

eXaMpLe: Αξιοποιώντας την τεχνολογία XML στη διαχείρισης της πληροφορίας

eXaMpLe: Exploiting XML technology in information management

Λίνα Μπουντούρη, Νικόλαος Τσόκας, Μιχάλης Σφακάκης
Εθνικό Κέντρο Τεκμηρίωσης (ΕΚΤ) / ΕΙΕ
Βασ. Κωνσταντίνου 48
11635, Αθήνα
{bountouri, tsokas, [msfaka](mailto:msfaka@ekt.gr)}@ekt.gr

Lina Bountouri, Nikolaos Tsokas, Michalis Sfakakis
National Documentation Center
{bountouri, tsokas, [msfaka](mailto:msfaka@ekt.gr)}@ekt.gr

Περίληψη

Οι σύγχρονες τάσεις στον τομέα των τεχνολογιών πληροφόρησης επηρεάζουν τους φορείς που παράγουν και διαχειρίζονται την πληροφορία αναφορικά με τα πρότυπα που χρησιμοποιούν και τα εργαλεία διαχείρισής τους. Η Επεκτάσιμη Γλώσσα Σήμανσης (eXtensible Markup Language - XML), γλώσσα για την αποτύπωση και κωδικοποίηση δεδομένων και μεταδεδομένων, ήλθε να αποτελέσει ένα νέο ευέλικτο εργαλείο για την περιγραφή και την εφαρμογή των προτύπων. Αποτέλεσμα της εξέλιξης αυτής είναι ο σταδιακός επαναπροσδιορισμός των παραδοσιακών μορφών προτύπων ως συνόλων κανόνων με τη χρήση της τεχνολογίας XML, χωρίς ωστόσο να διαφοροποιείται το περιεχόμενο ή ο σκοπός τους.

Το σύνολο των κανόνων που ρυθμίζουν την έγκυρη περιγραφή μιας οντότητας σε ένα έγγραφο XML έχει επικρατήσει να αποτυπώνεται στο λεγόμενο Σχήμα XML. Κατά αυτόν τον τρόπο σχηματίζονται ξεχωριστά Σχήματα XML για την ορθή περιγραφή οντοτήτων από διαφορετικούς επιστημονικούς ή άλλους τομείς και γνωστικά αντικείμενα, όπως αυτούς της βιβλιοθηκονομίας και της αρχειονομίας.

Την παραπάνω δυνατότητα της Επεκτάσιμης Γλώσσας Σήμανσης αξιοποιεί η εφαρμογή eXaMpLe, που αναπτύχθηκε και εξελίσσεται στο Εθνικό Κέντρο Τεκμηρίωσης (ΕΚΤ) με σκοπό να καλύψει τις νέες ανάγκες των φορέων πληροφόρησης σε ό,τι αφορά την αναδυόμενη στον κλάδο τεχνολογία. Δίνοντας, μέσω Σχημάτων XML, τη δυνατότητα στο χρήστη να καθορίσει το σύνολο των κανόνων που διέπουν την περιγραφή των οντοτήτων, τον οδηγεί στην έγκυρη αποτύπωση πληροφοριών, είτε πρόκειται για θεματικές ενότητες του ενδιαφέροντος συγκεκριμένων ομάδων χρηστών, είτε για διαδεδομένα διεθνή πρότυπα, με εύχρηστο τρόπο.

Η εφαρμογή eXaMpLe, μέσω της χρήσης Σχημάτων XML, έχει την ικανότητα να λειτουργήσει ως εργαλείο επεξεργασίας εγγράφων που υπακούουν σε πρότυπα όπως αυτά που έχουν αναπτυχθεί για χρήση από φορείς πληροφόρησης (EAD - Encoded Archival Description, MARCXML, MODS κλπ). Ο χρήστης καθοδηγείται διαρκώς σε εισαγωγή δεδομένων που συνάδουν τόσο με το ενεργό Σχήμα, όσο και με τα ήδη καταγεγραμμένα δεδομένα στο έγγραφο υπό επεξεργασία. Η επεξεργασία εγγράφων ενισχύεται από την ύπαρξη εγγράφων – οδηγών, τα οποία οδηγούν στην αυτόματη εισαγωγή πληροφοριών που αφορούν κοινά δεδομένα ομάδας

εγγράφων, ώστε να περιορίζεται η επανάληψη εισαγωγής όμοιων πληροφοριών από το χρήστη. Στην ευχρηστία της εφαρμογής συντείνει η δυνατότητα απόκρυψης της δομής του Σχήματος μέσω της διεπαφής εργασίας, η οποία καθίσταται φιλική ακόμα και σε χρήστες χωρίς εξοικείωση με τη τεχνική ορολογία του εκάστοτε γνωστικού αντικειμένου.

Η περιγραφή των οντοτήτων και η αποτύπωση των πληροφοριών σε ψηφιακή μορφή με τη βοήθεια της εφαρμογής συμπληρώνεται από το μηχανισμό προβολής και διάθεσής τους, ο οποίος εκμεταλλεύεται το φυσικό προσανατολισμό της γλώσσας XML στις τεχνολογίες του Διαδικτύου. Η εφαρμογή eXaMpLe επιχειρεί να καλύψει το διαφαινόμενο κενό μεταξύ των νέων τεχνολογιών, που ήδη επικρατούν στον επιστημονικό τομέα της διαχείρισης της πληροφορίας, των καθιερωμένων διεθνών προτύπων και των παραδοσιακών μεθόδων εργασίας.

Η παρούσα εργασία, πραγματοποιώντας αρχικά μία επισκόπηση στις νέες τάσεις και τεχνολογίες που επικρατούν στο χώρο αυτό, παρουσιάζει την εφαρμογή eXaMpLe, της οποίας τα λειτουργικά χαρακτηριστικά αναδεικνύονται μέσα από την κωδικοποίηση πληροφοριών με χρήση των προτύπων EAD και MARCXML.

Λέξεις-κλειδιά: XML εργαλεία, διαδικτυακή δημοσίευση, κωδικοποίηση μεταδεδομένων, επεξεργαστής δένδρου.

Abstract

The developments of information technology have always affected organizations that create and manage information, particularly in terms of the standards used and the tools for managing them. The eXtensible Markup Language (XML), a language for representing and encoding data and metadata, came to be established as a new flexible tool for describing and managing these standards. As a result of the above came the gradual redefinition of standards from their traditional form into XML syntax and technology, still without affecting their content or scope.

The set of rules that have to be applied, in order for data to be encoded in a valid and accurate manner, are defined in XML Schemas. Thus, a diversity of XML Schemas has been produced so as to correctly describe data from different scientific disciplines and domains, like librarianship, archival science, etc.

It is exactly this feature of XML that eXaMpLe application takes advantage of. eXaMpLe, which is being developed by the National Documentation Centre (NDC), aims at fulfilling the needs of informational organizations concerning emerging technologies. The application is giving users the opportunity to set the XML Schema they want to implement so as to describe data consistently, no matter if it is for the interest of a certain group of users, or about established international standards.

