

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ανυφαντής Νικόλαος

Μεταπτυχιακός Φοιτητής

Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

Επόπτης Μεταπτ. Εργασίας: Επικ. Καθηγητής Ι. Τζιτζικας

Δευτέρα, 29/2/2016, 12:00

Αίθουσα Β108, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

**" Διαχείριση Κανόνων Αντιστοιχίσεων για Ολοκλήρωση Πληροφοριών με
Οντολογίες"**

ΠΕΡΙΛΗΨΗ

Ο Σημασιολογικός Ιστός (ΣΙ) είναι μια εξελισσόμενη επέκταση του Παγκόσμιου Ιστού στην οποία το περιεχόμενο μπορεί να εκφραστεί όχι μόνο με φυσική γλώσσα αλλά και με τυπικές γλώσσες (π.χ. RDF/S) που επιτρέπουν την παροχή προηγμένων υπηρεσιών αναζήτησης, διαμοιρασμού και ολοκλήρωσης πληροφορίας. Στις τεχνολογίες αυτές εδράζεται η τάση για δημοσίευση περιεχομένου βάσει των αρχών των Διασυνδεδεμένων Δεδομένων. Δυστυχώς, το μεγαλύτερο μέρος του περιεχομένου που βρίσκεται στο διαδίκτυο βασίζεται σε τεχνολογίες που δεν ανταποκρίνονται στις αρχές του Σημασιολογικού Ιστού. Η πλειοψηφία των δεδομένων που είναι διαθέσιμα είναι σε μορφή σχεσιακών, object-oriented και NoSQL βάσεων δεδομένων. Για το λόγο αυτό έχουν αναπτυχθεί διάφορα εργαλεία και συστήματα τα οποία αναλαμβάνουν την μετατροπή αυτού του περιεχομένου σε συνδεδεμένα δεδομένα. Κάθε σύστημα ακολουθεί μια δική του μέθοδο, καθώς αξίζει να σημειωθεί και η γλώσσα συσχετίσεων R2RML που προτείνεται από το W3C ως επίσημη γλώσσα για την έκφραση αντιστοιχίσεων από σχεσιακές βάσεις

δεδομένων σε RDF σύνολα δεδομένων. Στην εργασία αυτή επικεντρωνόμαστε στην X3ML, μια γλώσσα αντιστοιχίσεων που σχεδιάστηκε από το Ινστιτούτο Πληροφορικής του ΙΤΕ, η οποία είναι αναγνώσιμη από τον άνθρωπο, αρκετά γενική ώστε να καλύπτει τα περισσότερα μοντέλα δεδομένων που χρησιμοποιούνται στις μέρες μας και πιο εκφραστική ως προς τη διαδικασία δημιουργίας URI.

