Automated Model Extraction
from Diverse Knowledge Sources

Aims

Many disciplines of contemporary computer science such as artificial intelligence, Semantic Web, software and web engineering, or information systems are based on different types of models. By models, we mean any representation that synthesizes knowledge in a form readable and manageable by a machine. Models can be classified according to many different criteria, for instance, some models such as domain ontologies, topic maps and concept maps (mostly used in the artificial intelligence discipline) are used to express general knowledge of a domain. Others such as class diagrams in UML (*Unified Modeling Language*) models, business rules in RIF (*Rule Interchange Format*) or SBVR (*Semantics of Business Vocabulary and Business Rules*), business process models in BPMN (*Business Process Modeling Notation*), or integrity constraints in OCL (*Object Constraint Language*), are specialized to represent different kinds of knowledge needed in very specific phases of the software engineering lifecycle from the initial requirements elicitation to the maintenance and performance analysis.

One common point of these models is that they are hard to build, maintain and exchange, due to their multiple formalisms and due to the well-known knowledge acquisition bottleneck. In fact, manual knowledge elicitation from domain experts is not a straightforward task and manually building a knowledge base from scratch is not easy. To overcome this bottleneck, using automatic methods for knowledge extraction and update seems to be a promising avenue. In particular, this special issue has a special interest on the extraction of models from unstructured knowledge sources such as natural language texts, but other knowledge sources (relational and non-SQL databases, XML documents, HTML pages …) and the composition of specific models (SBVR, OCL, topic maps …) to represent a more general view of the domain knowledge are also welcome. This special issue has the objective of increasing collaboration of researchers from the Semantic Web, Natural Language Processing, Data Mining, Information Retrieval, Information Systems, Business Process Management, and Software Engineering communities.

Among the questions that we seek to answer are the following:

- Do we require different extraction approaches based on the type of documents (e.g., requirements, general domain, or Web) that are being processed?
- Do we require different levels of analysis (e.g., shallow, deep, or patterns) depending on the type of documents?
- What are the methodological aspects of the knowledge extraction process?
- To what extent can we use reasoning to enrich/evaluate the extracted models?
- How can we integrate the information coming from various knowledge sources?
- How can we filter the important knowledge in a vast amount of data?
- How can we ensure the reuse of the developed models?
- Can we provide guidelines for the evaluation of the extracted models?
- How can we ensure that the extracted models are effective?

- What are the underlying principles for creating effective user interfaces for extraction and evolution of models?
- How can we deal with complexity aspects?
- What tool support is needed?

Topics

This special issue is concerned with all aspects of automatically acquiring, evolving, and translating knowledge models including (but not limited to):

- Model extraction from various (unstructured) sources including requirement documents, domain documents and Web documents with a special emphasis on domain-independent and unsupervised information extraction methods;
- Conversion from one formalism to the other through meta-models and standardized transformation approaches;
- Ontology learning and update including concept, hierarchical relations, conceptual relations and axioms;
- Ontology extraction from pre-existing software engineering models or unstructured knowledge sources;
- Ontology learning and update as a workflow;
- Generative and transformational techniques in software engineering for knowledge and information extraction;
- Meta-CASE tools used for information extraction purposes;
- Domain-specific languages for knowledge extraction, annotation and integration;
- Text mining and natural language processing for extraction of models;
- Deep and shallow semantic analysis in parsing (e.g., natural language) for extraction of models;
- Knowledge integration and reasoning for building models;
- Models for software engineering extraction and maintenance (such as UML, SBVR, and OCL);
- Important knowledge identification, ranking, and filtering including important assertions, concepts and relations in a model;
- The evaluation of the obtained models through metrics, quantitative methods and finally qualitative methods;
- Empirical studies related to the different issues of importance for automated model extraction such as (linguistic) analysis of vocabularies used for labelling model elements from diverse knowledge sources;
- User interfaces and usability studies for the extraction and integration of knowledge sources;
- Collaborative approaches for extraction of models (e.g., based on social networking, collaborative tagging, etc.)

