(8))+','+CAST(CensusData_1.Latitude AS varchar(8))) AS lat,
CAST(CensusData.Longitude AS varchar(8))+','+CAST(CensusData_1.Longitude AS varchar(8))
AS long, count(ArticleCountry.ArticleID) as NumberofArticles,
log(cast(count(ArticleCountry.ArticleID) as decimal(10,5))) as width FROM ArticleCountry inner
join ArticleCountry AS ArticleCountry_1 on ArticleCountry.ArticleID=ArticleCountry_1.ArticleID
inner join CensusData on ArticleCountry.CountryID=[CensusData].[ISO Alpha-2] inner join
CensusData AS CensusData_1 on ArticleCountry_1.CountryID=[CensusData_1].[ISO Alpha-2]
where ArticleCountry.CountryID<>ArticleCountry_1.CountryID AND
ArticleCountry_1.CountryID='{ $_POST['option']}' group by ArticleCountry.CountryName,
ArticleCountry.CountryName, ArticleCountry.CountryID, ArticleCountry_1.CountryID,
CensusData.Latitude, CensusData_1.Latitude, CensusData.Longitude, CensusData_1.Longitude
order by ArticleCountry.CountryID, ArticleCountry_1.CountryID";

```

```

$rs2=odbc_exec($conn,$sql2);
if (!$rs2) {exit("Error in SQL");}
$myString1 ='<lines>';
echo $myString1;
while (odbc_fetch_row($rs2))
{$myString2 = '<line lat='";
$myString3 = '" long='";
$myString4 = '" width='";
$myString5 = '" alpha="40" curved="true"></line>';
$lat=odbc_result($rs2,"lat");
$long=odbc_result($rs2,"long");
$width=odbc_result($rs2,"width");
echo $myString2.$lat.$myString3.$long.$myString4.$width.$myString5;}
odbc_close($conn);
$myString6='</lines></movie></movies></map>';
echo $myString6; ?>));
so.addVariable("settings_file", escape("ammap_settings.xml"));
    so.addVariable("preloader_color", "#999999");
    so.write("flashcontent");
    // ]]>
</script>
<!-- end of ammap script -->
<div>
<script type="text/javascript">

var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-23652679-1']);
_gaq.push(['_trackPageview']);

(function() {
var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-
analytics.com/ga.js';
var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
})();

</script>
</div>
</body>

```

```
</html>
```

9.3.12. form_keywords.html

```
<html>  
<head>  
<base target="main">  
</head>  
<body bgcolor="#E3E3FF">  
<form method="POST" action="keywords.php">  
Keyword: <input type="text" name="keyword" size="20"/>  
<input type="submit" value="Submit" name="Submit"/><br />  
</form>  
</body>  
</html>
```

9.3.13. keywords.php

```

<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<title>ammap</title>
<style>
  body{
    margin: 0px;
  }
</style>
</head>

<body onLoad="ammap.focus()" bgcolor="#E3E3FF" SCROLL=NO>
<div align="center">
<font face="Arial" size="6">
Major Elsevier Journals on Decision Sciences since 2002
</font>
<div align="center">
<font face="Arial" size="4">
Number Of Papers by Keywords And Country
</font>
<!-- saved from url=(0013)about:internet -->
<!-- ammap script-->
  <script type="text/javascript" src="../ammap/swfobject.js"></script>
    <div id="flashcontent">
      <strong>You need to upgrade your Flash Player</strong>
    </div>

    <script type="text/javascript">
      // 
        var so = new SWFObject("../ammap/ammap.swf", "ammap", "100%", "92%",
"8", "#FFFFFF");
        so.addVariable("path", "../ammap/");
        so.addVariable("map_data", escape('&lt;?php include ('connect2db.php'); if (!$conn)
{exit("Connection Failed: " . $conn);} $keyword = addslashes($_POST['keyword']);
$sql="SELECT ArticleCountry.CountryName, ArticleCountry.CountryID, count(article.id) as
count FROM article inner join ArticleCountry on article.id=ArticleCountry.ArticleID WHERE
</pre>
</div>
<div data-bbox="138 857 689 871" data-label="Page-Footer">Αντίληψη Δεδομένων Εκδοτικού Οίκου Elsevier και Απεικόνιση αυτών σε Διαδραστικό Χάρτη</div>
<div data-bbox="832 857 859 870" data-label="Page-Footer">68</div>
```

