

ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ

(DATA MINING)

ΑΠΟΘΗΚΕΣ ΔΕΔΟΜΕΝΩΝ

(DATA WAREHOUSING)

ΥΛΟΠΟΙΗΣΗ ΣΥΣΤΗΜΑΤΟΣ ΓΙΑ ΤΗ ΣΧΕΔΙΑΣΗ ΔΙΑΔΙΚΑΣΙΩΝ ΕΞΑΓΩΓΗΣ, ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΥ ΚΑΙ ΦΟΡΤΩΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΑΠΟΘΗΚΕΣ ΔΕΔΟΜΕΝΩΝ ΜΕ ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΤΟΥ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΙΣΤΟΥ 2

P-miner+: ΕΞΑΤΟΜΙΚΕΥΣΗ ΠΥΛΩΝ ΚΑΤΑΛΟΓΩΝ (PORTALS) ΜΕ ΥΠΟΣΤΗΡΙΞΗ ΔΙΑΔΙΚΑΣΙΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΧΡΗΣΗΣ ΚΑΙ ΑΝΑΛΥΣΗΣ ΠΡΟΦΙΛ ΧΡΗΣΤΩΝ 3

**ΥΛΟΠΟΙΗΣΗ ΣΥΣΤΗΜΑΤΟΣ ΓΙΑ ΤΗ ΣΧΕΔΙΑΣΗ ΔΙΑΔΙΚΑΣΙΩΝ ΕΞΑΓΩΓΗΣ,
ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΥ ΚΑΙ ΦΟΡΤΩΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΑΠΟΘΗΚΕΣ
ΔΕΔΟΜΕΝΩΝ ΜΕ ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΤΟΥ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΙΣΤΟΥ**

ΠΛΗΡΟΦΟΡΙΕΣ: Δ. Σκούτας, 210 772 1436, diskoutas@dbl-lab.ece.ntua.gr

ΠΕΡΙΛΗΨΗ: Η διπλωματική εργασία στοχεύει στην υλοποίηση ενός συστήματος που θα διευκολύνει τον σχεδιαστή ενός συστήματος Αποθήκης Δεδομένων να επιλέγει τις κατάλληλες πηγές και τους απαιτούμενους μετασχηματισμούς για τη φόρτωση δεδομένων στην αποθήκη. Χρησιμοποιώντας μία κατάλληλη οντολογία, το σύστημα θα περιγράφει σημασιολογικά τα σχήματα των πηγών και της αποθήκης και θα συνάγει μεταξύ τους συσχετίσεις.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Οι Αποθήκες Δεδομένων (Data Warehouses) κατέχουν μία σημαντική θέση στις σύγχρονες επιχειρησιακές εφαρμογές, λόγω της ανάγκης για απρόσκοπτη ενοποίηση πολλών ετερογενών πηγών δεδομένων, αποδοτική αποθήκευση και επερώτηση μεγάλου όγκου δεδομένων, καθώς και περιεκτική αναπαράσταση αποτελεσμάτων.

Μια από τις πιο σημαντικές εργασίες κατά τη σχεδίαση ενός συστήματος Αποθήκης Δεδομένων είναι η σχεδίαση διαδικασιών εξαγωγής, μετασχηματισμού και φόρτωσης (Extraction, Transformation and Loading - ETL) δεδομένων στην αποθήκη. Οι διαδικασίες αυτές είναι υπεύθυνες για λειτουργίες όπως: (α) ο εντοπισμός της σχετικής πληροφορίας στην πλευρά της πηγής, (β) η εξαγωγή της πληροφορίας αυτής, (γ) η προσαρμογή και η ενοποίηση πληροφορίας προερχόμενης από πολλαπλές πηγές σε μία κοινή μορφή, (δ) ο καθαρισμός του παραγόμενου συνόλου δεδομένων, με βάση ένα σύνολο κανόνων, (ε) η προώθηση των δεδομένων στην αποθήκη. Η σχεδίαση και υλοποίηση μιας τέτοιας διαδικασίας είναι αρκετά χρονοβόρα και επιρρεπής σε λάθη, συνεπώς η ανάγκη ύπαρξης μεθοδολογιών και εργαλείων που βοηθούν το σχεδιαστή σε αυτό το σκοπό είναι επιτακτική.

Σε αυτή την κατεύθυνση, η χρήση τεχνολογιών του Σημασιολογικού Ιστού (Semantic Web) μπορεί να αποτελέσει ένα χρήσιμο εργαλείο για την αυτοματοποίηση ενός μέρους των παραπάνω λειτουργιών. Συγκεκριμένα, μέσω της χρήσης μίας κατάλληλης οντολογίας, και της απεικόνισης των σχημάτων των πηγών σε αυτή, είναι δυνατό να εξαχθούν με αυτόματο τρόπο συμπεράσματα αναφορικά με τις συσχετίσεις των πηγών και τους μετασχηματισμούς που απαιτούνται για τη μεταφορά δεδομένων μεταξύ τους. Επιπλέον, για τη διαχείριση τόσο δομημένων όσο και ημι-δομημένων πηγών με ενιαίο τρόπο, ακολουθείται μία αναπαράσταση των σχημάτων των πηγών βασισμένη σε γράφους.

