ΜΟΝΑΔΕΣ ΑΡΙΣΤΕΙΑΣ
ΑΝΟΙΧΤΟΥ ΛΟΓΙΣΜΙΚΟΥ

ΤΕΙ ΑΘΗΝΑΣ

BibConvert μετατροπές LOM

Δημοσθένης Νικούδης

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Μονάδα Αριστείας ΕΛ/ΛΑΚ ΤΕΙ Αθήνας

cc BY NC SA

BibConvert

- Μετατρέπει μεταδεδομένα από άλλες μορφές σε MARC21 (ή πιο σωστά MARCXML)
- Command-line tool – Δεν έχει web interface
- Υποστηρίζει 2 modes για τα εισαγόμενα δεδομένα
  - Plain text-oriented mode
  - XML oriented mode

Plain text-oriented mode

- Το αρχείο με τα εισαγόμενα δεδομένα χωρίζεται με line-breaks σε γραμμές και με κάποιον ειδικό χαρακτήρα (πχ. ,) σε στήλες
- Γνωστότερο format αυτού του τύπου το CSV
  - Comma/Character separated values
- Μικρό μέγεθος, ελάχιστο περιττό περιεχόμενο
- Δεν μπορούν εύκολα να μπουν nested data

XML

- XML σημαίνει Extensible Markup Language
- Διαδεδομένο πρότυπο
- Γίνεται parse εύκολα από πολλές βιβλιοθήκες
- Ως markup language έχει ομοιότητες με την HTML αλλά είναι αυστηρότερη
  - Δε γίνονται parse όλες οι HTML σαν XML
  - Υπάρχουν όμως μορφές (XHTML) που γίνονται

Παράδειγμα XML


```
<?xml version="1.0" encoding="UTF-8"?>
<lom>
    <general>
        <title>
            <string language="el">Η θάλασσα στο έργο του Στράτη Μυριβήλη</string>
            <string language="en">The sea in his work</string>
        </title>
    </general>
</lom>
```

Τι θα κάνουμε

- Θέλουμε να βρούμε έναν τρόπο να μετατρέπουμε υπάρχοντα δεδομένα με μορφή LOM σε MARCXML ώστε να μπαίνουν στο Invenio
  - Τα δεδομένα LOM που έχουμε είναι σε μορφή XML
  - Από τη στιγμή που θα γίνουν convert με ποιο module θα εισαχθούν στο Invenio;

XSL

- XSL σημαίνει E**X**tensible **S**tylesheet **L**anguage
- Τα tags στην XML δεν είναι predefined – μπορούμε να χρησιμοποιήσουμε ότι tag name θέλουμε
  - Αντίθετα, στην HTML πχ. το <table> tag έχει συγκεκριμένη σημασία
- Αυτό σημαίνει ότι η σημασία του κάθε tag δεν είναι καλά κατανοητή
- Το XSL αναπτύχθηκε για να μπορούμε να δίνουμε σημασίες σε XML tags είτε βρίσκοντας πληροφορίες σε ένα document είτε μετατρέποντας το σε άλλες μορφές
- Το XSL περιλαμβάνει τρείς τεχνολογίες: XSLT, Xpath, XSL-FO

XPath

- Είναι ένα query language που μας βοηθάει να βρίσκουμε πληροφορίες μέσα σε ένα έγγραφο XML. Μια από τις πιο απλές μορφές του είναι:

```
<A>
   <B>
      <C>Test</C>
   </B>
</A>
```

- Για να βρούμε το περιεχόμενο του C μπορούμε να χρησιμοποιήσουμε αυτό το expression:
  - /A/B/C

XSLT

- Είναι μια γλώσσα για μετατροπή XML εγγράφων σε άλλες μορφές
  - Στην περίπτωση μας θα κάνουμε μετατροπή από ένα είδος XML σε ένα άλλο
- Το XSLT ορίζει τους κανόνες με τους οποίους θα γίνει η μετατροπή
- Η γλώσσα XSLT είναι υποσύνολο της XML
  - Οι κανόνες περιγράφονται με συγκεκριμένα XML tags