Submission

The all submissions will follow the submission, quality expectation, review, decision making, and publication rules and policies of the journal

Proposed Guest Editors

Jordi Cabot is currently leading the AtlanMod team, an INRIA research group at École des Mines de Nantes (France). Previously, he has been a post-doctoral fellow at the University of Toronto, a senior lecturer at the UOC (Open University of Catalonia) and a visiting scholar at the Politecnico di Milano. He received the BSc and PhD degrees in Computer Science from the Technical University of Catalonia. His research interests include conceptual modeling, model-driven and web engineering, formal verification and social aspects of software engineering. He has written more than 50 publications in international journals and conferences in the area and co-organized several national and international modeling events. He is currently the co-PC Chair for the Int. Conf. on Model Transformations (ICMT). Apart from his scientific work, he writes and blogs about all these topics in his Modeling Languages portal (http://modeling-languages.com). He is a member of the IEEE and the ACM.

Michel Gagnon is a professor at the Computer Engineering Department of Polytechnique Montreal since 2002. Previously, he worked as a team leader at Machina Sapiens inc., a company which at that time was a leader in the development of grammar checkers, and as a professor at the Univerdade Federal do Parana, in Brazil. He received his Ph.D. degree in computer science in 1993 from the Université de Montreal. Since then, he has been working on natural language processing, with a special attention to semantics. Since 2002, his research activities also include the semantic web, especially its industrial applications. He was a co-chair of the TALN 2010 conference, the main scientific event in French for researches in natural language processing. Currently, he is a co-leader of the Gitan project, whose objective is the processing of texts to produce 3D animations that convey their semantic contents.

Dragan Gašević (http://www.sfu.ca/~dgasevic/) is a Canada Research Chair in Semantic Technologies and an Associate Professor in the School of Computing and Information Systems at Athabasca University. He is also an Adjunct Professor in the School of Interactive Arts and Technology at Simon Fraser University and an associated research member of the GOOD OLD AI Research Network at the University of Belgrade. He is a recipient of Alberta Ingenuity's 2008 New Faculty Award. His research interests include semantic technologies, software language engineering, technology-enhanced learning, and service-oriented architectures. He has (co-)authored more than 200 research papers. He is the lead author of the book monograph Model Driven Architecture and Ontology Development published in two editions by Springer in 2006 and 2009.He was the founding program co-chair (2008) and general chair (2009), and is the current co-chair of the steering committee of the International Conference on Software Language Engineering (SLE). Dragan is also a steering committee member and a founding program co-chair (2011) of the new series of international conferences on Learning Analytics and Knowledge (LAK). Dragan has also been a (co-)chair of more than ten editions of workshops (ATEM, VORTE, DDBP, NFPinDSMLs) that have successfully been organized at major conferences such as MODELS and EDOC. Dragan has also organized and edited many books and special issues with major publishers (Springer) and prestigious international journals (e.g., IEEE TSE, IEEE SMCC, and Elsevier's Information Systems). Being an often invited and keynote speaker at many

events, Dragan has given tutorials at many international conferences (e.g., WWW, ISWC, and ESWC).

Amal Zouaq (http://azouaq.athabascau.ca/) is an Assistant Professor at the Royal Military College of Canada. Previously, she was a postdoctoral researcher at Simon Fraser University and Athabasca University and was funded by the FQRNT (Fonds Québécois de la Recherche sur la Nature et les Technologies). She also had a postdoctoral experience at the Ecole Polytechnique de Montréal. Her research interests include natural language processing, Semantic Web, ontology engineering, knowledge extraction and technology-enhanced learning. She serves as a member of the program committee of various conferences in technology-enhanced learning and Semantic Web and she is part of the editorial review board of the Interdisciplinary Journal of E-learning and Learning Objects. She also serves as a reviewer in many conferences and journals in knowledge and data engineering, natural language processing, eLearning and the Semantic Web.

Schedule

Deadline for paper submission: September 30, 2011

Completion of first review: January 30, 2012

Major revision due (first round): March 15, 2012

Completion of second round of reviews: May 1, 2012

Minor revision due (second round): May 31, 2012

Final decision notification: June 30, 2012

Publication date: TBD