Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Lafayette Ragsdale - Limber Up! A Limited Edition Civil War Print
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Outline

Introductions - Myself, Project, Thesauri, RDF

The Challenge - To Facilitate Cross-European Comparative Data Analysis

The Dream - A Computer Data-ing Service

The Reality - The Limber Approach

The Progress - Metadata, Thesaurus, Workshop

The End - A Possible Sponsor and Questions ???
Introductions

Ken Miller - Head of Information Systems Development, The UK Data Archive, The University of Essex, England

EU Human Language Technologies IST Project - Language Independent Metadata Browsing of European Resources.

CLRC Rutherford Appleton Laboratory, Intrasoft, UK Data Archive, Norwegian Data Archive, user group:- other European national archives

To Breakdown Linguistic and Discipline Barriers to interoperability between European Resources

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Searching

Go back to the

Softness. 324
Adj. soft, lissom, lim extensile, remollent

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University
Hierarchical Thesauri

- List of specific synonyms
- Broader and narrower relationships
- Subject concept
- Concept tree - hierarchies
- Shift focus within concept
- Ensured relevance
- Browsing and KWIC listings
- Apply to different parts of metadata
- Ranking
**Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing**

<table>
<thead>
<tr>
<th>NT</th>
<th>UF</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMPLOYMENT HISTORY,</td>
<td>JOBS, WORK</td>
<td>CONDITIONS OF EMPLOYMENT,</td>
</tr>
<tr>
<td>EMPLOYMENT PROGRAMMES,</td>
<td></td>
<td>EMPLOYMENT ABROAD,</td>
</tr>
<tr>
<td>FULL-TIME EMPLOYMENT,</td>
<td></td>
<td>EMPLOYMENT OPPORTUNITIES,</td>
</tr>
<tr>
<td>HOME-BASED WORK,</td>
<td></td>
<td>EMPLOYMENT POLICY,</td>
</tr>
<tr>
<td>IRREGULAR ECONOMIC ACTIVITY,</td>
<td></td>
<td>EMPLOYMENT SERVICES,</td>
</tr>
<tr>
<td>JOB CHANGING,</td>
<td></td>
<td>EMPLOYMENT STATUS,</td>
</tr>
<tr>
<td>JOB LOSSES,</td>
<td></td>
<td>LABOUR (WORK),</td>
</tr>
<tr>
<td>JOB SHARING,</td>
<td></td>
<td>LABOUR ECONOMICS,</td>
</tr>
<tr>
<td>JOB VACANCIES,</td>
<td></td>
<td>OCCUPATIONS,</td>
</tr>
<tr>
<td>MILITARY SERVICE,</td>
<td></td>
<td>PERSONNEL,</td>
</tr>
<tr>
<td>OFF-SHORE EMPLOYMENT,</td>
<td></td>
<td>PERSONNEL MANAGEMENT,</td>
</tr>
<tr>
<td>PART-TIME EMPLOYMENT,</td>
<td></td>
<td>RIGHT TO WORK,</td>
</tr>
<tr>
<td>SEASONAL EMPLOYMENT,</td>
<td></td>
<td>WORKING CONDITIONS</td>
</tr>
<tr>
<td>SHELTERED EMPLOYMENT,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPOUSES' EMPLOYMENT,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STUDENT EMPLOYMENT,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUBSIDIARY EMPLOYMENT,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEMPORARY EMPLOYMENT,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMPLOYMENT,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMEN'S EMPLOYMENT,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YOUTH EMPLOYMENT,</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:**

*Charting the Future for Social, Spatial and Government Data*

*June 7-10, Northwestern University*
RDF - Resource Description Framework

World-Wide Web Consortium (W3C)
A framework within which statements about Web resources can be recorded
Uses XML as its physical syntax
Based around directed labelled graphs
Triples - Resource, Property, Value
Can construct more complex models
Statements about statements
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data
June 7-10, Northwestern University
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