Comprehensively, through the use of XML Schemas, eXaMpLe functions as a tool for processing documents that conform to explicit standards, like the ones developed for use from the informational organizations (EAD, MARCXML, MODS etc). The user is continuously instructed to insert data that matches not only the XML Schema in use but also the actual document being processed. The editing of documents is also enforced by the use of templates, which allow for the automatic registration of information shared between groups of documents, so that unneeded repeated work is eliminated. User-friendliness can be also enhanced by hiding the underlying document structure from the actual interface, which in turn becomes easy to use even for those not familiar with the domain terminology.

The description of data in a digital form with eXaMpLe is complemented by a mechanism for preview and publishing that takes advantage of the orientation of XML towards Internet technologies. The application attempts to fill the apparent gap between the new technologies that are increasingly being employed in information management, the authoritative international standards and the traditional working methods.

The present study, reviewing at first place the new trends and technologies in information management, presents application eXaMpLe, the operational features of which are illustrated with the use of encoding examples based on the EAD and the MARCXML standards.

Keywords: XML tools, internet publishing, metadata encoding, tree editor.

1. Πρότυπα και Επεκτάσιμη Γλώσσα Σήμανσης (eXtensible Markup Language-XML)

Οι φορείς πληροφόρησης με σκοπό τη μεταξύ τους επικοινωνία και ανταλλαγή πληροφοριών χρησιμοποιούν κοινά πρότυπα για την επεξεργασία και τη διαχείριση των δεδομένων τους. Τα πρότυπα κωδικοποίησης δεδομένων προσφέρουν ενιαίο λεξιλόγιο και συντακτικό, αποτελώντας συνεπώς κοινό σημείο αναφοράς για λειτουργίες που πραγματοποιούνται μεταξύ ομάδων χρηστών ή φορέων, όπως η ανταλλαγή δεδομένων, η ταυτόχρονη αναζήτηση σε διαφορετικές πηγές, ο διαδανεισμός και η χρήση κοινών εργαλείων.

Η Επεκτάσιμη Γλώσσας Σήμανσης εμφανίστηκε το 1998 και έθεσε νέες βάσεις αναφορικά με την έκφραση, τη δόμηση και την ανταλλαγή της πληροφορίας (World Wide Web Consortium, 2004a). Καθώς οι σύγχρονες τάσεις και εξελίξεις στην επιστήμη της διαχείρισης και της τεχνολογίας της πληροφορίας επηρεάζουν σημαντικά τους φορείς που την καταγράφουν και τη διαχειρίζονται, η XML σύντομα υιοθετήθηκε από τους τελευταίους τόσο στην περιγραφή των προτύπων όσο και στη λειτουργική δομή των αντίστοιχων εργαλείων.

Η XML αρχικά σχεδιάστηκε με σκοπό να αποτελέσει μία μέση λύση μεταξύ της πολυπλοκότητας της SGML (World Wide Web Consortium, 2004b) και της επιφανειακότητας της HTML (World Wide Web Consortium, 2004c) ως γλωσσών σήμανσης εγγράφων που περιέχουν δομημένη πληροφορία. Η αυστηρή σύνταξη, η ευελιξία και ο πλουραλισμός που προσφέρει στην έκφραση των δεδομένων τράβηξε την προσοχή των επιστημόνων της πληροφόρησης, με αποτέλεσμα η XML να αποτελεί πλέον κοινή βάση για την έκφραση και τη δόμηση των προτύπων κωδικοποίησης δεδομένων. Σήμερα, τα περισσότερα πρότυπα έχουν ήδη καταγραφεί σε μορφή XML από όποια διάταξη κι αν προέρχονται, για παράδειγμα το TEI SGML έχει περάσει στη νέα του έκδοση σε XML μορφή (TEI Consortium, 2003). Το περιεχόμενο και ο σκοπός τους, ωστόσο, δεν έχουν σε καμία περίπτωση αλλοιωθεί, καθώς οι νέες δομές περιγράφουν τα ίδια ακριβώς δεδομένα. Σημαντικό είναι επίσης το γεγονός ότι η ομοιομορφία που προσφέρει η XML επιτρέπει στους φορείς πληροφόρησης τη χρήση κοινών εργαλείων για την επεξεργασία και διαχείριση δεδομένων σχετιζομένων με διαφορετικά πρότυπα.

Η μεταφορά των προτύπων κωδικοποίησης δεδομένων σε XML μορφή παρουσιάζει σημαντικά πλεονεκτήματα, τα οποία μεταξύ άλλων συντείνουν και στην αποτελεσματικότερη χρήση των ίδιων των περιγραφόμενων δεδομένων.

Κατ' αρχήν η XML παρουσιάζει το πλεονέκτημα της αξιοποίησης μεγάλου αριθμού ήδη υπαρχόντων εργαλείων επεξεργασίας και προβολής δεδομένων. Η δυνατότητα αυτή προκύπτει από την ανεξαρτησία της γλώσσας από το χρησιμοποιούμενο υλικό και λογισμικό.

Χαρακτηριστικό είναι ότι ένα έγγραφο XML μπορεί να παρουσιαστεί σε ένα φυλλομετρητή του Παγκόσμιου Ιστού χωρίς ενδιάμεση επεξεργασία. Επιπλέον, ποικίλες εφαρμογές μπορούν να χειριστούν ένα XML έγγραφο, όπως επεξεργαστές HTML εγγράφων, (π.χ. οι πρόσφατες εκδόσεις του Dreamweaver) και επεξεργαστές κειμένου.

Όντας ανεξάρτητη από πλατφόρμες, όπως προαναφέρθηκε, η XML αποτελεί ισχυρή βάση για τη διατήρηση της πληροφορίας στο χρόνο. Παράλληλα συνοδεύεται από υποστηρικτικά πρότυπα, όπως τις XPath εκφράσεις, που επιτρέπουν πρόσβαση στα δεδομένα με ευέλικτο τρόπο (π.χ. σε επίπεδο υποπεδίου μιας βιβλιογραφικής εγγραφής), και τα XSL έγγραφα που μετατρέπουν και εμφανίζουν τα δεδομένα σε διάφορες μορφές (π.χ. PDF, HTML, κτλ).

Τέλος, εκφράζοντας σε XML τα πρότυπα κωδικοποίησης δεδομένων διευκολύνεται η μεταξύ τους διαλειτουργικότητα τόσο σε επίπεδο δομής όσο και σε σημασιολογικό επίπεδο (schematic and semantic interoperability).

Οι βασικοί κανόνες δόμησης των XML εγγράφων είναι δυνατό να περιγραφούν μέσα από ένα Document Type Definition (DTD) ή από ένα XML Σχήμα. Το XML Σχήμα, διαθέτοντας πιο ευέλικτη δομή και ισχυρότερη εκφραστική μέθοδο, τείνει να υπερισχύσει του DTD και καθιερώνεται σταδιακά σε βάρος του. Η διαφορά ανάμεσα στις δύο μεθόδους διαφαίνεται έντονα στην περίπτωση που ζητείται να εκφραστούν περίπλοκοι δομικοί κανόνες για το περιεχόμενο ενός XML εγγράφου. Για παράδειγμα, μέσα σε ένα XML Σχήμα μπορούν να χρησιμοποιηθούν κανόνες από άλλα XML Σχήματα, καθώς και να τεθούν περιορισμοί σχετικά με το μήκος, τον τύπο και τη μορφή του περιεχομένου των στοιχείων και των γνωρισμάτων (Harold 2002: 225). Παρόλο που αναπτύχθηκαν διάφοροι κανόνες για τη σύνταξη XML Σχημάτων, όπως το RELAX NG (RELAX NG Technical Committee, 2003) και το Schematron (Jelliffe, 2003), η μορφή που έχει επικρατήσει είναι εκείνη που ορίζεται από το W3C (World Wide Web Consortium, 2004d).