Ο κύριος στόχος αυτής της μεταπτυχιακής εργασίας είναι η παροχή λειτουργιών για τη διαχείριση αντιστοιχίσεων εκφρασμένες στη γλώσσα X3ML. Αρχικά εντοπίσαμε τις κύριες περιπτώσεις χρήσης (business use cases), συγκεκριμένα τις α) *Δημιουργία μιας αντιστοίχισης*, κατά την οποία ο χρήστης θέλει να δημιουργήσει μια αντιστοίχιση από την αρχή, και β) *Βελτίωση μιας αντιστοίχισης* κατά την οποία ο χρήστης επιθυμεί να βελτιώσει μια υπάρχουσα αντιστοίχιση, να ελέγξει την ποιότητά της, να τη συγκρίνει με άλλες και να παρακολουθεί την εξέλιξή της. Βάσει αυτών ορίσαμε τις απαιτήσεις του συστήματος και εν συνεχεία σχεδιάσαμε και υλοποιήσαμε 11 βασικές λειτουργίες. Οι λειτουργίες αυτές αφορούν στην ανάλυση των σχημάτων που εμπλέκονται σε μία αντιστοίχιση, στον υπολογισμό ποσοστών κάλυψης, στη γραφική απεικόνιση και σύγκριση αντιστοιχίσεων, και στην παρακολούθηση της εξέλιξής τους. Συγκεκριμένα, η ανάλυση σχημάτων προσφέρει την δυνατότητα να περιηγηθούμε και να καταλάβουμε καλύτερα το περιεχόμενό τους. Τα ποσοστά κάλυψης παρέχουν μια συνολική εικόνα των στοιχείων που έχουν χρησιμοποιηθεί ενώ υπάρχει και η δυνατότητα εξαίρεσης των στοιχείων που θεωρούνται ανεπιθύμητα για μία αντιστοίχιση. Λειτουργίες γραφικής αναπαράστασης προσφέρονται για την οπτικοποίηση μιας αντιστοίχισης με έμφαση στα συντακτικά στοιχεία της γλώσσας, και επίσης παρέχεται μια τρισδιάστατη αλληλεπιδραστική αναπαράσταση. Όσον αφορά τη σύγκριση, λειτουργίες όπως κειμενική σύγκριση των αντίστοιχων XML αναπαραστάσεων, οπτικοποίηση της διαφοράς με γράφο και σύγκριση παραγόμενων στιγμιότυπων είναι εφικτές. Τέλος, η λειτουργία της παρακολούθησης της εξέλιξης βοηθάει το χρήστη να εντοπίσει τα σημεία που έχουν αλλάξει κατά την πάροδο του χρόνου. Συνοψίζοντας, η καινοτομία έγκειται στον διαδραστικό χαρακτήρα όλων των λειτουργιών και στην τρισδιάστατη απεικόνιση των αντιστοιχίσεων, καθιστώντας την παρούσα εργασία την πρώτη μελέτη εκτεταμένης διαχείρισης αντιστοιχίσεων στο συγκεκριμένο τομέα.

Anyfantis Nikolaos

M.Sc. Thesis

Computer Science Department

University of Crete

Master's Thesis Supervisor: Assistant Professor I. Tzitzikas

Monday, 29/2/2016, 12:00

Room B108, Computer Science dept., University of Crete

“Mappings Management for Ontology-based Integration”

ABSTRACT

The Semantic Web (SW) is an evolving extension of the World Wide Web in which the content can be expressed not only in natural language, but also in formal languages (e.g. RDF/S) that can be read and used by software agents, permitting them to find, share and integrate information more easily. Semantic Web technologies form the foundation for publishing data according to the principles of Linked Data. Unfortunately, most web content is based on technologies that do not meet the principles of the Semantic Web. The majority of the data is available in relational, object-oriented and NoSQL databases. For this reason several systems have been developed for converting and publishing content as linked data. Each system follows its own method, while it is worth noting the R2RML mapping language proposed by W3C as an official language for expressing mappings from relational databases in RDF datasets. In this work we focus on X3ML language; a mapping formalism designed by the Institute of Computer Science of FORTH, which is human readable, generic enough to cover most of the data models used nowadays and more explicit with the URI generating process.

The main objective of this MSc thesis is to provide a complete analysis of mappings expressed in X3ML language. Firstly, we identify the main business use cases, a) *Create a Mapping* in which the actor wants to generate a mapping from scratch, b) *Improve a Mapping* in which the actor needs to improve it, check its quality, compare with others and monitor its evolution. Based on these, we defined the system requirements and then designed and implemented 11 basic functionalities. These functionalities refer to the analysis of the schemas involved in a particular mapping, the calculation of coverage metrics, the graphical representation and comparison of mappings, and the monitoring of their evolution. More specifically, the analysis of schemas offers the opportunity of navigating and better understanding the content. The coverage metrics give an overview of the mapped data and allow excluding elements that are considered undesirable for a mapping. For the graphical representation, we introduce functionalities that visualize a mapping, focusing on syntactic elements of the language, and provide a three-dimensional interactive representation. As regards comparisons, several functionalities are offered such as textual comparison of mappings in XML format, representation of differences through graph and comparison of generated instances. Finally, the functionality of tracking the evolution of a mapping helps the user to identify the basic points that have changed over time. All in all, the main novelty of this work is the interactivity of all functionalities and 3D visualization of mappings.