AMMoRe 2018: First international workshop on analytics and mining of model repositories

Önder Babur
Eindhoven University of Technology
Eindhoven, The Netherlands
o.babur@tue.nl

Michel R. V. Chaudron
Chalmers | University of Gothenburg
Gothenburg, Sweden
chaudron@chalmers.se

Loek Cleophas
Eindhoven University of Technology
Eindhoven, The Netherlands
Stellenbosch University
Stellenbosch, Republic of South Africa
l.g.w.a.cleophas@tue.nl

Davide Di Ruscio
University of L'Aquila
L'Aquila, Italy
davide.diruscio@univaq.it

Dimitris Kolovos
University of York
York, United Kingdom
dimitris.kolovos@york.ac.uk

ABSTRACT

Model-based approaches promote the use of models and related artifacts (such as metamodels and model transformations) as central elements to tackle the complexity of building systems. Both in academia and in industry there is a growing need to efficiently i) store; ii) analyze; and ii) search & navigate, and iii) curate large collections of models. Such collections include for example large sets of software models such as the Lindholmen UML dataset [1], or of heterogeneous models in large MDE ecosystems and systems-of-systems, including e.g. software, hardware, and business models. The workshop Analytics and Mining of Model Repositories (AMMoRe) aims to gather modelling researchers and practitioners to discuss the emerging problems and propose solutions. The scope ranges from industrial reports and empirical analyses in the problem domain to novel cross-disciplinary approaches for large-scale analytics and management, e.g. exploiting techniques from data analytics, repository mining and machine learning.

1 OBJECTIVES AND SCOPE

Big data approaches are causing large changes in the way we can perform science and business. Big Data is also entering the arena of software engineering and software modelling. We want to bring together the communities of Big Data/Machine Learning and Software Modelling. Various datasets of models have now become available and now our community must learn methods, techniques and tools for analyzing these large datasets. Many such methods, techniques and tools are known from the Big Data/Machine Learning and Information Retrieval/Natural Language Processing communities. How they can be adapted and applied to models and model repositories is an open question. Conversely, the insights that come out of this may lead to insights for these communities that are usable beyond software modelling.

Undoubtedly, MODELS is the premier conference series for model-driven software and systems engineering. It tries to cover all aspects of modeling, yet analytics and mining of model repositories (and other large collections of models) has not been a prime focus or the topic of a MODELS workshop, nor of any other event in the community so far (to the best of our knowledge). Yet it is an increasingly timely and important topic.

AMMoRe aims to bring together researchers in software modelling, model repositories and model analytics, big-data and machine learning, information retrieval and natural language processing. Orthogonal to these domains, it welcomes contributions from a wide range of technical spaces to promote cross-fertilization: Model-Driven Engineering, Systems Engineering, Business Process Modelling, Software Architecture and so on. The topics of interest of AMMoRe 2018 include:

- Industrial reports, empirical studies on model corpora, applications of model corpora
- Repository mining and management for modelling artefacts
- Clone-, pattern-, aspect-mining for modelling artefacts
- Applications of exploratory or descriptive data analytics, predictive analytics, machine learning or deep-learning
- Large-scale model management and consistency checking
- Natural language processing for modelling
- Model searching, indexing, retrieval, storage
- Linking, enriching and labelling model-repositories
- Visualization of (possibly heterogeneous) large sets of modelling artifacts
- Techniques to analyze and automate (co-)evolution in modelling
- Variability mining and management, model-driven software product lines
- Distributed computing for modelling, with an eye towards Big Data
- Intelligent techniques for automating modelling tasks
- Building and composing model-analytics workflows (based on online services/repositories)

While this is the first edition of this workshop per se, a Dutch Symposium on Model Management and Analytics in October 2017 (with invited talks) was well-attended, attracting 23 participants from 6 universities and 7 companies, and 11 talks. The corresponding website [1], which is being used for a range of related events on similar topics (including AMMoRe), has so far attracted 2k+ views from 60+ countries.