2. eXaMpLe

Το Εθνικό Κέντρο Τεκμηρίωσης, στην προσπάθειά του να ανταποκριθεί στις τεχνολογικές εξελίξεις και να καλύψει τις ανάγκες των φορέων πληροφόρησης, αναπτύσσει την εφαρμογή eXaMpLe για τη δημιουργία και την επεξεργασία XML εγγράφων.

2.1. Γενικά

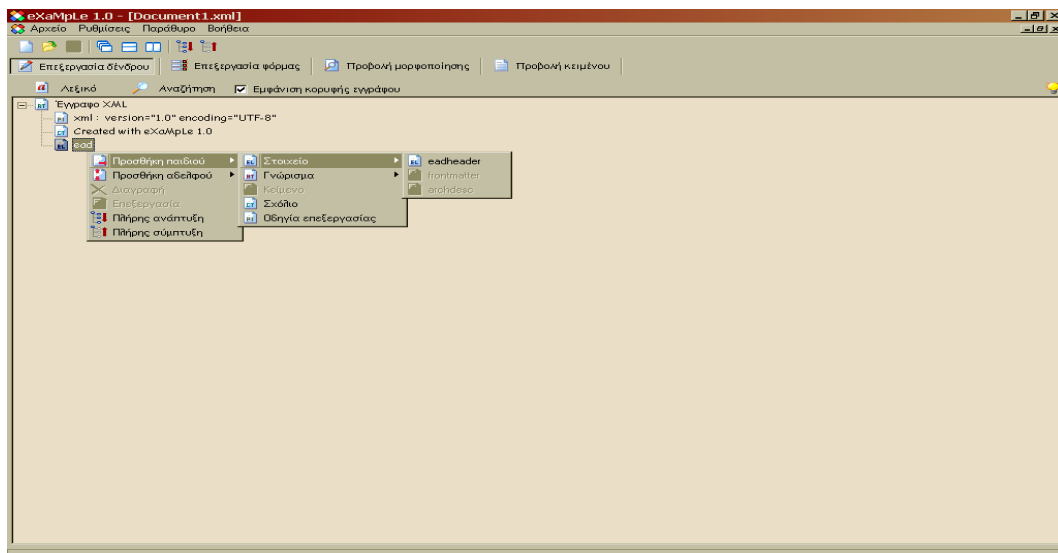
Η εφαρμογή eXaMpLe επεξεργάζεται τα XML έγγραφα στη φυσική τους μορφή, προβάλλοντάς τα σε μορφή δένδρου (tree editor). Κατ' αυτόν τον τρόπο ο χρήστης έχει τη δυνατότητα να εισάγει τα δεδομένα αντιλαμβανόμενος την πραγματική δομή και ιεραρχία μεταξύ τους.

Η βασική λειτουργία της εφαρμογής είναι η δημιουργία και επεξεργασία XML εγγράφων. Το eXaMpLe, αξιοποιώντας τις δυνατότητες των XML Σχημάτων, επιτρέπει στο χρήστη να ορίσει το σύνολο των κανόνων που διέπουν το υπό επεξεργασία έγγραφο. Ο χρήστης μπορεί να επιλέξει ένα συγκεκριμένο XML Σχήμα, όπως το MARCXML (Library of Congress, 2004a) ή το MODS (Library of Congress, 2004b), και να δημιουργήσει XML έγγραφα βασισμένα σε αυτό.

Όσον αφορά το περιβάλλον εργασίας, ο χρήστης μπορεί να επιλέξει μεταξύ δύο μορφών επεξεργασίας: την Επεξεργασία δένδρου (βλ. Εικόνα 1) και την Επεξεργασία φόρμας (βλ. Εικόνα 2).

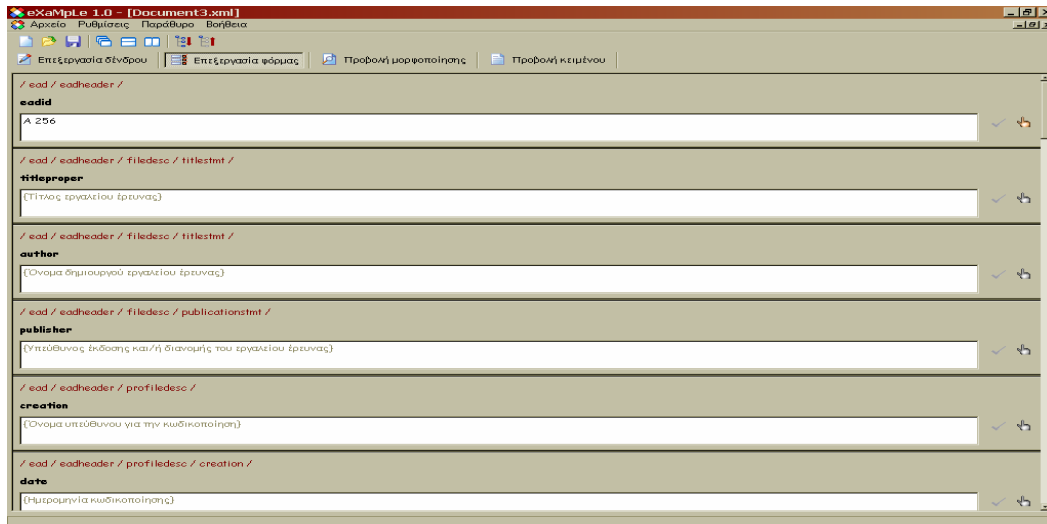
Η Επεξεργασία δένδρου επιτρέπει την εισαγωγή δεδομένων κατευθείαν στη δενδροειδή δομή του XML εγγράφου. Η εισαγωγή δεδομένων μέσω του δένδρου διευκολύνει το χρήστη, διότι του παρέχει διαφωτιστική εικόνα της πολύ-επίπεδης ιεραρχικής δομής του XML εγγράφου που διαχειρίζεται. Μέσω της επεξεργασίας δένδρου, ο χρήστης προσθέτει κόμβους (στοιχεία, γνωρίσματα, κείμενο, σχόλια και οδηγίες επεξεργασίας), σύμφωνα με το εκάστοτε XML Σχήμα που αυτός έχει επιλέξει να χρησιμοποιείται. Ορίζοντας ένα συγκεκριμένο Σχήμα, ο χρήστης καθοδηγείται κατά την εισαγωγή των στοιχείων, των γνωρισμάτων, καθώς και του κειμένου αυτών. Πιο αναλυτικά, κατά την προσθήκη νέων κόμβων, η εφαρμογή, λαμβάνοντας υπόψη το εκάστοτε επιλεγμένο XML Σχήμα, εμφανίζει ενεργά τα διάφορα στοιχεία ή γνωρίσματα που επιτρέπεται να χρησιμοποιηθούν σύμφωνα με την εκάστοτε κατάσταση του εγγράφου και ανενεργά εκείνα που δεν επιτρέπεται. Επιπλέον, αναφορικά με το κείμενο των γνωρισμάτων και των στοιχείων, συχνά προτείνονται στο χρήστη συγκεκριμένες τιμές προς επιλογή.