```

article.keywords LIKE '%{$keyword}%' COLLATE SQL_Latin1_General_CP1_CI_AS group by
ArticleCountry.CountryID, ArticleCountry.CountryName"; $rs=odbc_exec($conn,$sql); if (!$rs)
{exit("Error in SQL");} $myString1 = '<map map_file="maps/world.swf" zoom="100">
<areas>'; echo $myString1; while (odbc_fetch_row($rs)) {$myString2 = '<area title="';
$myString3 = 'mc_name="'; $myString4 = 'value="'; $myString5 = "'"; $myString6 = "'/>';
$CountryName=odbc_result($rs,"CountryName"); $CountryID=odbc_result($rs,"CountryID");
$NumberOfJournals=odbc_result($rs,"count"); echo $myString2.$CountryName.$myString5;
echo $myString3.trim($CountryID).$myString5; echo
$myString4.$NumberOfJournals.$myString6;} odbc_close($conn);
$myString7='</areas></map>'; echo $myString7; ?>));
so.addVariable("settings_file", escape("ammap_settings.xml"));
so.addVariable("preloader_color", "#999999");
so.write("flashcontent");
// ]]>
</script>
<!-- end of ammap script -->
<div>
<script type="text/javascript">

var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-23652679-1']);
_gaq.push(['_trackPageview']);

(function() {
var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-
analytics.com/ga.js';
var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
})();

</script>
</div>
</body>
</html>

```

9.3.14. ammap_settings.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<!-- Value between [] brackets, for example [#FFFFFF] shows default value which is used if
this parameter is not set -->

```

<!-- This means, that if you are happy with this value, you can delete this line at all and reduce file size -->

<!-- value or explanation between () brackets shows the range or type of values you should use for this parameter -->

<!-- x, y, widths and heights can be set in pixels or in percents -->

<settings>

<developer_mode></developer_mode> <!-- [false] (true / false) If set to true, developer information is displayed. You can also copy developer information to clipboard through right click menu -->

<projection>mercator</projection> <!-- [xy] (mercator, xy) Projection of your map. If you do not care about actual longitudes and latitudes I recommend to use xy projection. The default world map uses Mercator projection. -->

<width></width> <!-- [] (Number) if empty, will be equal to width of your flash movie -->

<height></height> <!-- [] (Number) if empty, will be equal to height of your flash movie -->

Tahoma <!-- [Tahoma] (font name) use device fonts, such as Tahoma, Times New Roman, Tahoma, Verdana... -->

<text_size>11</text_size> <!-- [11] (Number) text size of all texts. Every text size can be set individually in the settings below -->

<text_color></text_color> <!-- [#000000] (hex color code) main text color. Every text color can be set individually in the settings below-->

<decimals_separator></decimals_separator> <!-- [,] (string) decimal separator. Note, that this is for displaying data only. Decimals in data xml file must be separated with a dot -->

<thousands_separator></thousands_separator> <!-- [] (string) thousand separator -->

<redraw>true</redraw> <!-- [false] (true / false) if redraw is set to true, the map will be resized then screen size changes -->

<precision></precision> <!-- [2] (Number) shows how many numbers should be shown after comma for calculated values (percents) -->

<reload_data_interval></reload_data_interval> <!-- [0] (Number) how often data should be reloaded (time in seconds) -->

<preloader_on_reload></preloader_on_reload> <!-- [false] (true / false) Whether to show preloaded when data or settings are reloaded -->

<add_time_stamp></add_time_stamp> <!-- [false] (true / false) if true, a unique number will be added every time flash loads data. This will help to avoid data caching -->

<color_change_time_start></color_change_time_start> <!-- [2] (Number) How fast areas should change color after map is loaded -->

