Classification and automatic concept map creation in eLearning environment

Karol Furdík¹, Ján Paralič¹, Pavel Smrž²

¹: Technical University of Košice, Centre for Information Technologies,
Letná 9, 040 01 Košice, Slovakia
{Jan.Paralic, Karol.Furdik}@tuke.sk

²: Brno University of Technology, FIT, Božetěchova 2, 612 66 Brno,
University of Economics, Prague, W.Churchill Sq.4, 130 67 Praha, Czech Republic,
smrz@fit.vutbr.cz

13. - 15. 2. 2008
FIIT STU Bratislava
Contents

Context
• KP-Lab project: Collaborative eLearning
• PoZnaŤ project: Adaptation to the Slovak education

Semantic Annotation of Knowledge Artefacts

Text Mining Services
• Creation of Concept maps
• Classification
• Functional specification, Architecture
• Implementation, API - signatures of methods

Mining Engine Console
Testing, Evaluation
Future work
Collaborative eLearning

Attributes:
- Collaborative work
- Sharing of learning materials (knowledge artefacts)

Trialogical learning, Activity theory:
- knowledge creation activities rely on usage, manipulation, and evolution of shared knowledge artefacts

KP-Lab project, www.kp-lab.org:
- integrated EU funded FP6 IST project (2006 - 2011)
- objective: development of eLearning system aimed at facilitating innovative practices of sharing, creating, and working with knowledge in education and workplaces
Project PoZnaŤ

Full title: Support of the processes of innovative knowledge creation
Duration: 30 months (February 2007 - July 2009)
Web: www.tuke.sk/fei-cit/poznat/

Project is supported by the Slovak Research and Development Agency under the contract No. RPEU-0011-06

Builds on the integrated EU project KP-Lab

Focused on: eLearning, Knowledge management.

Main goal: adaptation of the KP-Lab tools for trialogical learning to the conditions of Slovak higher education.
Knowledge artefacts are semantically annotated (by means of ontologies/taxonomies/concept maps) and are collaboratively investigated in the shared learning space.

Text mining services - intelligent access and manipulation with the knowledge artefacts; to assist users in creating or updating the semantic descriptions of KP-Lab knowledge artefacts.

TMS fundamental tasks:

• **Ontology learning** - extraction of conceptual maps (clustering), i.e. an automatic extraction of significant terms from KA's textual descriptions and converting them to a structure of concepts and their relationships.

• **Classification of knowledge artefacts** - grouping a given set of artefacts into predefined or ad hoc categories.
Classification and automatic concept map creation in eLearning environment
Karol Furdík, Ján Paralič, Pavel Smrž
13. - 15. 2. 2008, FIIT STU Bratislava
Concept map creation

Concept maps can be used in learning for:

- open-ended research questions about a certain topic,
- questions coming out of practical experiences that are to be explored.

Procedure:
- Related materials are uploaded to the Shared Space.
- System analyses the data (using Text Mining services) and provides a hierarchy of concepts extracted from documents.
General Functionality (1)

Concept map services:

- Pre-process documents, produce internal representation, store it into the Mining Object Repository,
- Manage the Mining Object Repository (insert / update / delete),
- Find clusters in a set of documents,
- Identify concept candidates and rank them according to the estimated relevance,
- Given a set of concepts, find related concepts from the documents provided by the user. Return a ranked list of candidate relations together with their types,
- Build the concept map and generate the RDF graph.
Classification services:

• Creation of a training data set from already annotated knowledge artefacts to a pre-defined set of categories,

• Creation of a classification model, based on the selected algorithm and on a given training data set,

• Modification (tuning) of the classification model, by changing the texts and/or categories in the training data set, as well as by editing the settings of the algorithm or switching to another algorithm,

• Provision of basic measures for existing classification model, e.g. by means of precision and recall,

• Verification and validation of the classification model.

• Classification of unknown (not annotated) artefacts to the categories used for training. The output of this function is a set of weighted categories (concepts, terms) for each of the classified artefacts.
Implementation

Java 1.5, Service-oriented architecture - Web Services

J Bowl Library: http://sourceforge.net/projects/jbowl/

• platform for pre-processing (incl. NLP methods) and indexing of large textual collections;
• functions for creation and evaluation of text mining models (for both supervised or unsupervised algorithms).