Η Επεξεργασία φόρμας παρέχει στο χρήστη τη δυνατότητα επεξεργασίας ενός XML εγγράφου μέσω ενός συνόλου πλαισίων κειμένου, στα οποία καταχωρείται το κείμενο των στοιχείων.



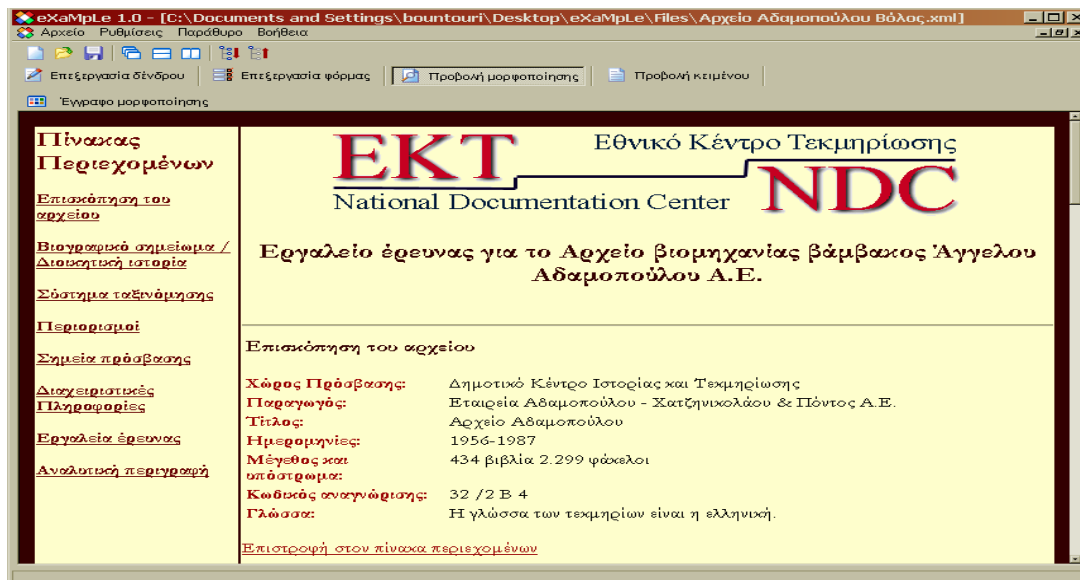
Εικόνα 1. Επεξεργασία δένδρου

Η συγκεκριμένη δυνατότητα επεξεργασίας εμφανίζεται στο χρήστη στην περίπτωση που ορίσει ένα έγγραφο – οδηγό (template). Τα έγγραφα – οδηγοί είναι XML έγγραφα μέσα από τα οποία ορίζονται τα στοιχεία προς επεξεργασία που θα εμφανιστούν στη φόρμα. Παράλληλα, μέσα σε αυτά μπορούν να οριστούν επιπλέον κόμβοι, οι οποίοι και καταχωρούνται αυτόματα στο δένδρο του XML εγγράφου, ώστε να περιορίζεται η επανάληψη εισαγωγής όμοιων πληροφοριών από το χρήστη (βλ. Παρ. 2.2.2).

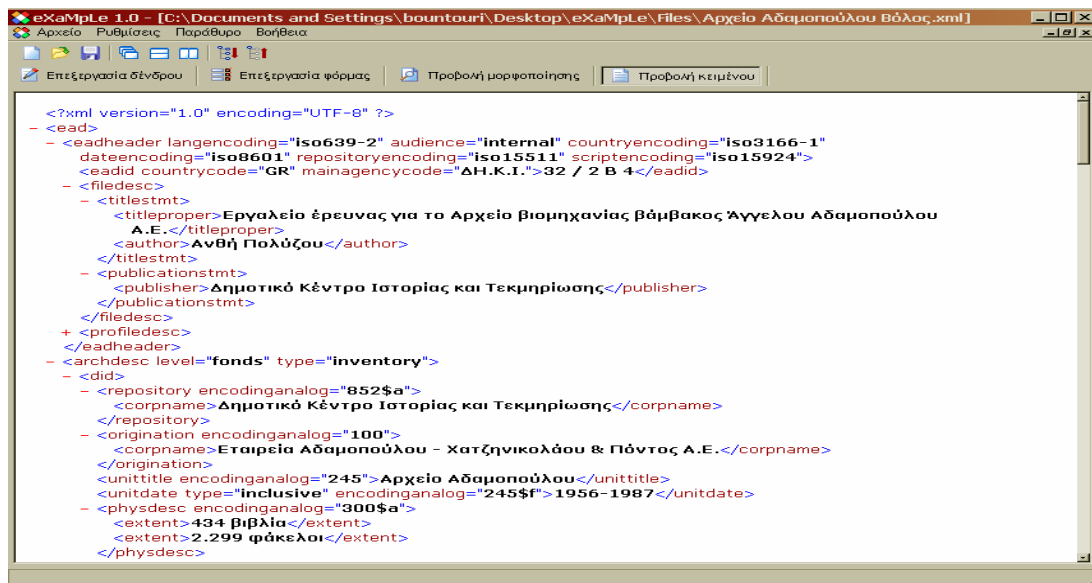


Εικόνα 2. Επεξεργασία φόρμας

Η εφαρμογή eXaMple παρέχει στο χρήστη τη δυνατότητα προβολής του XML εγγράφου σε μορφή HTML με τη βοήθεια ενός XSLT εγγράφου (World Wide Web Consortium, 1999a), το οποίο ορίζεται από το χρήστη (βλ. Εικόνα 3). Συγχρόνως, το eXaMple χρησιμοποιεί ένα φυλλομετρητή του παγκόσμιου ιστού (Internet Explorer 5+), προκειμένου να προβάλλει το έγγραφο σε μορφή κειμένου (βλ. Εικόνα 4).



Εικόνα 3. Προβολή μορφοποίησης



Εικόνα 4. Προβολή κειμένου

2.2. Τεχνικές λεπτομέρειες

2.2.1. Document Object Model (DOM)

Η εφαρμογή υλοποιήθηκε με χρήση της τεχνολογίας DOM (World Wide Web Consortium, 2004e), σύμφωνα με την οποία το XML έγγραφο αναπαρίσταται ως δένδρο. Ως κορυφή του δένδρου τοποθετείται το αρχικό στοιχείο του XML εγγράφου. Οι υπόλοιποι κόμβοι του εγγράφου (στοιχεία, γνωρίσματα, σχόλια κλπ) τοποθετούνται ως κλαδιά κάτω από την κορυφή του δένδρου, ούτως ώστε κάθε ένας από αυτούς να βρίσκεται στο αμέσως κατώτερο επίπεδο από εκείνον, στον οποίο περιέχεται. Κατ' αυτόν τον τρόπο, ένας κόμβος που αντιπροσωπεύει στοιχείο μπορεί να έχει ως «παιδιά» άλλα στοιχεία, γνωρίσματα κλπ. Αντίθετα, οι κόμβοι που δεν αντιπροσωπεύουν στοιχεία, όπως τα γνωρίσματα και τα σχόλια, δεν μπορούν να είναι «γονείς» κανενός άλλου κόμβου σε κατώτερο επίπεδο. Η αποθήκευση του XML εγγράφου στη μνήμη του υπολογιστή σε DOM μορφή γίνεται αυτόματα κατά τη διάρκεια ανάγνωσης του εγγράφου από το δίσκο.