`<color_change_time_hover></color_change_time_hover>` `<!-- [0.2] (Number) How fast areas and movies should change color on roll out -->`
`<fit_to_screen></fit_to_screen>` `<!-- [true] (true / false) whether to resize map or not to fit it on screen -->`
`<drag_map></drag_map>` `<!-- [true] (true / false) whether map is draggable or not -->`
`<always_hand></always_hand>` `<!-- [false] (true / false) if set to true, mouse cursor will always be hand but only if drag_map=true -->`
`<mask></mask>` `<!-- [true] (true / false) this is important if you use print or printAsBitmap JavaScript function and want only visible part of a map to be printed. Also, if you load the map into another swf movie. -->`
`<js_enabled></js_enabled>` `<!-- [true] (true / false) In case you don't use any flash - JavaScript communication, you can set this setting to false - this will save some CPU and will disable the security warning message which appears when opening the chart from hard drive. -->`
`<force_smoothing></force_smoothing>` `<!-- [false] (true / false) If you use bitmap images as your maps, then you can set this to "true" - the quality of the image will be better (might slow the zooming) -->`
`<layers></layers>`

`<!-- [movies,labels,lines] Use this setting to change the sequence of layers. The left-most layer will be behind all the other layers. Please note that objects with x/y coordinates will still be placed on top of objects with long/lat coordinates. -->`

`<background>` `<!-- background -->`
`<color></color>` `<!-- [#444444] (hex color code) -->`
`<alpha></alpha>` `<!-- [0] (0 - 100) use 0 if you are using custom swf or jpg for background -->`
`<border_color></border_color>` `<!-- [#000000] (hex color code) -->`
`<border_alpha></border_alpha>` `<!-- [0] (0 - 100) -->`
`<file></file>` `<!-- [] (filename) swf or jpg file of a background. This file must be located in path/ folder (set in html file) -->`
`<stretch></stretch>` `<!-- [false] (true / false) should we stretch background over the whole width/heigh? -->`
`<overlay_file></overlay_file>` `<!-- [] (filename) swf or jpg file to use as an overlay for the map. This file must be located in path/ folder (set in html file) -->`
`<overlay_stretch></overlay_stretch>` `<!-- [false] (true / false) should we stretch overlay over the whole width/heigh? -->`
`</background>`

`<zoom>` `<!-- zoom control -->`
`<enabled>true</enabled>` `<!-- [true] (true / false) whether to display zoom control or not -->`

`<locked></locked>` `<!-- [false] (true / false) whether it`
would be possible to drag zoom bar when dragging "home" icon -->

`<mouse_wheel_enabled></mouse_wheel_enabled>` `<!-- [true] (true / false)`
whether to enable zooming with mouse wheel. Mouse wheel work only after user clicks
somewhere on a flash. To workaround this issue, you have to add: `onLoad="ammap.focus()"` to
your html `<body>` -->

`<arrows_enabled>false</arrows_enabled>` `<!-- [true] (true / false)`
whether to display navigation arrows or not -->

`<home_link_enabled></home_link_enabled>` `<!-- [true] (true / false)`
whether to display home link or not (home link is only displayed if arrows are enabled) -->

`<step_size></step_size>` `<!-- [10] (Number) when moving map`
with navigation arrows, the map moves by `step_size%` of flash width or height -->

`<color></color>` `<!-- [#990000] (hex color code)`
background color of "+", "-", scroller buttons and navigation arrows -->

`<color_hover></color_hover>` `<!-- [#CC0000] (hex color code) -->`

`<alpha></alpha>` `<!-- [100] (0 - 100) alpha of zoom`
control buttons -->

`<outline_color></outline_color>` `<!-- [#FFFFFF] -->`

`<outline_alpha></outline_alpha>` `<!-- [100] (0 - 100) -->`

`<x></x>` `<!-- [15] (Number or Number%) x position`
of zoom control -->

`<y></y>` `<!-- [15] (Number or Number%) y position`
of zoom control -->

`<height></height>` `<!-- [200] (Number or Number%) zoom`
control height -->

`<min></min>` `<!-- [] (Number%) minimum zoom. if left`
empty, the minimum value will be calculated automaticly (to fit all the map to the stage) -->

`<max></max>` `<!-- [2000%] (Number%) maximum`
zoom -->

`<grid_every></grid_every>` `<!-- [100%] (Number%) zoom tool`
grid will be displayed every "grid_every" value and the dragger height will be equal to the
distance between two grids -->

`<time></time>` `<!-- [1] time of zoom in seconds -->`

`<effect></effect>` `<!-- [easeout] (easeout / easein) zoom`
motion effect -->

`<rotate></rotate>` `<!-- [false] (true / false) whether to`
rotate zoom control by 90 degrees or not. If set to true, zoom control becomes horizontal -->