<web:Description about="http://elsst/concepts/ClD_6">
<web:type resource="http://rdf-dot/Thes/Thes.xrdf#Concept"/>
<rdfs:isDefinedBy web:resource="http://elsst/concepts/">
<thes:indicator web:resource="http://elsst/terms/TID_3"/>
<thes:notation>620</thes:notation>
<thes:scope web:resource="http://elsst/scopenotes/SN_12"/>
<thes:relatedConcept web:resource="http://elsst/concepts/ClD_15"/></web:Description>

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data
June 7-10, Northwestern University
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

<web:Description about="http://elsst/terms/TID_3">
<web:type resource="http://rdf-dot/Thes/Thes.xrdf#Term"/>
<thes:lang>en</thes:lang>
<web:value>Friends</web:value>
<thes:termType web:resource="http://rdf-dot/Thes/Thes.xrdf#preferred"/>
</web:Description>

<web:Description about="http://elsst/scopenotes/SN_12">
<web:type resource="http://rdf-dot/Thes/Thes.xrdf#ScopeNote"/>
<thes:lang>en</thes:lang>
<web:value>To be used only for platonic relationships</web:value>
</web:Description>

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data
June 7-10, Northwestern University
The Challenge - To Facilitate Cross-European Comparative Data Analysis

A metadata model and representation to allow integration within and across data archives.

A multilingual thesaurus to index and access social science datasets in data archives.

A multilingual query and retrieval tool to allow queries to datasets in archives to be made in several languages with keyword and phrase translation in the retrieved metadata.

Tools to support the construction and maintenance of datasets in an archive using automatic indexing.

An XML metadata server using the metadata model and representation.
“Effect of Diet on Children’s Teeth”

Greenhouse effect, Diet, Diet and nutrition, Diet therapy, Children’s rights, False teeth, teeth

TEETH
nt DENTURES
rt DENTAL DISEASES
rt DENTAL HEALTH

“National Diet, Nutrition and Dental Survey”

Show me more
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Computer dating service for datasets
Metadata elements to search for a perfect match
Metadata to display to confirm a perfect match
Analysing tools to ensure a perfect match
Resource Independent
Language Independent
Thesaurus Independent
Dream on
The Dream -
A Computer
Data-ing Service

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data
June 7-10, Northwestern University
<table>
<thead>
<tr>
<th>Documentation</th>
<th>Table</th>
<th>Graphics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

42. DDL_NOW: Domestic division of labour: now
7: MUMOCC: Mother's occupation

<table>
<thead>
<tr>
<th></th>
<th>semi/uns</th>
<th>Skilled</th>
<th>low prog</th>
<th>higher prog</th>
<th>entrep</th>
<th>housewit</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>traditional</td>
<td>41.7</td>
<td>60.6</td>
<td>53.8</td>
<td>73.9</td>
<td>60.0</td>
<td>50.0</td>
<td>55.9</td>
</tr>
<tr>
<td>shared</td>
<td>33.3</td>
<td>24.2</td>
<td>26.6</td>
<td>13.0</td>
<td>0.0</td>
<td>32.1</td>
<td>26.0</td>
</tr>
<tr>
<td>men do most</td>
<td>8.3</td>
<td>0.0</td>
<td>5.1</td>
<td>0.0</td>
<td>0.0</td>
<td>3.6</td>
<td>3.3</td>
</tr>
<tr>
<td>third party</td>
<td>4.2</td>
<td>0.0</td>
<td>7.7</td>
<td>4.3</td>
<td>20.0</td>
<td>7.1</td>
<td>5.3</td>
</tr>
<tr>
<td>lives alone</td>
<td>8.3</td>
<td>15.2</td>
<td>7.7</td>
<td>8.7</td>
<td>20.0</td>
<td>7.1</td>
<td>9.9</td>
</tr>
<tr>
<td>6</td>
<td>4.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>N=</td>
<td>24</td>
<td>33</td>
<td>39</td>
<td>23</td>
<td>5</td>
<td>28</td>
<td>152</td>
</tr>
</tbody>
</table>

