STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LII**, Number 1, 2007

THE FIRST INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING PRINCIPLES AND TECHNIQUES (KEPT 2007)

DOINA TĂTAR, HORIA F. POP, MILITON FRENŢIU, AND DUMITRU DUMITRESCU

1. Introduction

The Faculty of Mathematics and Computer Science of the Babeş-Bolyai University in Cluj has organised during June 6-8, 2007, the First International Conference on Knowledge Engineering Principles and Techniques (KEPT 2007), a beginning of what we would like to be a permanent series of events on theoretical foundations and real-world applications of knowledge engineering. This conference, organised on the platform of Knowledge Engineering, has been a forum for intellectual, academic, scientific and industrial debate to promote research and knowledge in this key area, and to facilitate interdisciplinary and multidisciplinary approaches, more and more necessary and useful today.

The conference was honoured by leading class keynote lecturers, to present their invited lectures in two plenary sessions: Prof. Bruno Buchberger (Johannes Kepler University Linz), with a lecture on "Algorithm Synthesis by Lazy Thinking: A Case Study in Mathematical Knowledge Engineering", Prof. Diana Inkpen (University of Ottawa, Canada), with a lecture on "Semantic Similarity Knowledge and its Applications", Prof. Alain Lecomte (University Paris 8, France), with a lecture on "Some Representation Structures for Computational Linguistics", Prof. Rada Mihalcea (University of North Texas, USA), with a lecture on "Using Wikipedia for Automatic Word Sense Disambiguation", and Dr. Constantin Orasan (University of Wolverhampton, UK), with a lecture on "The Role of Linguistic Information for Shallow Language Processing".

The organisation of this conference reflects three major areas of concern: Natural Language Processing, Artificial Intelligence, and Software Engineering. Oral presentations of 41 regular papers are organized in 12 dedicated sessions, planned with sufficient length to encourage the direct dialogue and exchange of ideas among researchers.

Received by the editors: July 12, 2007.

Knowledge engineering refers to the building, maintaining, and development of knowledge-based systems. It has a great deal in common with software engineering, and is related to many computer science domains such as artificial intelligence, databases, data mining, expert systems, decision support systems and geographic information systems. Knowledge engineering is also related to mathematical logic, as well as strongly involved in cognitive science and socio-cognitive engineering where the knowledge is produced by socio-cognitive aggregates (mainly humans) and is structured according to our understanding of how human reasoning and logic works.

Since the mid-1980s, knowledge engineers have developed a number of principles, methods and tools that have considerably improved the process of knowledge acquisition and ordering. Some of the key issues follow: there are different types of knowledge, and the right approach and technique should be used for the knowledge required; there are different types of experts and expertise, and methods should be chosen appropriately; there are different ways of representing knowledge, which can aid knowledge acquisition, validation and re-use; there are different ways of using knowledge, and the acquisition process can be goal-oriented; there are structured methods to increase the acquisition efficiency.

2. Opening lectures

KEPT 2007 conference has been opened with three very actual and interesting lectures. First is called "Semantic Similarity Knowledge" and was done by Dr. Diana Inkpen from University of Ottawa, Canada. It present several methods for computing the similarity of two words, following two directions: dictionary-based methods that use WordNet, Roget's thesaurus, or other resources, and corpus-based methods that use frequencies of co-occurrence in corpora (cosine method, latent semantic indexing, mutual information, etc). Also, several applications of word similarity knowledge are presented: detecting words that do not fit into their context (real-word error correction), detecting speech recognition errors, solving TOEFL-style synonym questions, and synonym choice in context (for writing aid tools).

The second lecture is "The Role of Linguistic Information for shallow Language Processing" and the author is Dr. Constantin Orasan from School of Humanities, Languages and Social Sciences, University of Wolverhampton. The lecture argues that the advantage of shallow methods in comparison to deep processing methods is that they do not require the building of elaborate representations of the text to be processed or to perform reasoning on this data, and as a result they can be more easily implemented. The talk shows how a shallow method for automatic

summarization can improve its performance by adding different types of linguistic information.

The final talk was that presented by Dr. Rada Mihalcea, from the University of North Texas with the name "Using Wikipedia for automatic Word Sense Disambiguation". Starting with the hyperlinks available in Wikipedia, the author shows how one can generate sense annotated corpora that can be used for building accurate and robust sense classifiers.

3. The Natural Language Processing section

To continue with the Natural Language Processing section, the following papers have been presented, accordingly with the order of presentation.

It is known that he task of function tagging involves labeling certain nodes in an input parse tree with a set of functional marks such as logical subject, predicate, etc. The paper "Large Scale Experiments with Function Tagging" shows haw two Decision Trees based approaches to the task of function tagging outperform baseline approaches when the most frequent tag is assigned.

The second paper in NLP section was "A chain dictionary method for Word Sense Disambiguation and applications". The approach of WSD, which is one of the most important open problems in NLP, uses the lexical base WordNet for a new algorithm originated in Lesk's, namely "chain algorithm for disambiguation" of all words (CHAD). Some experiments and evaluations with CHAD for Semcor and Senseval 1 corpora are described, which prove the performance of the algorithm. Conclusions of using the CHAD for machine translation (here from Romanian language to English) and for text entailment verification are discussed.