`<background_zooms_to_top></background_zooms_to_top>` `<!-- [false] (true /`
false) If this is set to true, clicking on background would be the same as clicking on "home"
button -->

`<zoom_on_click></zoom_on_click>` `<!-- [false] (true / false) If this is`
set to true, the map will be centered and zoomed in to the place you click. The
`background_zooms_to_top` will be set to false automatically if you set this to true. -->


```

    <grid_click_enabled></grid_click_enabled>          <!-- [true] (true / false) by
    default clicking on zoom control grid will move slider to that position. Set this setting to "false"
    to disable this behavior. -->
  </zoom>

```

```

  <small_map>                                          <!-- small map -->
    <enabled>false</enabled>                          <!-- [true] (true / false) whether to
    display small map or not -->
    <locked></locked>                                  <!-- [false] (true / false) whether it
    would be possible to drag small map when dragging border of small map -->
    <active></active>                                  <!-- [true] (true / false) whether clicking
    on a small map will result position change of big map -->
    <x></x>                                             <!-- [] (Number or Number%) x position of
    small map. If not defined, will be aligned to right side -->
    <y></y>                                             <!-- [] (Number or Number%) y position of
    small map. If not defined, will be aligned to bottom -->
    <width></width>                                     <!-- [25%] (Number or Number%) width
    of small map (height is calculated automatically) -->
    <color></color>                                    <!-- [] (hex color code) leave empty if you
    want to keep original map color -->
    <border_width></border_width>                      <!-- [5] (Number) -->
    <border_color></border_color>                     <!-- [#FFFFFF] (hex color code) -->
    <rectangle_color></rectangle_color>               <!-- [#FFFFFF] (hex color code)
    color of rectangle, indicating active area on a small map -->
    <collapse_button_color></collapse_button_color>    <!-- [#000000] (hex color
    code) collapse button arrow color (background color is the same as small map border color) -->
    <collapse_button_position></collapse_button_position> <!-- [BR] (BR / BL / TR / TL)
    BR - bottom right, BL - bottom left TR - top right TL - top left -->
  </small_map>

```

```

  <navigation_path>                                   <!-- navigation path -->
    <enabled></enabled>                                <!-- [false] (true / false) whether to
    display navigation path or not -->
    <x></x>                                             <!-- [70] (Number or Number%) x position
    of path -->
    <y></y>                                             <!-- [30] (Number or Number%) y position
    of path -->
    <color></color>                                    <!-- [#000000] (hex color code) path
    background color -->
    <alpha></alpha>                                    <!-- [0] (0 - 100) path background alpha
    -->

```

```

    <padding></padding>                                <!-- [0] (Number) horizontal padding
of path buttons -->
    <text_color></text_color>                          <!-- [#FFFFFF] (hex color code) -->
    <text_color_hover></text_color_hover>              <!-- [#CC0000] (hex color code) -
->
    <text_size></text_size>                            <!-- [text_size] (Number) -->
    <separator></separator>                          <!-- [] symbol which separates path
entries -->
    <home_text></home_text>                            <!-- [] (text) Home text -->
</navigation_path>

<area>                                                <!-- areas are movie clips which are already
in map swf file, usually countries. areas should be listed in data file -->
    <balloon_text>
        <![CDATA[{value} Papers from {title}]]>      <!-- [] ({title} {value} {percents}
{description}) You can add your own HTML formatted text (only some of html tags are
supported) -->
    </balloon_text>
    <active_only_if_value_set>true</active_only_if_value_set> <!-- [false] (true / false) if
set to true, areas (countries) without defined value in data file will not display balloon and will
not change color on roll over -->
    <color_solid>#FF0000</color_solid>                <!-- [#DD0000] (hex color
code) color of area with highest value-->
    <color_light>#FFDD00</color_light>                <!-- [#FFCC00] (hex color
code) color of areas with value = 0 or value is not set -->
    <color_hover>#FF0000</color_hover>                <!-- [] (hex color code) -->
    <color_selected></color_selected>                  <!-- [] (hex color code) If not
specified, original area color will be used -->
    <color_unlisted>#DADADA</color_unlisted>          <!-- [] (hex color code) The
color of the areas which are in the map file, but not listed in the data file. If you don't set any
color here, the area will keep it's original color. Setting a color allows keeping data file smaller. -
->
    <color_no_value></color_no_value>                  <!-- [] (hex color code) The color
of the areas which have no value attribute. Leave this empty if you do not use value attributes
in areas. -->
    <disable_when_clicked>true</disable_when_clicked> <!-- [false] (true / false) if
this area is currently selected, roll over and click events of this area will be disabled if this set to
true -->
    <preserve_color></preserve_color>                 <!-- [false] (true / false) if this is
set to true color_solid and color_light settings will be ignored - the area will preserve colors as
they are in the original SWF -->
</area>