Με τη χρήση του μοντέλου DOM η εφαρμογή μπορεί να μετακινείται γρήγορα μεταξύ κόμβων και να ανακτά εύκολα πληροφορίες για το XML έγγραφο, καθώς όλοι οι κόμβοι είναι ιεραρχικά συνδεδεμένοι μεταξύ τους. Για παράδειγμα, μπορεί με απλό τρόπο να συλλεχθεί το σύνολο των γνωρισμάτων ενός στοιχείου χωρίς να είναι απαραίτητη η ανάγνωση του αρχείου από το δίσκο και η ανάλυσή του σε μορφή κειμένου (parsing).

Η τεχνολογία DOM προσφέρει ένα σύνολο μεθόδων επεξεργασίας XML εγγράφων, οι οποίες παραμένουν ανεξάρτητες από το λογισμικό ή την πλατφόρμα, στην οποία αναπτύσσεται η εφαρμογή. Στα πλαίσια του eXaMple χρησιμοποιήθηκε η υλοποίηση MSXML του μοντέλου DOM που διατίθεται από τη Microsoft (Microsoft, 2004) και η υλοποίηση Xerces που διατίθεται από την Apache (Apache Software Foundation, 2003).

Με άλλα λόγια, η μορφή DOM προσεγγίζει τη φυσική δομή ενός XML εγγράφου δίνοντάς του ταυτόχρονα μια παραστατική διάσταση. Η δένδροειδής εικόνα του εγγράφου που εμφανίζεται στην Επεξεργασία δένδρου του eXaMple δεν είναι παρά η DOM μορφή του προβαλλόμενη στην επιφάνεια εργασίας.

2.2.2. Έγγραφα - οδηγοί (templates)

Οι οδηγοί είναι XML έγγραφα, τα οποία χρησιμεύουν ως γνώμονας για τη δημιουργία και τη συμπλήρωση άλλων XML εγγράφων. Υπακούουν σε κάποιο XML Σχήμα και περιέχουν τις πληροφορίες εκείνες, οι οποίες θεωρούνται απαραίτητες για μία συγκεκριμένη εφαρμογή. Με τη χρήση εγγράφων – οδηγών αποφεύγεται η επανάληψη της εισαγωγής των ίδιων δεδομένων σε ομάδα εγγράφων, αφού κάθε νέο έγγραφο αυτόματα αποκτά ορισμένα σταθερά στοιχεία, γνωρίσματα ή κόμβους άλλου τύπου. Για παράδειγμα, κοινό στοιχείο μεταξύ εγγράφων μπορεί να αποτελεί το όνομα του δημιουργού και η ημερομηνία επεξεργασίας τους. Οι κόμβοι αυτοί, εφόσον το επιλεγμένο Σχήμα προβλέπει την ύπαρξή τους, μπορούν να περιέχονται σε ένα έγγραφο – οδηγό και να συμπληρώνονται αυτόματα σε κάθε νέο έγγραφο.

Εκτός της αυτόματης συμπλήρωσης δεδομένων, όπως προαναφέρθηκε η χρήση οδηγών παρέχει τη δυνατότητα επεξεργασίας των XML εγγράφων διαμέσου μίας πιο απλής σε χρήση επιφάνειας εργασίας, η οποία αποτελείται από πλαίσια κειμένου. Στα πλαίσια κειμένου εισάγεται το κείμενο των στοιχείων που προβλέπονται από το έγγραφο – οδηγό. Τα στοιχεία αυτά προστίθενται δυναμικά στο νέο έγγραφο κατά τη συμπλήρωση των αντίστοιχων πλαισίων.

Η δημιουργία εγγράφων – οδηγών μπορεί να πραγματοποιηθεί κατευθείαν από το χρήστη, ενώ η σύνταξή τους αρκεί να ακολουθεί τους εξής κανόνες: α) οι οδηγοί είναι απαραίτητο να υπακούουν σε κάποιο XML Σχήμα και να είναι έγκυρα έγγραφα σύμφωνα με αυτό και β) τα κείμενα που θα συμπληρωθούν μελλοντικά από το χρήστη και που θα εμφανίζονται με τη μορφή πλαισίων στην Επεξεργασία φόρμας θα πρέπει στο έγγραφο – οδηγό να περιγραφούν ως οδηγίες επεξεργασίας (processing instructions).

Δημιουργώντας ένα νέο XML έγγραφο με χρήση οδηγού, το έγγραφο αποκτά μορφή όμοια με του τελευταίου εφόσον συμπληρωθούν όλα τα πλαίσια κειμένου στην επεξεργασία φόρμας. Η μόνη διαφορά έγκειται στο ότι το νέο έγγραφο περιέχει κόμβους κειμένων στη θέση των οδηγιών επεξεργασίας του οδηγού. Σύμφωνα με τα παραπάνω, εφόσον ο οδηγός είναι έγκυρο έγγραφο, το νέο έγγραφο που θα προκύψει θα είναι και αυτό έγκυρο.

Στην περίπτωση κατά την οποία δεν συμπληρωθούν όλα τα πλαίσια κειμένου το νέο έγγραφο είναι πιθανό να μην είναι έγκυρο, ωστόσο ο χρήστης έχει τη δυνατότητα να συνεχίσει την επεξεργασία του άλλη στιγμή μέσα, είτε από την επεξεργασία φόρμας είτε από την επεξεργασία δένδρου.

2.2.3. Λεξικό

Το λεξικό είναι ένα αρχείο, το οποίο μπορεί να γραφεί σε έναν απλό επεξεργαστή κειμένου (π.χ. Notepad). Επιτρέπει τη μετάφραση ή επεξήγηση όρων στη δενδροειδή μορφή του XML εγγράφου, καθώς και στην Επεξεργασία φόρμας. Η σύνταξή τους είναι η εξής:

όρος προς επεξήγηση – επεξήγηση

Παράδειγμα λεξικού:

<abbr> - Abbreviation
<abstract> - Abstract
<accessrestrict> - Conditions Governing Access
<accruals> - Accruals
<acqinfo> - Acquisition Information
<address> - Address

2.2.4. Έλεγχος εγκυρότητας (Validation)

Η εγκυρότητα του εγγράφου ως προς το προεπιλεγμένο XML Σχήμα ελέγχεται αυτόματα κατά το άνοιγμά του. Στην περίπτωση κατά την οποία το ανοιγμένο έγγραφο δεν είναι έγκυρο, τα σφάλματα αναφέρονται στο χρήστη μέσω σχετικού μηνύματος. Ωστόσο, ακόμα και στην περίπτωση αυτή ο χρήστης μπορεί να προχωρήσει στη επεξεργασία του εγγράφου.