The following paper, "Text entailment verification with text similarities", presents a method based on lexical resolution and supposes the word sense disambiguation of the two texts (text and hypothesis). The method also relies on a recent directional measure of semantic similarity between two texts and is applied to the dataset of PASCAL RTE-2.

The paper "Syntagma Processing for incomplete answers" refers to the syntagma as an incomplete answer that resumes to a phrase not to a grammatically correct sentence. SPEL (Syntactic Parser for English Language) system is introduced as an method that is able to reconstruct the answers from the given syntagma and the initial question, without affecting the semantic information given by the answer. The paper proposes a solution to incomplete answer processing, answers that are very frequent in a usual communication scenario based upon question-answer pattern.

Named "A Text Analysis Based Approach for the Compliance between the Specification and the Software Product", the next paper proposes a new approach in evaluating the compliance between software documentation (expressed on natural language) and the final software product. The authors define two evaluation measures and present some case studies.

"Text Categorization experiments using Wikipedia" shows how to use Wikipedia articles to give word distributional representation for documents. Since the word-distributional representation causes dimensionality increase, dimensionality reduction is needed to make the problem computationally tractable. The authors use in this respect a method known as latent semantic indexing (LSI) and combine this with the processing of the training corpus. The results of experimenting with the method on real-world dataset is presented.

The paper "The 'Integral' Model of Language Functioning (E. Coseriu)" explores the framework of Coseriu's "integral linguistics", focusing mainly on the three planes of language and their corresponding "linguistics" - the three directions in language investigation that Coseriu postulated. It is argued that, in the panorama of contemporary approaches to language, Coseriu's integral linguistics offers one of the most comprehensive and finely articulated frameworks for investigating the functioning of language in a dynamic perspective. This paper relies on experiences of a pure linguistic team and could seed a further collaboration with our NLP researchers.

"Enhancing the Invisible Web" is the title of next presented paper. This article describes the architecture of an Invisible-Web Extractor, whose primal goal is to enhance the value of the hidden Web data. The author considers three main issues of the tool: how to access the Invisible Web information, how to extract information from the gathered data and how to create new knowledge from it.

In "Chain Algorithm used for Part of Speech Recognition" the author shows how CHAD algorithm (see above) can be used to identify the part of speech of words from a text written in a single language.

Natural Language Generation (NLG) and Foreign Language Writing Aid (FLWA) are two important tasks of Natural Language Processing (NLP), which deal with obtaining natural language from a machine representation system and building computer programs that assists a non-native language user in writing decently in a target language, respectively. The paper "Natural Language Generation - Applications for Romanian Language" uses an affix grammar to construct the Romanian language grammar and a semantic which gives us information about the words we use to build a sentence. It shows how one can construct, starting from a set of words, correct sentences from syntactic and semantic point of view.

The short enumeration of the conference papers presented at KEPT2007 gives an image of the diversity and the depth of tackled problems. We can say that NLP section of Kept2007 was a success and we hope for a continue development in the further editions.

The second day started with the wonderful conference of Prof. Bruno Buchberger (Johannes Kepler University Linz), with a lecture on "Algorithm Synthesis by Lazy Thinking: A Case Study in Mathematical Knowledge Engineering". The lecture stated that Mathematical Knowledge Engineering equals the Algorithm-supported Mathematical Theory Exploration: invent axioms and definitions for notions (functions and predicates), invent and prove theorems about notions, invent problems about notions. The talk focused to the author's method of invention (synthesis) of correct algorithms from problem specifications by systematic reasoning by the "lazy thinking" method. The method is implemented in the Theorema system on top of Mathematica. An example of non-trivial problem is the construction of Grbner bases, which the author presented in details.

4. The Artificial Intelligence section

The Artificial Intelligence section of the conference is focused on Computational Intelligence techniques. The main topics concern Evolutionary Computing, probabilistic neural networks, multi-agent systems and complex networks. Some papers in this section introduce new evolutionary models and concepts applying them for solving complex optimization problems.

A new selection operator based on the family line of each individual is proposed and tested for solving TSP problems (paper "Collaborative Selection for Evolutionary Algorithms").

A novel trajectory based technique intended to solve complex problems via hierarchical decomposition is presented in the paper 'Exact Model Building in Hierarchical Complex Systems'. The potential of the method resides on the ability to aggressively explore the building block space. Two papers deal with the application of genetic chromodynamics metaheuristic in data mining, training and job scheduling. The paper 'Multi-agent Distributed Computing' investigates the potential of intelligent agents to support distributed collaboration environments. A new multi-agent knowledge management and support system in designed and evaluated.

The potential of hybridization between multi-agent systems and nature-inspired metaheuristics such as Ant Colony Systems is explored in the papers 'Stigmergic Agent Systems for solving NP-hard Problems' and 'Sensitive Ant Systems in

Combinatorial Optimization'. Numerical experiments indicate that the emerging models are very promising for solving search problems.