Στο σύστημα που θα υλοποιηθεί θα πρέπει να δοθεί έμφαση στη λειτουργικότητα και φιλικότητα προς το χρήστη, μέσω κατάλληλων γραφικών διεπαφών (GUI(s)) και αυτοματισμών.

**P-miner+: ΕΞΑΤΟΜΙΚΕΥΣΗ ΠΥΛΩΝ ΚΑΤΑΛΟΓΩΝ (PORTALS) ΜΕ ΥΠΟΣΤΗΡΙΞΗ
ΔΙΑΔΙΚΑΣΙΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΧΡΗΣΗΣ ΚΑΙ ΑΝΑΛΥΣΗΣ ΠΡΟΦΙΛ
ΧΡΗΣΤΩΝ**

ΠΛΗΡΟΦΟΡΙΕΣ: Π. Μπούρος, 210 772 1436, pbour@dblab.ece.ntua.gr, Θ. Δαλαμάγκας, 210 772 1402, dalamag@dblab.ece.ntua.gr

ΠΕΡΙΛΗΨΗ: Η διπλωματική εργασία στοχεύει στην ανάπτυξη του πρότυπου συστήματος εξατομίκευσης πυλών καταλόγου P-miner+ για συγκεκριμένο πεδίο γνώσης, υποστηριζόμενη από διαδικασίες εξόρυξης δεδομένων χρήσης της πύλης

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java/Apache/PHP/MySQL

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Τα **Portal Catalog Web Sites** είναι πύλες στον Ιστό που παρέχουν πληροφορίες σχετικές με μία θεματολογία. Παραδείγματα αποτελούν οι πύλες με υλικό για μουσεία, πώληση βιβλίων κ.λ.π. Η πληροφορία είναι οργανωμένη σε γράφο με σχέσεις κατηγορίας/υποκατηγορίας μεταξύ θεματικών κατηγοριών. Οι χρήστες πλοηγούνται στο γράφο για να βρουν την πληροφορία που θέλουν.

Στο εργαστήριο έχει ήδη αναπτυχθεί μία πύλη καταλόγου με βάση το περιεχόμενο της πύλης <http://www.dmoz.org/> καθώς και το σύστημα P-miner για την πλήρη διαχείρισή της. Το σύστημα μελετά, με δυνατότητες εξόρυξης γνώσης, τις **πλοηγήσεις χρηστών** στις θεματικές κατηγορίες. Συγκεκριμένα, εντοπίζει περιοχές της πύλης **πολύ δημοφιλείς** ή περιοχές που παρατηρούνται **προβλήματα στην πλοήγηση** των χρηστών και προτείνει στο διαχειριστή τρόπους, με τη μορφή **συντομεύσεων, επιτάχυνσης και βελτίωσης** της πλοήγησης. Επίσης δίνεται η δυνατότητα στο διαχειριστή να ομαδοποιεί τους χρήστες ανάλογα με την **ομοιότητα των πλοηγήσεών τους** εφαρμόζοντας τεχνικές συσταδοποίησης (clustering), και να εξατομικεύει την πύλη.

Η διπλωματική θα επεκτείνει το υπάρχον σύστημα P-miner, προσφέροντας ισχυρότερο μηχανισμό εξατομίκευσης της πύλης πέραν των προτάσεων δημιουργίας συντομεύσεων που προσφέρει το παρόν σύστημα. Θα μελετηθεί η ενσωμάτωση προφίλ χρηστών που θα χρησιμοποιείται από το μηχανισμό πρότασης κατηγοριών/σελίδων που θα επισκέπτεται ο χρήστης. Η διπλωματική αυτή επομένως, θα ασχοληθεί με τα παρακάτω θέματα:

1. Μελέτη και εξοικείωση με το σύστημα P-miner που έχει αναπτυχθεί στον εργαστήριο.
2. Ανάλυση απαιτήσεων εξατομίκευσης πυλών καταλόγων.
3. Σχεδίαση συστήματος P-miner+ για την εξατομίκευση πυλών με υποστήριξη διαδικασιών εξόρυξης γνώσης από πλοηγήσεις χρηστών και επισκέψεις σελίδων.
4. Υλοποίηση συστήματος P-miner+.

Για περισσότερες πληροφορίες δείτε την αναφορά:

<http://www.dbnet.ece.ntua.gr/pubs/uploads/DIPL-2006-10.pdf>