Παράδειγμα XLST

- Έστω ότι έχουμε το παρακάτω XML και θέλουμε να το κάνουμε ένα όμορφο HTML:

```
<?xml version="1.0" encoding="UTF-8"?>
<catalog>
  <cd>
    <title>Empire Burlesque</title>
    <artist>Bob Dylan</artist>
    <country>USA</country>
    <company>Columbia</company>
    <price>10.90</price>
    <year>1985</year>
  </cd>
</catalog>
```

Παράδειγμα XSLT

Το περνάμε από το εξής XLST

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

<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

<xsl:template match="/">
  <html>
  <body>
  <h2>My CD Collection</h2>
  <table border="1">
    <tr bgcolor="#9acd32">
      <th>Title</th>
      <th>Artist</th>
    </tr>
    <xsl:for-each select="catalog/cd">
    <tr>
      <td><xsl:value-of select="title"/></td>
      <td><xsl:value-of select="artist"/></td>
    </tr>
    </xsl:for-each>
  </table>
  </body>
  </html>
</xsl:template>

</xsl:stylesheet>
```

Ποιο θα είναι το αποτέλεσμα;

Ποιο θα είναι το αποτέλεσμα;


```
<html>
  <body>
  <h2>My CD Collection</h2>
  <table border="1">
    <tr bgcolor="#9acd32">
      <th>Title</th>
      <th>Artist</th>
    </tr>
    <tr>
      <td>Empire Burlesque</td>
      <td>Bob Dylan</td>
    </tr>
  </table>
  </body>
</html>
```

Άσκηση

- Θα φτιάξουμε ένα XSL με όνομα lom2marcxml.xsl που να μετατρέπει τα πεδία title και description από το lom_example.xml σε MARCXML
- Θα δοκιμάσουμε το XSL που δημιουργήσαμε με την εντολή:
  - bibconvert -clom2marcxml.xsl < lom_example.xml > marcxml_example.xml

Template


```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
   <xsl:output method="xml" indent="yes" encoding="UTF-8"/>
   <xsl:template match="/">
  Περιεχόμενο εδώ
   </xsl:template>
</xsl:stylesheet>
```

Αναμενόμενο αποτέλεσμα


```
<?xml version="1.0" encoding="UTF-8"?>
<record>
  <datafield tag="245" ind1=" " ind2=" ">
    <subfield code="a">Η θάλασσα στο έργο
  του Στράτη Μυριβήλη</subfield>
  </datafield>
  <datafield tag="500" ind1=" " ind2=" ">
    <subfield code="a">Αυτό το κλιπ
  εστιάζει στην αγάπη του Μυριβήλη για τη
  θάλασσα και την κυριαρχική παρουσία της
  στο έργο του.</subfield>
  </datafield>
</record>
```

Επιβεβαίωση και ανέβασμα

- Μπορούμε να ελέγξουμε την ορθότητα του παραγόμενου xml ως MARCXML με το εργαλείο xmlmarclint
  - xmlmarclint --verbose=9 marcxml_example.xml
- Αν είναι έγκυρο μπορούμε να το ανεβάσουμε στο Invenio με το BibUpload
  - bibupload -i marcxml_example.xml
- Το BibUpload βάζει το αρχείο σε μια ουρά η οποία καταναλώνεται από ένα άλλο module – το BibSched

Επιβεβαίωση και ανέβασμα

- Σε περίπτωση που δεν τρέχει ο worker του bibsched τον ξεκινάμε με την εντολή:
  - bibsched start
- Για να βρούμε το id του εγγράφου που ανεβάσαμε μπορούμε να δούμε τα log του BibSched στον φάκελο:
  - /opt/invenio/var/log/bibsched/0
- Γιατί δεν εμφανίζεται το νέο έγγραφο στην αναζήτηση;
  - Δεν έχει μπει ακόμα στο index

Χρήσιμα links

- Xpath tutorial: http://www.w3schools.com/XPath/
- Xpath tester: http://www.xpathtester.com/xpath
- MARC21 tag reference: http://www.loc.gov/marc/bibliographic/ecbdhome.html
- Έτοιμο xsl από Invenio: https://github.com/inveniosoftware/invenio/blob/master/modules/bibconvert/etc/crossref2marcxml.xsl

Σας ευχαριστώ πολύ

Ερωτήσεις;