An interesting approach for generating blood networks is described in the paper 'Simulating Microcapillary Networks using Random Graphs'. Solving significant NP-hard problems using evolutionary algorithms and hybrid techniques is an important issue of Computational Intelligence that has been addressed by four papers. The differential evolution technique is used for unsupervised clustering of documents. The paper 'An Evolutionary Model for Solving Multiplayer Noncooperative Games' proposes a technique for detecting multiple Nash equilibria. The method can be extended to deal with cooperative games.

A hybrid technique for parameter setting in probabilistic neural networks is applied for hepatic cancer diagnosis. A supervised version of learning vector quantization induces a neural network for mining a toxin database.

The papers in this section form an important contribution to the field of Natural Computing and address some significant practical applications. New powerful computational models for search and optimization are proposed and some interesting hybridization techniques with a great potential are investigated.

5. The Software Engineering section

Software Engineering refers to specification, design, coding, verification, and maintenance of software. It is connected to knowledge engineering since it implies transforming the requirements of the clients (i.e. knowledge from their domains) into specifications of corresponding software product. In other words, software engineering is dealing with the transformation of knowledge into software. That's why the third section of KEPT2007 was Software Engineering , with its various subdomains.

In the paper "On Software Attributes Relationship Using a New Fuzzy C-Bipartitioning Method", the authors introduce a new data analysis method, the fuzzy bipartitioning method, and use this method to study the dependence between software attributes.

The second paper, "Some Formal Approaches for Dynamic Life Session Management", introduces three formal approaches for determining, establishing, and maintaining the lifetime of a HTTP session.

The third paper, "Management of Web Pages Using XML Documents" describes a method of automatic management of a WEB site, by memorizing in a database the information sources in different pages.

"A View on Fault Tolerant Techniques Applied for Mediogrid", analyses the characteristics of fault tolerance for grid systems. Some directions for enhancing the fault tolerance of the MedioGrid are suggested.

Then "A New Graph-Based Approach in Aspect Mining" presents a new graph-based approach in aspect mining. The cross cutting concerns is viewed as a search problem in a graph, and an algorithm, GRAM, is given to solve the problem.

Next paper, "Introducing Data-Distributions into PowerList Theory", introduces data-distributions into PowerList theory in order to reconcile abstraction of this theory with performance.

The notion of internally and externally stable set for the G-complex of multi-ary relations are defined in "The Stable Sets of a G-complex of Multi-ary Relations and its Applications" Also, some properties of these sets are proven.

The paper "Multi-Agent System for Competence Modeling" gives a model for competencies of people. This model is useful to the universities, and companies, to study the competencies.

Then, "Data Verification in ETL Processes" uses metrics on partitions based on Shannon entropy in the verification of consistency of data loaded into the data warehouse by ETL processes.

Next paper, "Data Predictions using Neural Networks", proposes the Artificial Neural Networks as trainable tools that attempt to mimic information processing patterns in the brain, and use them for data analysis and prediction.

The paper "Evaluating Dynamic Client-Driven Adaptation Decision Support in Multimedia Proxy-Caches" evaluates the use of dynamic client-driven adaptation decision support in multimedia proxy-caches through the use of the Adaptation-awareMultimedia Streaming Protocol.

In "Automated Proof of Geometry Theorems Involving Order Relation in the Frame of the Theorema Project" the author proposes a method that combines the area method for computing geometric quantities and the Cylindrical Algebraic Decomposition method, in order to prove geometry theorems. An implementation of this method as part of Geometry Prover in the frame of Theorema project is done.

The following paper, "A Hierarchical Clustering Algorithm for Software Design Improvement", presents a new hierarchical clustering algorithm that can be used for improving software system design. This approach may be used to assist software engineers in refactoring software systems.

In the paper "Metrics-Based Selection of a Component Assembly" some software metrics are used to select the specified components required by the design of a system.

The authors of the paper "Architecting and Specifying a Software Component using UML" suggest a component-based architecture for LCD Wallet Travelling Clock case study.

In the next paper, "A TSpaces Based Framework for Parallel-Distributed Applications", a framework to deploy and execute parallel-distributed applications is suggested.

Finally, in the paper N. Magariu, Applying Transition Diagram Systems in Development of Information Systems Dynamic Projects a model of complex software development based on the usage of transition diagrams system.

6. Conclusions

The First International Conference on Knowledge Engineering Principles and Techniques (KEPT 2007) was an exciting and useful experience and exchange of knowledge for our department. The possibility to communicate our most recent studies, the commitment with the results of others colleagues, the emulation of new ideas and research, all these mean a great gain of experience in our professional life. We hope that the next edition of KEPT (in 2009) will be even more successful and more enthusiastic than this one.

So, let us see you at KEPT 2009!

E-mail address: dtatar@cs.ubbcluj.ro
E-mail address: hfpop@cs.ubbcluj.ro
E-mail address: mfrentiu@cs.ubbcluj.ro
E-mail address: ddumitr@cs.ubbcluj.ro

Babeş-Bolyai University, Department of Computer Science, 1, M. Kogălniceanu St., 400084 Cluj-Napoca, Romania