Study title: Women's Employment and Family Life in the United Kingdom, Norway, France, Czech Republic and Russia
Copyright: Copyright R. Crompton and F. Harris.
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data
June 7-10, Northwestern University
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data
June 7-10, Northwestern University
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

The Reality - The Limber Approach

Discipline boundaries

RDF

Mappings between metadata standards

Linguistic boundaries

Multi-lingual Thesaurus Interface

Relevance feedback

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Convert DDI codeBook to RDF
Translate DDI Headings/ Tag Library
Reduce HASSET
Add Methodology Terms
Translate Hierarchies
Check individual languages
Convert to XML/ RDF
Multi-mapped 4 Language Thesaurus
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

The Progress - Metadata, Thesaurus, Workshop

Reduction of Thesaurus
Automation NOT possible
Removal of Cultural specificity
Removal of Institutional specificity
Removal of hierarchies in existing thesauri
Initial reduction of major 20 hierarchies
A top down reduction to very broad base
User group evaluation of reductions

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data
June 7-10, Northwestern University
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Metadata Additions to Thesaurus methodology kind of data universe spatial unit file structure, format, type access conditions age groups
Translation of Thesaurus

Reduced monolingual thesaurus
Specialist teams with social science background
Backwards and sideways translations
Additions of new terms
Allow for different structures
Allow for non-equivalence
Extensive use of synonyms and scope notes
Other EU projects (PACO, CHINTEX etc)
Thesaurus defined in XML and RDF

<thesTerm classCode="L10.10" type="pref">
  <termName xml:lang="en">ADDICTION</termName>
  <termName xml:lang="fr">DÉPENDANCE</termName>
  <scopeNote xml:lang="en" type="ambig">Use a more specific term if possible</scopeNote>
  <ufTerm fterm="L10.10NP"/>
  <ntTerm fterm="L10.10.10"/>
  <btTerm fterm="L10"/>
  <ttTerm fterm="L"/>
  <rtTerm fterm="R70.40"/>
</thesTerm>
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Limber Workshop - April 17-18th 2000

Alta-vista style searches
Free text / controlled vocabulary option
Logic of multi-word combination explained
Interface and hit list in chosen language
Hit list ranked by where word found
Customised prioritising of ranking
Search history to refine and combine searches
Other EU projects (e.g. PACO, EPAG etc)
Search via classification code
Automatic use of thesaurus only for synonyms
Complexity of thesaurus hidden unless requested
Simple alphabetic listings except for browsing
Searches performed first with suggestion list
Failed searches employ stemming and truncation
Keywords are useful for searching and ranking
Keywords are NOT so useful for relevance
Serious analysis requires translation
Title and summary in English
Controlled vocabularies on specific elements
  access conditions, methodology, kind of data, universe and spatial unit
Languages of metadata displayed and selectable
Translation of DDI element headings
Great interest in Automatic Indexing Tool
Great scepticism as to if it would work
To work from DDI metadata
Use linguistic techniques to draw concepts out
Assign keywords that express those concepts
Learn from metadata with manual keywords
Other EU projects (RENAIRDUS) + OCLC
Some doubts over RDF
Some doubts over DDI / NESSTAR
Some doubts over imbalances
Some doubts over multiple resources
Some doubts over mapping
IASSI ST conference 7-10th June 2000
New doubts ????

IASSI ST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data
June 7-10, Northwestern University
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Mother Earth Herbals

The End - A Possible Sponsor and Questions ???

Limber Up – Ache and Pain Salve

Here’s help for all you weekend warriors! Apply this all natural healing rub to sore joints, muscles, tendons, ligaments or sprains. 2 oz. Jar.

Ingredients: Olive oil, St John's Wort, arnica, beeswax and essential oils.

Regular price: $6.50  Sale price: $5.50

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:
Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University
Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Lafayette Ragsdale - Limber Up! A Limited Edition Civil War Print