2.3. Χρηστικότητα

Λαμβάνοντας υπόψη ότι η XML και οι υπόλοιπες σχετικές με αυτή τεχνολογίες είναι πρόσφατες και ότι στην ελληνική πραγματικότητα δεν υπάρχει μέχρι στιγμής η κατάλληλη υποδομή, η ομάδα ανάπτυξης βιβλιογραφικών εφαρμογών του Εθνικού Κέντρου Τεκμηρίωσης δίνει ιδιαίτερη έμφαση στη χρηστικότητα του συγκεκριμένου εργαλείου.

Με γνώμονα τη χρηστικότητα δημιουργήθηκαν δύο διαφορετικές επιλογές σχετικά με την επεξεργασία των XML εγγράφων. Η Επεξεργασία δένδρου είναι πιο κατάλληλη για χρήστες εξοικειωμένους με την τεχνολογία XML και τη δομή του επιλεγμένου XML Σχήματος, ενώ αντίθετα η Επεξεργασία φόρμας απευθύνεται σε λιγότερο έμπειρους χρήστες.

Συγχρόνως, η χρηστικότητα της εφαρμογής eXaMpLe έγκειται σε μεγάλο βαθμό στον ορισμό ρυθμίσεων από την πλευρά του χρήστη. Βασικό εξάλλου χαρακτηριστικό της εφαρμογής είναι η δυνατότητα εναλλαγής των XML Σχημάτων. Ο χρήστης μπορεί να εργαστεί με ποικίλα σχήματα στην ίδια επιφάνεια εργασίας, ανάλογα με τις ανάγκες του, επιλέγοντας την αντίστοιχη ρύθμιση. Υπάρχει επίσης η δυνατότητα ορισμού του στοιχείου που χρησιμοποιείται ως αρχικό (root element) κάθε φορά που δημιουργείται νέο έγγραφο.

Η καθοδήγηση που παρέχει για εισαγωγή στοιχείων και γνωρισμάτων, τα οποία συμφωνούν με το ορισμένο XML Σχήμα, διευκολύνει το χρήστη στη δημιουργία των εγγράφων επιταχύνοντας σημαντικά την εργασία του, ειδικά στην περίπτωση που αυτός δε γνωρίζει τη δομή του χρησιμοποιούμενου σχήματος. Επιπλέον, βοηθά στη δημιουργία σωστά διαμορφωμένων (well – formed) και έγκυρων (valid) εγγράφων.

Στη χρηστικότητα του εργαλείου συντείνει και η χρήση του λεξικού. Καθώς οι όροι που χρησιμοποιούνται για να οριστούν τα στοιχεία μέσα σε ένα XML έγγραφο είναι συχνά δυσνόητοι, το eXaMpLe δίνει τη δυνατότητα αυτοί να αντικατασταθούν από περιγραφικούς όρους. Συγκεκριμένα, δίνεται στο χρήστη η δυνατότητα να ορίσει ένα λεξικό, σύμφωνα με το οποίο, όπως προαναφέρθηκε, πραγματοποιείται η συσχέτιση και η αντικατάσταση στο περιβάλλον εργασίας των δυσνόητων όρων είτε πρόκειται για την Επεξεργασία δένδρου είτε για την Επεξεργασία φόρμας.

2.4. XML και Διαδίκτυο - XIndexer

Η κωδικοποίηση δεδομένων σε XML είναι φυσικά προσανατολισμένη στη διάθεση του περιεχομένου τους μέσω του Διαδικτύου. Με σκοπό την προβολή των κωδικοποιημένων δεδομένων, το Εθνικό Κέντρο Τεκμηρίωσης έχει αναπτύξει μία συνοδευτική του eXaMpLe εφαρμογή, το XIndexer.

Η εφαρμογή αυτή μετατρέπει σύνολα XML εγγράφων σε HTML σελίδες, οι οποίες παράλληλα ευρετηριάζονται. Η διαδικασία μετατροπής βασίζεται στον ορισμό ενός συγκεκριμένου στοιχείου ως κλειδιού για την αλφαβητική ταξινόμηση του συνόλου των εγγράφων, καθώς και ενός εγγράφου μορφοποίησης (XSLT) για την προβολή τους. Το στοιχείο αυτό αναζητείται στα XML έγγραφα και, εφόσον εμφανίζεται σε αυτά, χρησιμοποιείται ως σύνδεσμος από την τελική HTML σελίδα παρουσίασης στο αντίστοιχο έγγραφο. Το έγγραφο μορφοποίησης λειτουργεί ως μηχανισμός δημιουργίας της HTML

σελίδας. Αποτέλεσμα αυτής της διαδικασίας είναι μία HTML σελίδα στην οποία παρουσιάζεται μία λίστα που έχει παραχθεί με βάση το στοιχείο – κλειδί από κάθε XML έγγραφο. Για παράδειγμα, μία λίστα με βάση τα ονόματα συγγραφέων, τα σημεία πρόσβασης, τους τίτλους κτλ. Με αυτό τον τρόπο, τα XML δεδομένα μπορούν να γίνουν εύκολα προσβάσιμα και αναζητήσιμα από φυλλομετρητές και μηχανές αναζήτησης στο Διαδίκτυο.

3. Πρότυπα και eXaMpLe

Η eXtensible Markup Language έχει χαρακτηριστεί ως “*meta-markup language*”. Λειτουργεί με άλλα λόγια ως βάση για τη δημιουργία νέων γλωσσών σήμανσης που αφορούν συγκεκριμένα θεματικά πεδία.

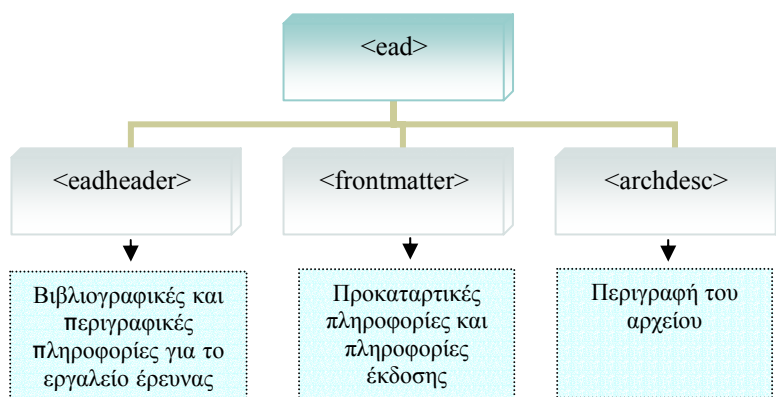
Η εφαρμογή eXaMpLe έχει τη δυνατότητα να χειριστεί ικανό αριθμό προτύπων που έχουν βασιστεί σε XML Σχήμα. Ενδεικτικά θα αναφερθούν δύο από τα πλέον βασικά πρότυπα που χρησιμοποιούνται από τα αρχεία και τις βιβλιοθήκες, το Encoded Archival Description (Library of Congress, 2004c) και το MARCXML Schema (Library of Congress, 2004a). Τα συγκεκριμένα πρότυπα κωδικοποίησης δεδομένων διατίθενται από τη Βιβλιοθήκη του Κογκρέσου.