---

[1] http://modelanalytics.wordpress.com

2 WORKSHOP ORGANIZATION

Önder Babur, Michel Chaudron, Loek Cleophas, Davide Di Ruscio and Dimitris Kolovos organized and chaired the program committee (PC) for the first edition of AMMoRe. Each of the submissions was reviewed by at least four PC members and the papers were selected based on their relevance to the workshop's topics and the reviews provided by PC members. The AMMoRe 2018 PC consisted of:

- Nour Assy (Eindhoven University of Technology, NL)
- Alessandra Bagnato (Softeam, FR)
- Ludovico Iovino (Gran Sasso Science Institute L'Aquila, IT)
- Yannis Korkontzelos (Edge Hill University, UK)
- Henrik Leopold (Vrije Universiteit Amsterdam, NL)
- Ivano Malavolta (Vrije Universiteit Amsterdam, NL)
- Nicholas Matragkas (University of York, UK)
- Richard Paige (University of York, UK)
- Alfonso Pierantonio (University of L'Aquila, IT)
- Ivan Polasek (Slovak University of Technology & Gratex International, SK)
- Gregorio Robles (Universidad Rey Juan Carlos, ES)
- Christoph Seidl (Technische Universität Braunschweig, DE)
- Matthew Stephan (Miami University, US)
- Daniel Strüber (University of Koblenz and Landau, DE)
- Bedir Tekinerdogan (Wageningen University, NL)
- Mark van den Brand (Eindhoven University of Technology, NL)
- Maurice van Keulen (University of Twente, NL)
- Barbara Weber (Technical University of Denmark, DK)
- Manuel Wimmer (Vienna University of Technology, AT)

3 PROGRAM

AMMoRe attracted 5 submissions on a variety of topics related to the analytics and mining of model repositories. The final program included one invited talk and 3 paper presentations:

- *Process model management and analytics* (invited talk, by Barbara Weber) gives an overview of the invited speaker's past 10+ years research on business process modeling and model management, including the adaptation and evolution of process models, the management of process model repositories, as well as model analytics conducted online as needed for e.g. personalized modeling environments.
- *Automatic model repair using reinforcement learning* (by Angela Barriga, Adrian Rutle and Rogardt Heldal) proposes to automatically fix broken models by using machine learning. Since appropriate data sets are seen as lacking, the authors' prototype tool uses reinforcement learning, which does not need initial data sets. A case study was performed with this prototype on a set of broken models.
- *Model analytics for feature models: case studies for S.P.L.O.T. repository* (by Önder Babur, Loek Cleophas and Mark van den Brand) adapts the model analytics framework SAMOS, which uses information retrieval techniques to compare models, to the setting of feature models, and discusses two case studies in which the resulting framework was used to get insight into collections of feature models.
- *Exploring model repositories by means of megamodel-aware search operators* (by Francesco Basciani, Juri Di Rocco, Davide Di Ruscio, Ludovico Iovino and Alfonso Pierantonio) aims to improve model artefact reuse by presenting a novel approach to model search. Repositories are structured into megamodels, with specific search operators allowing effective exploration and browsing of the model repositories. A case study implements this in the MDEForge platform, using the Lucene search library.

4 OUTLOOK

With this first edition of AMMoRe, we aim to build and strengthen an audience from various domains and communities working on analytics and mining of model repositories. We hope to increase our effort and organize further iterations of this workshop and other follow-up events, with the goal of attracting more attention to these timely and important topics.

ACKNOWLEDGEMENTS

REFERENCES

[1] Regina Hebig, Truong Ho-Quang, Michel R. V. Chaudron, Gregorio Robles, and Miguel Angel Fernández. 2016. The quest for open source projects that use UML: mining GitHub. In *Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems, Saint-Malo, France, October 2-7, 2016*, Benoit Baudry and Benoît Combemale (Eds.). ACM, 173–183. https://doi.org/10.1145/2976767