GATE Framework: http://www.gate.ac.uk

• an architecture, or organisational structure, for NLP software;
• a framework, or class library, which implements the architecture;
• a development environment built on top of the framework
Pre-process service - implemented as a pipeline of processing resources on top of the GATE engine. Additional NLP resources integrate language-dependent tasks such as parsing, keyword extraction, co-occurrence statistics and semantic-distance computation:

- URI preprocess(String[] artefacts, String[] seed)
- void delete(URI preprocessedData)

Clustering and concept map service - interfaces to unsupervised text mining methods:

- String[] findClusters(String[] artefacts)
- String[] findConceptCandidates(URI preprocessedData)
- String[] findRelatedConcepts(URI preprocessedData, String[] concepts)
- String[] buildConceptMap(URI preprocessedData)

API reference: [http://pcnlp1.fit.vutbr.cz:8080/ConceptMapCreation/]
Pre-process Service

Classification and automatic concept map creation in eLearning environment
Karol Furdík, Ján Paralič, Pavel Smrž

13. - 15. 2. 2008, FIIT STU Bratislava
Classification and automatic concept map creation in eLearning environment

Karol Furdík, Ján Paralič, Pavel Smrž

13. - 15. 2. 2008, FIIT STU Bratislava
**API - signatures of methods (2)**

**LearningClassification service** - methods for creation, verification, modification, and removal of a classification model:

- String createModel(String settings, String[] artefactURI)
- String verifyModel(String modelURI, String[] artefactURI)
- String modifyModel(String[] settings, String[] artefactURI)
- String deleteModel(String modelURI)

**Classify service** - method for classification of artefacts:

- String[] classify(String modelURI, String[] artefactURI)

API reference, WSDL, demo web interface:

http://147.232.75.67:8080/TMSClassify/
Classification and automatic concept map creation in eLearning environment

Karol Furdík, Ján Paralič, Pavel Smrž

13. - 15. 2. 2008, FIIT STU Bratislava
Classify Service

Output: Array [URI of classified artefact, category, weight]

Pre-processing

Plain text

Internal Classify Implementation

Retrieve from the Mining Object Repository

properties of artefact

Content item

Gateways to Content Repository

URI of Content item

SWKM Knowledge Repository

moURI

artefactURI[]

artefactURI

Classification and automatic concept map creation in eLearning environment
Karol Furdík, Ján Paralič, Pavel Smrž
13. - 15. 2. 2008, FIIT STU Bratislava
Mining Engine Console

- Web interface (JSP),
- Access to the TMS for users,
- Maintenance of the Mining Object Repository and classification models,
- Perform classification tasks,
- View statistics, generate reports.
Three categorized sets of documents:

- **EID UU**: University of Utrecht, The Netherlands. 138 PDF and MS Word documents classified into 5 categories. Language: Dutch.
- **UHE**: University of Helsinki, Finland. 56 MS Word and plain text documents classified into 5 categories. Language: English.
- **VSE**: University of Economics in Prague, Czech Republic. 97 PDF and MS Word documents classified into 7 categories. Language: English.

<table>
<thead>
<tr>
<th>Test case</th>
<th>Documents All/Train/Test</th>
<th>Cat.</th>
<th>Algorithm</th>
<th>Recall [%]</th>
<th>Precision [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>EID UU</td>
<td>138 / 90 / 48</td>
<td>5</td>
<td>kNN</td>
<td>70,4</td>
<td>27,8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Perceptron</td>
<td>74,3</td>
<td>24,7</td>
</tr>
<tr>
<td>UHE</td>
<td>56 / 35 / 21</td>
<td>5</td>
<td>kNN</td>
<td>67,7</td>
<td>21,2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Perceptron</td>
<td>65,4</td>
<td>18,6</td>
</tr>
<tr>
<td>VSE</td>
<td>97 / 65 / 32</td>
<td>7</td>
<td>kNN</td>
<td>84,0</td>
<td>42,3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Perceptron</td>
<td>82,3</td>
<td>37,4</td>
</tr>
</tbody>
</table>
Future Work

Full integration of the Text Mining Services into the whole KP-Lab system is planned for spring 2008.

Enlarge data collection for testing: build a corpora of learning materials, create the suite of tools for processing of Slovak language.

Implement more text mining algorithms, compare their efficiency: a mechanism of automatic selection of learning algorithms and their settings is currently investigated.

Enhance the Mining Engine Console for usage within the real learning process: the console will be tested in summer semester 2008 on Technical University of Košice, within the lessons of Knowledge Management.
Further information:
http://www.kp-lab.org
www.tuke.sk/fei-cit/poznat/

Thank you!
Questions?