Facilitating the creation of on-line courses with existing Internet material

Marco Alfano
*Anghelos* Centre on Communication Studies, Italy
Computer Science – University of Palermo, Italy

Biagio Lenzitti
Computer Science – University of Palermo, Italy

Natalia Visalli
SISSIS – University of Palermo, Italy
Contents

1. background
2. creating new online courses with existing learning objects
3. saxef: A system for automatic extraction of learning object features
4. saxef development
5. text analysis module
6. experimental results
7. conclusions & future work
Background

- Huge amount of learning material available in Internet
  - educational (school, university)
  - professional
  - cultural

- Different types of materials
  - content presentation (text, multimedia, etc.)
  - length and level of details
  - interactivity degree

- Different scenarios
  - student works alone
  - student interacts with a tutor or a virtual class
Creating new online courses with existing learning objects

“Standard” approach (SCORM, IMS, ...)

Is it viable?
Many LOs already existing (with no metadata)
Are LO authors willing to insert metadata?
Do metadata provide didactic description of LO?
LO creation more technological than didactic
Our proposal: Use what exists!

What changes?
LO used “as is”
Basic indicators totally describe LOs (also didactically)
Teacher selects suitable LOs just by looking at their indicators
What can we extract from an existing LO?

• LO basic indicators
  – main and secondary topics
  – theoretical or practical LO
  – synthetic or analytical LO
  – media types and multimediality level
  – complexity level
  – links to other LO with same topics
  – links to other LO with related topics
  – ...

A sort of DNA that totally describes the LO (hypothesis to be verified)
SAXEF: A System for Automatic eXtraction of E-Learning object Features

- Analyzes whole on-line courses or single LOs
- Extracts LO basic indicators
- Produces E-learning Identification Cards (EIC)
- Stores EICs in a database and links them

A user can either search a topic into the DB or analyze new LOs and insert related EICs into the DB.
SAXEF Architecture (3 levels)

Low Level:
- Learning Object
- Image Analysis
- Text Analysis
- Sound Analysis
- Identifies media
- Extracts specific features of each media

Medium Level:
- Pattern Fusion
- Creates EICs

High Level:
- Fuses media features

E.I.C.
A teacher has to organize a new course and wants to use existing on-line courses and LOs

**Without SAXEF**

- Uses a search engine and obtains a list of URLs that deal with the subject of interest
- Manually analyzes the contents and the didactic nature of URLs and decides whether to include them into the course

**Is it viable?**

- Much time consuming!
- Analysis limited to the first few URLs?
- New LOs will likely be created
- No storage of the analysis results
SAXEF usage example (2/2)

With SAXEF

• Uses a search engine and obtains URLs...
• Uses SAXEF to analyze found URLs and obtains LO basic indicators
• Examines LO indicators, chooses and stores LOs of interest
• Directly analyzes chosen LOs and decides whether to include them into the course

What changes?

Time saved!
Suitable LOs more likely to be found
Full control kept on the contents of new course
Storage of the analysis results for re-use by others
SAXEF development

- The first phase of SAXEF development is being devoted to confirm SAXEF viability
- Development progress
  - low level to separate the different media of a learning object (given in a standard html format)
  - first module of the medium level to perform a **text analysis** that provides the most frequent words and their distances
  - The objective is to find **main and secondary topics** of the learning object or the whole course
The module performs the following tasks:

- Input parameters management
- Internal link discovery (in case of a whole course)
- Text analysis
- Insert of output into the database
Input parameters management

- **Five fields**
  - URL of the web page to be parsed
  - Minimal percentage of single words
  - Minimal percentage of word couples
  - Analysis of single page or whole course
  - Output level (summary or details of the analysis)
Text analysis steps

• Elimination of the words of common use
• Computation of single word occurrences
• Computation of word couples occurrences
• Research of words inside relevant tags
• Word selection (based on input %)
• Weight assignment to words

\[
\text{weight} = \frac{\text{occ}_w}{\text{tot}_w} + \frac{M}{\sqrt{\text{tot}_w}} + 3 \times \left( \frac{T}{\sqrt{\text{tot}_w}} \right) + \frac{\text{occ}_d}{\sqrt{\text{tot}_w}}
\]

• Choice of main and secondary topics based on weighted words
The output shows the main and secondary topics with their scores and the links of parsed pages. The user can delete unrelated topics and links, and store the remaining ones in a MySQL database.
Experimental results

We ran fifty tests using heterogeneous topics (arts, sciences, ...) and choosing whole courses with at least three LOs

Experiment methodology

• Search of a topic in the net (with google)
• Selection of some LOs found in the net
• Human analysis of chosen LOs for determining the main topic
• Text analysis of same LO through SAXEF
Experiment Methodology (cont.)

- Score attribution to words found by SAXEF
  - thematic correspondence between the word and the main topic determined through the human analysis
  - 1 point to first and second found words (main topics)
  - 0.5 points to the next four words (secondary topics)

- Computation of the proximity index (from 0 to 4) as sum of the individual scores

<table>
<thead>
<tr>
<th>URL</th>
<th><a href="http://www.synapses.co.uk/genetics/index.html">http://www.synapses.co.uk/genetics/index.html</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Analysis</td>
<td>Genetic</td>
</tr>
</tbody>
</table>
| SAXEF Analysis and score attribution | 1. chromosomes (1)  
2. cell (1)  
3. meiosis (0.5)  
4. genetics (0.5)  
5. mitosis (0.5)  
6. number (0) |
| Proximity index | 3.5 |
Test results

- ~ 90% of proximity indexes > 2 (mean value)
- 70% of proximity indexes ≥ 3
  - at least 4 words have a thematic correspondence with the chosen topic
Conclusions

• SAXEF development is planned in more steps
• First phase has been mainly devoted to dividing a LO in its media components and performing a textual analysis to find main and secondary topics
• First test results are very encouraging and suggest us to progress further in this direction
• The analysis must be more accurate and less linked to human judgments
Future work

• Presently only html documents can be analyzed through SAXEF but soon also doc, pdf and ppt documents will be permitted

• A second module is being developed to analyze the media relationships and to evaluate the synthetic or analytical nature of LOs and their multimediality level

• A graphical map showing the relationships among stored LOs is being created

• SAXEF is available at http://altair.math.unipa.it/saxef

Thanks!