```

`<movie>` `<!-- movies are swf, jpg, gif or png files which you load on top of a map or on stage. You should define your movies in data file -->`
`<balloon_text>`
`<![CDATA[{title}]]>` `<!-- [] ({title} {value} {description}) You can add your own HTML formatted text (only some of html tags are supported) -->`
`</balloon_text>`
`<active_only_if_value_set></active_only_if_value_set>` `<!-- [false] (true / false) if set to true, movies without defined value in data file will not display balloon and will not change color on roll over -->`
`<color></color>` `<!-- [] (hex color code) If not specified, original area color will be used -->`
`<color_hover></color_hover>` `<!-- [] (hex color code) If not specified, original area color will be used -->`
`<color_selected></color_selected>` `<!-- [] (hex color code) If not specified, original area color will be used -->`
`<hover_focus></hover_focus>` `<!-- [false] (true / false) If set to true hovered movie will be brought in front of other movies -->`
`</movie>`

`<label>` `<!-- labels are texts which you set in data file. labels can be bounded to a map or placed on stage -->`
`<balloon_text>`
`<![CDATA[]]>` `<!-- [] ({title} {value} {description}) You can add your own HTML formatted text (only some of html tags are supported) -->`
`</balloon_text>`
`<bg_margins_horizontal></bg_margins_horizontal>` `<!-- [10] if your labels have background, then this parameter sets horizontal margins of background box -->`
`<bg_margins_vertical></bg_margins_vertical>` `<!-- [10] if your labels have background, then this parameter sets vertical margins of background box -->`
`<hover_focus></hover_focus>` `<!-- [false] (true / false) If set to true hovered label will be brought in front of other labels -->`
`</label>`

`<line>` `<!-- you can bound lines to a map or draw them on the stage. Lines are defined in data file -->`
`<balloon_text>`
`<![CDATA[{title}]]>` `<!-- [] ({title}) You can add your own HTML formatted text (only some of html tags are supported) -->`
`</balloon_text>`

```

    <color></color>                                <!-- [#990000] (hex color code) default
line color -->
    <alpha></alpha>                                <!-- [100] (Number) default line alpha --
>
    <width></width>                                <!-- [0] (Number) default line width -->
    <dashed></dashed>                              <!-- [false] (true / false) dashed lines
are recommended only for stage lines, not for lines inside a map -->
    <arrow></arrow>                                <!-- [] (end / start / middle / both)
arrow position -->
    <arrow_size></arrow_size>                      <!-- [8] (Number) arrow size -->
    <arrow_color></arrow_color>                    <!-- [] (line.color) default arrow
color. If empty, is the same as line color -->
    <arrow_alpha></arrow_alpha>                    <!-- [100] (0 - 100) arrow alpha --
>
    <curved></curved>                              <!-- [false] (true / false) If this is set to
true, the lines will be "smoothed". The lines will be smoothed only if they have at least 3 points.
You can set curved="true" or curved="false" for individual line in the data file -->
    <hover_focus></hover_focus>                    <!-- [false] (true / false) If set to
true hovered line will be brought in front of other lines -->
</line>

<legend>                                          <!-- LEGEND -->
    <enabled></enabled>                            <!-- [true] (true / false) -->
    <x></x>                                         <!-- [20] (Number / Number% / !Number)
-->
    <y></y>                                         <!-- [!55] (Number / Number% / !Number)
-->
    <width></width>                                <!-- [] (Number / Number%) -->
    <max_columns></max_columns>                    <!-- [] (Number) the maximum
number of columns in the legend -->
    <color></color>                                <!-- [#FFFFFF] (hex color code) -->
    <alpha></alpha>                                <!-- [50] (0 - 100) background opacity --
>
    <border_color></border_color>                  <!-- [#000000] (hex color code)
border color -->
    <border_alpha></border_alpha>                  <!-- [30] (0 - 100) border alpha --
>
    <text_color></text_color>                       <!-- [text_color] (hex color code) -->
    <text_size></text_size>                         <!-- [text_size] (Number) -->
    <spacing></spacing>                            <!-- [10] (Number) vertical and
horizontal gap between legend entries -->

