I-TUTOR Maps
Exploring the theoretical background

Arianna Pipitone, Vincenzo Cannella, and Roberto Pirrone

Department of Chemical, Mechanical, Computer, and Mechanical Engineering (DICGIM)
I-TUTOR overview

• Intelligent Tutoring for Lifelong Learning

• An *AI enriched* VLE, which supports
  • Monitoring
  • Instructional design
  • Self-regulation in students
I-TUTOR overview – The plugin

• I-TUTOR plugin functionalities:
  • Maps
  • Monitoring
  • Profiling
  • Alerting

• I-TUTOR supports multilingual Moodle courses.
I-TUTOR overview - Users

- Three kinds of users with different needs
  - Instructional designer
  - Tutor
  - Student
Users - Tutor

- Monitoring students
  - Single student
  - Entire class

- Student and class activities
  - Time spent on studying
  - Contents produced by the student/class
  - Social activities of the student/class
## Users - Tutor

### Monitoring (students vs time)

<table>
<thead>
<tr>
<th></th>
<th>At a given time</th>
<th>Over a period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single student</td>
<td>Disaggregate data analysis</td>
<td>Disaggregate diachronic data analysis</td>
</tr>
<tr>
<td>Entire course</td>
<td>Aggregate synchronic data analysis</td>
<td>Aggregate diachronic data analysis</td>
</tr>
</tbody>
</table>
Users - Student

• Self-monitoring
  • Through time
    - advances throughout the course
  • Proper access to contents
    - Referencing materials to the topics of the course

• Self-regulation
Users – Instructional designer

- Knowledge Domain Authoring
- Course Authoring
  - Overview of the contents of a course
  - Topics of the course
  - Relationships between topics
    - Semantic similarity
    - Pre-requisite (Timing of the contents)
Relevant Processes in I-TUTOR

• Authoring
  • Domain representation
  • Semantic technologies
  • Visualization

• Information Retrieval
  • Navigation
  • Accessing materials
  • Visualization
  • Semantic technologies

• (Self-)assessment
  • Visualization
Domain representation

• How to represent knowledge about a domain
  • A set of facts and events

• Explicit representation
  • Ontologies
  • Conceptual Maps (hypertext…)

• Implicit representation through verbose texts
  • Definitions
  • Learning materials
Explicit Domain Representation

• **Pros**
  - Based on formal description of domain facts and events

• **Cons**
  - Requires meta knowledge about the kind of representation (ontologies, ERD, general taxonomies)
  - High complexity
Implicit Domain Representation

• Pros
  - Direct use of texts
  - Verbose
  - Not structured
  - Easy to implement
  - No technical skills needed

• Cons
  - Needs intensive information analyses techniques
Information Retrieval and Assessment

• Many facets to be managed:
  • Content
  • Course
  • Student
  • Class
  • Studying Activities
  • Social Activities

• …and relations between above-mentioned facets…
Semantic Technologies

- Symbolic analysis and linguistic approaches for NLP
  - Semantic parsing
  - Named entity recognition
  - ...
- Sub-symbolic analysis
  - Machine learning and statistical evaluation
- Explicit vs Latent Semantic
Course Visualization

• Overview of the course
  • Topics
  • Semantic relations between topics
    • Similarity
    • Adjacency
    • Overlapping
    • Hierarchy
  • Topics’sequencing
Content Visualization

• Different kinds of contents
  • Learning materials
  • Contents produced by the students
    • Homeworks
    • Social activities

• Topic-based classification

• Distribution over topics
Activities Visualization

• Studying activities
  • Amount of documents accessed and/or produced by the user

• Social activities
  • Amount of discussions inside the social media and their relation with course topic
The Proposed Solution

A sub-simbolic statistical method for classifying concepts and didactical documents of a course

Creation of a semantic space representing the course domain where data analysis can be performed

New documents and/or activities can be projected into the space or a new classification can be made

Graphic rendering of the space through a ZUI map
I-TUTOR Maps pipeline
Documental Corpora

- Weighted keywords
- Hidden database and keywords definition
- Didactical documents
- Teacher learning materials
- Documents by students
  - Social (forum, chat)
  - Didactical (test answers, notes, and so on)
Preprocessing

• Stemming

• Stop-words removal
TF-IDF

TF-IDF is a numerical statistic evaluation which reflects how important a word is into a collection of document or corpus.

It is computed through the two numbers:

\[ tf_{i,j} = \frac{n_{i,j}}{|d_j|} \]
\[ idf_j = \log \frac{|D|}{|\{d : t_i \in d\}|} \]

where \( n_{i,j} \) is the number of occurrences of term \( t_i \) in the document \( d_j \) and \(|D|\) is the number of documents.