3.1. Encoded Archival Description (EAD)

Το Encoded Archival Description είναι ένα διεθνές πρότυπο αρχειακής περιγραφής. Η υλοποίησή του επιτρέπει τη δημιουργία ηλεκτρονικών εργαλείων έρευνας, τα οποία αποτελούν το τελικό προϊόν της περιγραφής των αρχείων. Αναλυτικότερα, τα εργαλεία έρευνας είναι τα πληροφοριακά εργαλεία, τα οποία περιλαμβάνουν μεταδεδομένα για την ταύτιση, διαχείριση και ερμηνεία ενός αρχείου και, συγχρόνως, αναλύουν το περιβάλλον δημιουργίας του.

Απεικονίζοντας τη δομή του Διεθνούς Προτύπου Αρχειακής Περιγραφής (Γενικό), το EAD δίνει έμφαση στην ιεραρχική φύση της αρχειακής περιγραφής και στην κληρονομικότητά της. Η αξιοποίηση της ευελιξίας και της δένδροειδούς σύνταξης της XML, επιτρέπει στο EAD την αποτύπωση σε ηλεκτρονική μορφή της πολύ-επίπεδης δομής των αρχείων.

Το EAD διαθέτει ποικιλία στοιχείων και γνωρισμάτων για την πλήρη περιγραφή ενός αρχείου ή μιας συλλογής. Η βασική δομή του προτύπου απεικονίζεται στην Εικόνα 5. Αναλυτικότερα, τα στοιχεία που χρησιμοποιούνται για την περιγραφή του συνόλου του αρχείου ή της συλλογής είναι διαθέσιμα και για την περιγραφή των συστατικών μερών τους. Παράλληλα, κάθε κατώτερο επίπεδο περιγραφής «κληρονομεί» την περιγραφή του επιπέδου στο οποίο περιέχεται ή του ανώτατου επιπέδου (Pitti, 1999). Για παράδειγμα, στην περίπτωση που ο δημιουργός είναι κοινός για όλο το αρχείο και το όνομά του έχει αναφερθεί στην περιγραφή του συνόλου δε θα επαναληφθεί στην περιγραφή των συστατικών μερών. Επιπλέον, το EAD περιλαμβάνει πληροφορίες για το ηλεκτρονικό εργαλείο έρευνας. Σε αυτές τις πληροφορίες καταχωρούνται τα μεταδεδομένα του EAD εγγράφου, όπως το όνομα του δημιουργού και η ημερομηνία δημιουργίας του.



Εικόνα 5. Βασική δομή ενός EAD εγγράφου

Το σύνολο των κανόνων που διέπουν το πρότυπο EAD εκφράζεται στο EAD Document Type Definition. Για τη μετατροπή του EAD DTD σε μορφή XML Σχήματος, είναι δυνατό να χρησιμοποιηθούν διαθέσιμα στο Διαδίκτυο εργαλεία που επιτρέπουν τη μετατροπή DTD εγγράφων σε XML Σχήματα. Συγχρόνως διατίθενται ανεπίσημες εκδόσεις του EAD Schema, οι οποίες είναι συνήθως αξιόπιστες.

Το EAD χρησιμοποιείται από διαρκώς αυξανόμενο αριθμό αρχειακών φορέων, κατά συνέπεια προκύπτει ανάγκη για τη δημιουργία εργαλείων διαχείρισης και οδηγιών υλοποίησής του. Το eXaMPLe έχει υιοθετήσει σε μεγάλο βαθμό τα περιεχόμενα του EAD COOKBOOK 2002 (EAD Round Table of the Society of American Archivist, 2004), το οποίο υποστηρίζεται και ανανεώνεται από τη Βιβλιοθήκη του Κογκρέσου, και αποτελεί ένα σύνολο οδηγιών υλοποίησης, εγγράφων – οδηγιών και εγγράφων μορφοποίησης.

3.2. MARCXML

Η τάση για μεταφορά των βιβλιοθηκονομικών συστημάτων σε περιβάλλον XML ώθησε τη Βιβλιοθήκη του Κογκρέσου στην ανάπτυξη ενός XML Σχήματος για βιβλιογραφικά δεδομένα με το όνομα MARCXML.

Οι πρώτες προσπάθειες της Βιβλιοθήκης του Κογκρέσου για τη μεταφορά των βιβλιογραφικών δεδομένων σε SGML και XML μορφή ξεκίνησαν με τη δημιουργία, αρχικά, ενός SGML DTD και στη συνέχεια ενός XML DTD. Η εξέλιξη των DTD είναι άμεσα συνδεδεμένη με τους ορισμούς και τις δομές των Bibliographic/Holdings/Community Information και Authority/Classification MARC 21 σχημάτων.

Το MARCXML Schema ή MARC “Slim”, όπως αρχικά ονομαζόταν, αποτελεί ένα γενικότερο πλαίσιο εργασίας για τη δημιουργία, παρουσίαση και μετατροπή των MARC (2709) εγγραφών σε XML μορφή και την ανταλλαγή τους σε XML περιβάλλον. Η μεταφορά των εγγραφών σε XML μορφή επιτρέπει τη δημιουργία ισχυρών και ευέλικτων συνδέσεων μεταξύ τους, διευκολύνει τη διαδικασία ελέγχου της εγκυρότητάς τους (validation), ενώ μπορεί να λειτουργήσει ως εύχρηστος διάυλος για τη μετατροπή τους σε άλλα πρότυπα κωδικοποίησης δεδομένων.

Το συγκεκριμένο πρότυπο περιλαμβάνει έξι βασικά στοιχεία: συλλογή (collection), εγγραφή (record), ετικέτα εγγραφής (leader), πεδίο ελέγχου (controlfield), πεδίο (datafield) και υποπεδίο (subfield). Το στοιχείο «συλλογή» δε διαθέτει αντιστοιχία στο MARC, διότι υλοποιήθηκε με σκοπό να περιγράψει μία ομάδα εγγραφών (records), διευκολύνοντας με αυτό τον τρόπο την ανταλλαγή πολλών εγγραφών μέσα από ένα αρχείο καθώς και την

συλλογική μετατροπή τους σε άλλα σχήματα μεταδεδομένων (Keith, 2004). Το MARCXML Schema, πέρα από την περιγραφή ενός συνόλου εγγραφών κάτω από το αρχικό στοιχείο <collection>, επιτρέπει και την ατομική περιγραφή εγγραφών ορίζοντας ως αρχικό στοιχείο το <record>.

Το MARCXML Schema διατηρεί τα βασικά δεδομένα που περιλαμβάνονται σε μία MARC εγγραφή, ενώ αποκλείει δομικά δεδομένα, όπως το μήκος του πεδίου, τα οποία δε θεωρούνται απαραίτητα σε μία XML εγγραφή (Library of Congress, 2002).

Ενδεικτικά δίνεται ένα παράδειγμα κωδικοποίησης βιβλιογραφικών δεδομένων σε MARCXML Schema και η αντιστοιχία του σε MARC 21.