```

```

    <argins></argins>                                <!-- [10] (Number) legend margins
(space between legend border and legend entries, recommended to use only if legend border is
visible or background color is different from chart area background color) -->
    <key>                                             <!-- KEY (the color box near every legend
entry) -->
    <size></size>                                    <!-- [16] (Number) key size-->
    <border_color></border_color>                    <!-- [] (hex color code) leave empty
if you don't want to have border -->
    </key>
    <entries>                                         <!-- legend entries go here-->
    <!-- <entry color="#3366CC">Some title</entry> --> <!-- You can have any
number of legend entries, uncomment this line and copy as many times as you need -->
    </entries>
</legend>

<object_list>                                       <!-- OBJECT LIST -->
    <enabled>true</enabled>                          <!-- [false] (true / false) -->
    <type></type>                                     <!-- [dropdown] (dropdown / listbox) -->
    <levels></levels>                                <!-- [3] (Number) How many levels of
objects should be included in the list. You can exclude individual objects from a list by setting
object_list="false" in the object's settings in data file. -->
    <include_areas>>false</include_areas>            <!-- [true] (true / false)
whether areas must be included in a list -->
    <include_movies></include_movies>                <!-- [true] (true / false) whether
movies must be included in a list -->
    <include_labels></include_labels>                <!-- [true] (true / false) whether
labels must be included in a list -->
    <x></x>                                           <!-- [!200] (Number / Number% /
!Number) -->
    <y></y>                                           <!-- [0] (Number / Number% / !Number)--
>
    <width></width>                                   <!-- [200] (Number / Number%) -->
    <height></height>                                 <!-- [100%] (Number / Number%) -->
    <color></color>                                   <!-- [#FFFFFF] (hex color code)
Background color-->
    <color_hover></color_hover>                       <!-- [#DADADA] (hex color code)
background color of a list item when mouse is over it -->
    <color_selected></color_selected>                <!-- [#000000] (hex color code)
background color of a selected list item -->
    <alpha></alpha>                                  <!-- [100] (0 - 100) background opacity
-->

```

```

    <border_color></border_color>                <!-- [#DADADA] (hex color code) --
>
    <border_alpha></border_alpha>                  <!-- [100] (0 - 100) border opacity
-->
    <text_color></text_color>                      <!-- [text_color] (hex color code) -->
    <text_color_hover></text_color_hover>          <!-- [#000000] (hex color code) -
->
    <text_color_selected></text_color_selected>    <!-- [#FFFFFF] (hex color code)
-->
    <text_size></text_size>                        <!-- [text_size] (Number) -->
    <scroller_color></scroller_color>              <!-- [#990000] (hex color code) -->
    <scroller_bg_color></scroller_bg_color>        <!-- [#DADADA] (hex color code) -
->
    <home_text></home_text>                        <!-- [Home] (text) Set "none" if you
don't want the home text at all -->
    </object_list>

    <text_box>                                     <!-- text box is displayed when user clicks
on an object and the object contains description. You can disable text box for individual objects
in data file -->
        <enabled></enabled>                       <!-- [true] (true / false) -->
        <locked></locked>                         <!-- [false] (true / false) whether it
would be possible to drag text box -->
        <selectable></selectable>                 <!-- [true] (true / false) whether it is
possible to select the text in text box or not -->
        <show_on_hover></show_on_hover>           <!-- [false] (true / false) text
box can be showed on roll over too - set this parameter to "true" -->
        <hide_on_roll_out></hide_on_roll_out>     <!-- [false] (true / false) if text
box is showed on roll over, it can be hidden on roll out -->
        <x></x>                                   <!-- [60%] (Number or Number%) x
position of text box. You can set individual x position for every object in data file -->
        <y></y>                                   <!-- [20%] (Number or Number%) y
position of text box. You can set individual y position for every object in data file -->
        <width></width>                           <!-- [35%] (Number or Number%) text
box width. You can set individual width for every movie or area -->
        <height></height>                         <!-- [40%] (Number or Number%) text
box height. You can set individual height for every movie or area -->
        <color></color>                           <!-- [#FFFFFF] (hex color code)
background color -->
        <alpha></alpha>                          <!-- [100] (0 - 100) background alpha --
>