Finally

\[ (tf-idf)_{i,j} = tf_{i,j} \times idf_j \]
Latent Semantic Analysis

LSA analyzes relationships between a set of documents and the terms they contain.

LSA produces a set of concepts related to the documents and terms. LSA assumes that words that are closed in meaning will occur in similar pieces of text.
**LSA Occurrence Matrix**

The LSA Occurrence Matrix describes the occurrences of terms in documents. It is a sparse matrix whose rows correspond to terms and whose columns correspond to documents; We use TF-IDF for weighting the elements of the matrix.

\[
\begin{bmatrix}
    d_1 \\
    \vdots \\
    d_j \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
    x_{1,1} & \cdots & x_{1,n} \\
    \vdots & \ddots & \vdots \\
    x_{m,1} & \cdots & x_{m,n}
\end{bmatrix}
\]
**LSA Decomposition**

For reducing LSA matrix dimension the Singular Value Decomposition (SVD) is applied.

\[
X \rightarrow (U) \rightarrow \Sigma \rightarrow (V^T)
\]

\[
(t_i^T) \rightarrow \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix} = (v_i^T) \rightarrow \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_t \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_t \end{bmatrix} \cdot \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_t \end{bmatrix}
\]
**LSA Spaces**

Document’s space

Concept’s space

I-TUTOR Conceptual and Activity spaces
Self-Organizing Maps

- A type of artificial neural network
  - it is trained through unsupervised learning for producing a map
  - map is a low-dimensional (typically 2D) representation of the input space

- Two operating modes
  - Training: builds the map using input examples
  - Mapping: automatically classifies a new input vector

- Vectors from the semantic space are placed into the map by finding the node with the closest weight vector (in the euclidean sense).
Clustering

• K-means clustering

• Parametric clustering
  • changing keywords weights
I-TUTOR Process Pipe

Generating maps

- Multilinguism Problem
- Graphic Communication
- Visual Code
# Interface – Goals and Solutions

<table>
<thead>
<tr>
<th>Goal</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Looking at contents and topics together in one shot</td>
<td>Concept map</td>
</tr>
<tr>
<td>Easy to understand</td>
<td>Choice of suitable metaphors in the GUI</td>
</tr>
<tr>
<td>Easy to use</td>
<td>Zooming User Interface</td>
</tr>
<tr>
<td>Expressive</td>
<td>Visual Code</td>
</tr>
</tbody>
</table>
Interface - Concept Map and Metaphor

- *Topics* as *Concepts of the domain*
- *Topics and Documents* as *Points in a map*
- *Starry sky* as metaphor for enabling *quick access to contents*
- *Topology and Metrics* as *metaphors* to depict the *Conceptual Space*
Interface - GUI

- Zooming User Interface
  - Recursive nesting
  - Arbitrary level of zoom
  - Easy to interact
    - Reduced number of actions
      - Click
      - Drag
    - Familiarity (Google Maps, ...)

## Interface - Visual Code

<table>
<thead>
<tr>
<th>Graphical element</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colours code (distinct colours for distinct region)</td>
<td>Cluster of documents sharing a common topic</td>
</tr>
<tr>
<td>Brightness</td>
<td>Number of documents in a cluster</td>
</tr>
<tr>
<td>Shapes</td>
<td>Markers to locate studied documents</td>
</tr>
<tr>
<td>Size</td>
<td>Number of studied documents – Spread of a topic in the course</td>
</tr>
<tr>
<td>Spatial closeness</td>
<td>Semantic similarity</td>
</tr>
</tbody>
</table>
Evaluation

- First piloting round for enabling deep technical upgrades
- Second piloting round for making intense evaluation of the maps
  - More than 100 students involved in the courses owned by the partners
- First results are encouraging
  - More than 60% of interviewed people appreciated I-TUTOR as a whole
Future works

• NLP techniques for processing corpora
  • Topic Categorization (Ontology learning)
  • Symbolic approach
  • Semantic annotation

• NLP techniques for social activities
  • Pattern definition and matching
  • Co-reference resolution
  • Anaphors
Future works

• Corpora Clustering
  • Sub-symbolic (Hierarchical clustering, multi-clustering)
  • Symbolic (faceted classification)
Future works

- Visualization
  - New metaphors
  - 3D visualization
  - New facets to describe a student
    - Social Interactions (nets, information flows, roles)
    - Complex Behaviours described as combinations of different facets
  - The task at hand