METALIS: A federated-search tool to aggregate Data Providers in LIS field

METAreresearch in Library and Information Science
Instruments used

- Cheshire II
- Perl 5.8 scripts (XML::Twig and HTML::Entities)
- TCL scripts
- HTML pages
Operations

- Harvest
- Conversion to an internal schema
- Search and display
Harvest

• OAI-PMH sources

• Every day an OAI-PMH query

• Use OAI-PMH to harvest a section of a Data Provider

• All sources available in LIS field
Conversion

• One format (oai_dc format)
• Many implementation
• Many interpretations
• Key point:
  ▪ Subject
  ▪ Language
  ▪ Type
  ▪ Rights
Conversion

• Internal format has more than metadata only

• There are:
  – Metadata
  – Display information
  – Sort information's
Conversion - ArchiveSIC

• Type conversion
  'Text' => 'Article'

• Language conversion
  'fr' => 'fre',
  'en' => 'eng',
  'de' => 'ger',
  'es' => 'spa',
  'it' => 'ita'
Conversion - ArchiveSIC


• ‘Mass media’, ‘Scientific communication and information’ => ‘C’
  ‘Museology’ => ‘D’
• ‘Information/communication law’ => ‘E’
• ‘Economy, cultural industry’, ‘Education e-learning training’ => ‘G’
• ‘Electronic publishing’ => ‘H’
• ‘Hypertext, hypermedia’, ‘Information retrieval’ => ‘I’
• ‘Information system engineering’, => ‘L’
Display modification

sub mod_title # modification of the title for the display
    my($t, $title) = @_;    # Import values in the routine
    $title->set_gi('dc:title');    # change the tag dc:title into dctitle
    my $title_value = $title->text;    # Work the value
    if (!($title_value)) {$title_value = 'Not title available'} ;
    my $decoder = Encode::Guess->guess($title_value);    # Find char set
    die $decoder unless ref($decoder);
    my $utf8 = $decoder->decode($title_value);    # All into utf-8
    $utf8 =~ s/&/&/g;
    $utf8 =~ s/&quot;/"/g;
    $utf8 = HTML::Entities::decode($utf8);    # all chars into &#nnnn;
    my $title_value = HTML::Entities::encodenumeric($utf8);
    $title->set_text($title_value);    # Put out the changed value
Sorting

• Put all data in one tag.
• For the authors: <sortauth></sortauth>
• Only for limit of cheshire 2
• To improve in the future
Search

• Use every information to improve the search
• Use a familiar interface.
• An interface like an OPAC
• On the top links to navigate
• Down informative links
METALIS is a Service Provider for the Library and Information Science field. We collect (harvest) metadata from institutions that offer full-text papers and documents about Library and Information Science. The general framework is known as "Open Access Movement"; the AEPIC project maintains a list of useful resources about all aspects of Open Access.

The different options available have these meanings:

- **Simple Search**: To search by keywords in all fields. It is possible to filter by year, type of document, origin of metadata and class.
- **Advanced Search**: To search by keywords and phrases in different fields. It is also possible to filter by one or more years, type of document, origin of metadata and class.
- **Setup OpenUrl Resolver**: METALIS builds an OpenURL with the metadata of the document. By default METALIS uses its resolver; if you prefer you can specify your local resolver.

METALIS harvests from these Data Providers:

- Archive Ouverte en Sciences de l'Information et de la Communication.
- ArXiv: Only set about Computer Science and for the subjects "Digital Libraries" and "Information Retrieval".
- Caltech Library System Papers and Publications.
- CNR Bologna Research Library.
- Digital Library of Information Science and Technology.
- E-Prints in Library and Information Science.
- Librarians' Digital Library: Only two sets, Publications / Articles and Theses / Dissertations.
- Mémoires de 3 cycle en Sciences de l'Information et de la Communication.
- Thèse-EN-ligne: Only sets about Library and Information Science.

This work is licensed under a Creative Commons License.

Service developed by Zeno Tajoli. Credits of S. Warner, A. Tugnoli, UKOLN and RDN. Sources
Advanced search link

http://metalis.cilea.it/advanced.html
OpenUrl

• A standard to find a specific full-text
• A reverse use
• To find correlated full-texts
• METALIS can use your OpenUrl resolver
The Future

• Now Cheshire II

• In the future:
  – Cheshire3 http://www.cheshire3.org/ ?
  – Zebra with OCKHAM
    http://wiki.osuosl.org/display/OCKPub/Home ?
  – CDSware ?
  – Others ?
Question?