Παράδειγμα 1

MARCXML Schema

```
<record>
  <datafield tag="100" ind1="1" ind2=" " >
    <subfield code="a">Καζαντζάκης, Νίκος,</subfield>
    <subfield code="d">1883-1957.</subfield>
  </datafield>
  <datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">Οι αδελφοφάδες </subfield>
    <subfield code="c">Νίκος Καζαντζάκης.</subfield>
  </datafield>
</record>
```

MARC 21

```
100 1 |aΚαζαντζάκης, Νίκος,|d1883-1957.
245 10|aΟι αδελφοφάδες /|c Νίκος Καζαντζάκης.
```

4. Μελλοντικές υλοποιήσεις

Η XML είναι μία γλώσσα σήμανσης που συνεχώς εξελίσσεται και ενισχύεται από ένα σύνολο υποστηρικτικών τεχνολογιών και προτύπων. Στο νέο στάδιο ανάπτυξης της εφαρμογής eXaMρLe, η ομάδα ανάπτυξης βιβλιογραφικών εφαρμογών του ΕΚΤ σκοπεύει στην υλοποίηση των XML Namespaces (World Wide Web Consortium, 1999b). Μέσα από τη συγκεκριμένη υλοποίηση θα αυξηθούν οι δυνατότητες της εφαρμογής αναφορικά με την ποικιλία των XML Σχημάτων που θα μπορεί να διαχειριστεί.

Αναλυτικότερα, τα XML Namespaces είναι μία συλλογή ονομάτων, προσδιορισμένα από μία URI αναφορά, τα οποία χρησιμοποιούνται στα XML έγγραφα ως τύποι στοιχείων και ονόματα γνωρισμάτων. Κάθε στοιχείο και κάθε γνώρισμα σε ένα XML Namespace προσδιορίζεται μοναδικά από δύο μέρη: το όνομα του XML Namespace και το τοπικό όνομα. Πολλές φορές, μέσα σε ένα XML έγγραφο ή σε μία ομάδα XML εγγράφων, το ίδιο στοιχείο ή γνώρισμα μπορεί να χρησιμοποιείται για να δηλώσει δεδομένα με διαφορετική σημασία. Για παράδειγμα, το στοιχείο <creator> μπορεί σε μία περίπτωση να δηλώνει το δημιουργό του XML εγγράφου, ενώ σε μία άλλη περίπτωση να δηλώσει το δημιουργό ενός τεκμηρίου που περιγράφεται μέσα από το XML έγγραφο. Επιπλέον, ένα XML έγγραφο μπορεί να συνδυάζει σήμανση από διαφορετικά XML Σχήματα.

Τα XML Namespaces αντιμετωπίζουν επιτυχημένα το παραπάνω πρόβλημα. Η λειτουργία τους έγκειται στο διαχωρισμό των στοιχείων και των γνωρισμάτων που προέρχονται από διαφορετικά λεξιλόγια και έχουν διαφορετική σημασιολογία όμως μοιράζονται το ίδιο όνομα.

Συγχρόνως, επιτρέπουν την ομαδοποίηση όλων των σχετικών στοιχείων και γνωρισμάτων με σκοπό την αναγνώρισή τους από το λογισμικό.

Ορίζοντας το συγκεκριμένο λεξιλόγιο από το οποίο προέρχεται το συγκεκριμένο στοιχείο ή γνώρισμα είναι δυνατό να συνδυαστούν πληροφορίες από διαφορετικές πηγές και διαφορετικά πρότυπα κωδικοποίησης δεδομένων χωρίς να δημιουργείται ονομαστική και σημασιολογική σύγχυση.

5. Επίλογος

Η eXtensible Markup Language αποτελεί μία νέα ευέλικτη βάση για την έκφραση και δόμηση των προτύπων κωδικοποίησης δεδομένων. Συνέπεια της εξέλιξης αυτής είναι η καταγραφή σε μορφή XML των περισσότερων προτύπων. Η XML επιτρέπει στους φορείς πληροφόρησης τη χρήση κοινών εργαλείων για την επεξεργασία και διαχείριση των δεδομένων. Η εφαρμογή eXaMpLe αναπτύσσεται με σκοπό να αξιοποιηθεί από τους φορείς πληροφόρησης που υλοποιούν πρότυπα βασισμένα σε XML ως ένα εύχρηστο εργαλείο.

Βιβλιογραφία

Apache Software Foundation (2003) “Xerces-C++ Parser” [Διαθέσιμο στο <http://xml.apache.org/xerces-c/>]

EAD Round Table of the Society of American Archivists (2004) “THE EAD COOKBOOK 2002” [Διαθέσιμο στο <http://www.iath.virginia.edu/ead/ead2002cookbookhelp.html>]

Harold, Elliotte Rusty and W. Scott Means (2002) *XML: in a nutshell*. O’ Reilly

Jelliffe, Rick (2003) “Schematron” [Διαθέσιμο στο <http://www.schematron.com/>]

Keith, Corey (2004) “Using XSLT to manipulate MARC metadata” *Library Hi Tech*, 22(2) pp. 122-130

Library of Congress (2002) “MARC XML Design Considerations” [Διαθέσιμο στο <http://www.loc.gov/standards/marcxml/marcxml-design.html>]

Library of Congress (2004a) “MARC XML” [Διαθέσιμο στο <http://www.loc.gov/standards/marcxml/>]

Library of Congress (2004b) “MODS” [Διαθέσιμο στο <http://www.loc.gov/standards/mods/>]

Library of Congress (2004c) “Encoded Archival Description (EAD)” [Διαθέσιμο στο <http://www.loc.gov/ead/>]

Microsoft (2004) “Processing XML with MSXML” [Διαθέσιμο στο <http://msdn.microsoft.com/XML/BuildingXML/ProcessXMLwMSXML/>]

Pitti, Daniel (1999) “Encoded Archival Description: an Introduction and Overview” *D-Lib Magazine*, 5(11) [Διαθέσιμο στο <http://www.dlib.org/dlib/november99/11pitti.html>]

RELAX NG Technical Committee (2003) “RELAX NG” [Διαθέσιμο στο <http://www.relaxng.org/>]

TEI Consortium (2003) “Text Encoding Initiative” [Διαθέσιμο στο <http://www.tei-c.org>]

World Wide Web Consortium (1999a) “XSL Transformations (XSLT)” [Διαθέσιμο στο <http://www.w3.org/TR/xslt>]

World Wide Web Consortium (1999b) “Namespaces in XML” [Διαθέσιμο στο <http://www.w3.org/TR/REC-xml-names/>]

World Wide Web Consortium (2004a) “Extensible Markup Language (XML)” [Διαθέσιμο στο <http://www.w3.org/XML/>]

World Wide Web Consortium (2004b) “Overview of SGML Resources” [Διαθέσιμο στο <http://www.w3.org/MarkUp/SGML/>]

World Wide Web Consortium (2004c) “HyperText Markup Language (HTML)” [Διαθέσιμο στο <http://www.w3.org/MarkUp/>]

World Wide Web Consortium (2004d) “XML Schema” [Διαθέσιμο στο <http://www.w3.org/XML/Schema>]

World Wide Web Consortium (2004e) “Document Object Model (DOM)” [Διαθέσιμο στο <http://www.w3.org/DOM/>]