```

```

    <shadow_color></shadow_color>                <!-- [#000000] (hex color code)
shadow color -->
    <shadow_alpha></shadow_alpha>                <!-- [50] (0 - 100) shadow alpha
-->
    <shadow_blur></shadow_blur>                <!-- [5] (Number) shadow
bluriness -->
    <shadow_distance></shadow_distance>        <!-- [5] (Number) shadow
distance -->
    <corner_radius>7</corner_radius>          <!-- [0] (Number) corner radius --
>
    <border_width></border_width>                <!-- [0] (Number) border width -->
    <border_color></border_color>                <!-- [#DADADA] (hec color code)
border color -->
    <border_alpha></border_alpha>                <!-- [100] (Number) border alpha -
->
    <margin_width></margin_width>                <!-- [10] (Number) margin width -
->
    <text_size></text_size>                    <!-- [text_size] (Number) -->
    <text_color></text_color>                    <!-- [text_color] (hex color code) -->
    <close_button_color></close_button_color>    <!-- [#000000] (hex color code)
-->
    <close_button_color_hover></close_button_color_hover> <!-- [#CC0000] (hex color
code) -->
    <scroller_color></scroller_color>            <!-- [#990000] -->
    <scroller_bg_color></scroller_bg_color>      <!-- [#DADADA] -->
</text_box>

    <balloon>                                    <!-- balloon is shown when user roll overs
the objects -->
    <enabled></enabled>                          <!-- [true] (true / false) -->
    <max_width></max_width>                      <!-- [30%] (Number or Number%)
-->
    <color></color>                              <!-- [#880000] (hex color code) balloon
background color -->
    <alpha></alpha>                              <!-- [100] (0 - 100) -->
    <text_color></text_color>                    <!-- [0xFFFFFFFF] (hex color code) -->
    <text_size></text_size>                    <!-- [text_size] (Number) -->
    <border_color></border_color>                <!-- [balloon.color] (hex color code)
-->
    <border_alpha></border_alpha>                <!-- [balloon.alpha] (0 - 100) -->
    <border_width></border_width>                <!-- [0] (Number) -->

```

```

    <corner_radius></corner_radius>
    <arrow></arrow>
    balloon's arrow position -->
  </balloon>

  <export_as_image>
  only on a web server -->
    <file></file>
    context menu (then user right clicks on flash movie) "Export as image" will appear. This will
    allow user to export map as an image. Collected image data will be posted to this file name
    (use ammap/export.php or ammap/export.aspx) -->
    <target></target>
    window in which export file must be called -->
    <x></x>
    data" text -->
    <y></y>
    data" text. If not set, will be aligned to the bottom of flash movie -->
    <color></color>
    background color of "Collecting data" text -->
    <alpha></alpha>
    <text_color></text_color>
    <text_size></text_size>
  </export_as_image>

  <strings>
    <loading_map></loading_map>
    <export_as_image></export_as_image>
    text for right click menu -->
    <collecting_data></collecting_data>
    is displayed while exporting map to an image -->
  </strings>
</settings>

```

<!-- [0] (Number) -->
 <!-- [vertical] (vertical / horizontal)
 <!-- export_as_image feature works
 <!-- [] (filename) if you set filename here,
 <!-- [] (_blank, _top ...) target of a
 <!-- [0] (Number) x position of "Collecting
 <!-- [] (Number) y position of "Collecting
 <!-- [#BBBB00] (hex color code)
 <!-- [0] (0 - 100) background alpha -->
 <!-- [text_color] (hex color code) -->
 <!-- [text_size] (Number) -->
 <!-- [Loading map] (text) -->
 <!-- [Export as image] (text)
 <!-- [Collecting data] (text) this text