KNOWLEDGE ENGINEERING GROUP

Contact details

<table>
<thead>
<tr>
<th>Name</th>
<th>Knowledge Engineering Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acronym</td>
<td>KEG</td>
</tr>
<tr>
<td>Logo</td>
<td><img src="image" alt="Knowledge Engineering Group Logo" /></td>
</tr>
<tr>
<td>Site</td>
<td><a href="http://keg.utcluj.ro">http://keg.utcluj.ro</a></td>
</tr>
<tr>
<td>Address</td>
<td>26-28 Baritiu Str., rooms C09, D01, M03; 400027, Cluj-Napoca, Romania</td>
</tr>
<tr>
<td>Faculty</td>
<td>Faculty of Automation and Computer Science</td>
</tr>
<tr>
<td>Department</td>
<td>Computer Science Department</td>
</tr>
<tr>
<td>Telephone</td>
<td>+40 264 202389</td>
</tr>
<tr>
<td>Fax</td>
<td>+40 264 594491</td>
</tr>
<tr>
<td>Director</td>
<td>Prof. Dr. Eng. Rodica Potolea</td>
</tr>
<tr>
<td>e-mail</td>
<td><a href="mailto:rodica.potolea@cs.utcluj.ro">rodica.potolea@cs.utcluj.ro</a></td>
</tr>
</tbody>
</table>

Areas of expertise

**Knowledge extraction from data**: Data Mining and Classification (incomplete records and irrelevant and/or redundant pieces of information, imbalanced class distribution and error costs; metric, algorithm and model selection; domain-specific application constraints) and Information Retrieval (schema mapping and data fusion; IR from unstructured sources; context-sensitive IR; user profiling; data quality, trust)

Team and key skills

**Prof. Dr. Eng. Rodica Potolea**, Head of the CS Department, PhD Supervisor

- Competence areas: Data mining, classification, information retrieval, logic programming, algorithms and complexity
- Member in 20 past research projects, more than 100 published papers in Journals and Conference proceedings (ISI and international databases indexed)

**Assoc. Prof. Dr. Eng. Mihaela Dinsoreanu**

- Competence areas: software engineering, data integration, information retrieval
- Member in 7 past research projects, more than 50 published papers in Journals and Conference proceedings (ISI and international databases indexed)

**Assist. Prof. Dr. Eng. Camelia Lemnaru**

- Competence areas: cost-sensitive learning, learning from imbalanced/large data
- Member in 3 past research projects, more than 40 published papers in Journals and Conference proceedings (ISI and international databases indexed)

Development strategy

The basic objective of our group is to advance our knowledge base by tackling real-life problems (medical, financial, social). In this respect we aim the development of collaborative, interdisciplinary research proposals, involving early career researchers by continuously mentoring high-quality BSc and MSc students in research leading to relevant PhD thesis.

Representative projects

**NOKIA** – Context-sensitive recommendation systems

The aim of this project is the research and development of a general context-sensitive recommendation system considering the following context features: content topics and user profiles. The project involves the design of a general context model and the associated similarity metrics based on lexical and semantic features.


The main goal of the project was to define a model of an adaptive e-learning environment, using Concept Maps. Our main concern was the correct and continuous identification of the user learning style, and the provision of the most
Significant results

The most representative publications of the past 5 years


Remarkable results achieved in the projects in the last 5 – international patents, products used by beneficiaries, technologies adopted by companies

Our efforts to integrate outstanding students in the research process has the following outcomes:
- Early publication of the results in papers at International IEEE conferences
- Recognition of the quality of the work: awards at Computer Science Students Conference, and normalization for the best paper award at KDIR 2012
- The research work is considered with relevance for the industry, as one of our bachelor thesis has been awarded by industry (ISDC) as best thesis in the 2011-2012 academic year.

Each of the projects provided a functional prototype that can be used by the intended beneficiaries:

**Nokia prototype**

We designed and developed a three-layered matching model that aims to identify the most relevant content in a given context. Our model consists of three components: a lexical, a semantic, and a user-profile component and the corresponding composed similarity metrics. We instantiated our model for the online advertisement problem and the preliminary evaluations showed promising results in terms of information relevance.

**SEARCH prototype**

User profiling under dynamic environment: identify the profile of each individual interacting with the system, based on initial (static) features and the interaction with the system (dynamic features) and providing learning content according to the actual profile. The user classification module – which has been implemented by means of a SOM – allows for a seamless transition from the initial static to the static and dynamic user assessment.
GridMOSI prototype
Benchmarking numerical algorithms: parallel/distributed deployment of a set of numerical algorithms frequently employed in applications. A set of best practice techniques has been developed as a result of the experiments, to relate the context of the problem to be solved with the specific algorithm to be employed, the size of the input data with the number of processors on which to run the application, and the communication type/quantity with the size of data transferred.

The offer addressed to the economic environment

<table>
<thead>
<tr>
<th>Research &amp; development in core areas</th>
<th>Context-sensitive content retrieval and (dis)similarity measures</th>
<th>extracting relevant content for a given context; data sources: free-text documents or data gathered via measuring/sensors/investigations. Main methods: lexical and/or syntactical analysis, ontology-driven concept identification, query expansion, dealing with negation and/or (in)consistencies, (dis)similarity measures. Supervised/semi-supervised/unsupervised machine learning techniques – dealing with fundamental learning techniques, their current trends and applications, and the specific instantiations for very large, noisy, incomplete, imbalanced, and/or unstructured data. Text mining – extracting knowledge from unstructured text via text processing methods, taxonomy and/or ontology support, learning techniques and domain-specific information. Potential beneficiary fields: medical, economic, legal, social, legal. Social networks mining – deriving social relationships and behaviours from existing explicit or implicit connections. Identifying clusters, membership relations, habits, social patterns, trends and needs for individuals belonging to a group. Opinion mining and Sentiment analysis – extracting opinions from reviews and/or other unstructured free-text; dealing with deceptive content, negation. Identify trends in different groups from their indirect opinions. Data integration – integrating heterogeneous data by ontology-driven, (semi-) automatic design of unified data structures and automatic design of the corresponding ETL processes.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research &amp; development in applied fields</td>
<td>Recommendation systems in different areas – developing prototype recommendation systems according to state of the art techniques in the field and up-to-date technologies. Topic extraction and representation – identifying concepts to be representative in a given context; projecting (very) large (un)structured data to relevant dimensions and providing representation to allow knowledge extraction. Community detection – identifying clusters from implicit and/or explicit connections. User profiling – finding groups of individuals with similar features, finding/defining patterns for various profiles, predicting trends and future behaviour. Schema mapping and data fusion – designing unified data(warehouse) structures to integrate heterogeneous data sources, designing corresponding ETL processes. Decision support systems – extracting knowledge from organizational data, predicting evolutions, trends, identify relationships and correlations.</td>
<td></td>
</tr>
<tr>
<td>Consulting</td>
<td>Knowledge extraction for decision making, Data integration, Data structure design</td>
<td></td>
</tr>
<tr>
<td>Applied engineering services</td>
<td>Knowledge engineering solutions applied in industrial fields</td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>N/a</td>
<td></td>
</tr>
</tbody>
</table>