

Workshops and Tutorials at K-CAP2017

Proceedings Preface

Giuseppe Rizzo

ISMB
Turin, Italy
giuseppe.rizzo@ismb.it

The International Conference on Knowledge Capture (K-CAP) provides a forum that brings together members of diverse research communities who are interested in efficiently capturing knowledge from a vast range of sources and in creating representations that can be useful for building knowledge-intensive autonomous applications. Numerous research fields are investigating and applying these aforementioned research lines and they include natural language processing, machine learning, knowledge management, and semantic web. Besides the traditional research track, K-CAP usually hosts workshops and tutorials on topics related to the theme of the conference. In particular, workshops aim to provide opportunities for exchanging views, advancing ideas, and discussing preliminary results in an atmosphere that fosters the active exchange of ideas. Workshops are usually held before the conference and prepare the attendees to the discussions during the conference. Tutorials enable attendees to fully appreciate current research trends, main schools of thoughts, and possible application areas.

The 2017 conference, also known as the Ninth International Conference on Knowledge Capture,¹ aimed at attracting researchers from diverse areas of Artificial Intelligence, including knowledge representation, knowledge acquisition, intelligent user interfaces, problem-solving and reasoning, planning, agents, text extraction, and machine learning, information enrichment and visualization, as well as researchers interested in cyber-infrastructures to foster the publication, retrieval, reuse, and integration of data. Today these data come from an increasingly heterogeneous set of resources that differ with regards to their domain, media format, quality, coverage, viewpoint, bias. More than the sheer amount of these data, their heterogeneity allows us to arrive at better models and answer complex questions that cannot be addressed in isolation but require the interaction of different scientific fields or perspectives. In most cases, knowledge is not captured as a means to an end but to, for instance, enable better user interfaces, improve retrieval beyond simple keyword search. For K-CAP 2017, we

¹<http://k-cap2017.org>

focused on the creation, enrichment, querying, and maintenance of knowledge graphs out of heterogeneous data sources.

The 2017 conference welcomed in total two workshops and three tutorials scheduled the day before the conference started. Workshops and tutorials opened the discussions: the workshops covered the crucial task of capturing knowledge from scientific content and by investigating the need to go beyond the traditional macro-reading processes for extracting knowledge from documents. In detail:

Second International Workshop on Capturing Scientific Knowledge²

From the early days of Artificial Intelligence, researchers have been interested in capturing scientific knowledge to develop intelligent systems. There are a variety of formalisms used today in different areas of science. Ontologies are widely used for organizing knowledge, particularly in biology and medicine. Process representations are used to do qualitative reasoning in areas such as physics and chemistry. Probabilistic graphical models are used by machine learning researchers, e.g., in climate modeling. In addition to enabling more advanced capabilities for intelligent systems in science, capturing scientific knowledge enables knowledge dissemination and open science practices. This is increasingly more important to enable the reuse of scientific knowledge across scientific disciplines, businesses and the public. Although great advances have been made, scientific knowledge is complex and poses great challenges for knowledge capture. This workshop provided a forum to discuss existing forms of scientific knowledge representation and existing systems that use them, and to envision major areas to augment and expand this important field of research. The increasing emphasis in open science has had a major focus on data sharing but it needs to encompass knowledge as well. There are many research challenges in open sharing and reuse of scientific knowledge that need to be addressed in future research. The workshop had as opening an invited talk by Suzanne Pierce and seven papers presented.

Machine Reading³ Machine reading holds significant potential for automating knowledge capture, especially given the continuing improvements in natural language processing technologies. Macro-reading techniques (skimming many documents) now enable collecting large databases of facts, while modern micro-reading techniques (comprehension of individual paragraphs) have proven effective at factoid question answering. In this workshop, participants will discuss ways to develop new capabilities in macro- and micro-reading to take these to the next level, in particular to extract useful representations of text (be they symbolic, neural, or a hybrid) that enable, for example, automated reasoning to answer non-trivial questions. This workshop provided a forum to researchers in discussing themes related to knowledge-based approaches applied to deep processing of con-

²<https://sciknow.github.io/sciknow2017>

³<http://www.cs.utexas.edu/users/porter/kcap-machinereadingworkshop.php>

tent. It also addressed the topic of assessing at large scale the quality of knowledge graphs. Five papers were presented at the workshop.

In addition to these workshops, three tutorials were included in the program. Also the tutorials attracted a lot of interest, they all shared the same format alternating depth analyses of topics with practical demonstrations. Three main topics were covered: representation learning, knowledge graphs, and deep learning. In detail:

Semantic data mining for knowledge acquisition⁴ The tutorial provided a synthetic, unifying view on semantic data mining and its application to knowledge acquisition. Semantic data mining is a data mining approach where domain ontologies are used as background knowledge. The challenge is to mine knowledge encoded in domain ontologies and knowledge graphs in addition to purely empirical data. The tutorial aimed to present major research challenges arising from peculiarities of semantic data mining such as proper consideration of the semantics of background knowledge, dealing with Open World Assumption, and semantic similarity measures. In addition, it covered also some of the recent advances in the area, namely semantic embeddings (embedding ontological background knowledge into neural networks).

DOing REusable MUSical data⁵ This tutorial firstly provided an in-depth explanations of the DOREMUS model (and its underlying foundations, CIDOC-CRM and FRBRoo) as well as the necessary controlled vocabularies. It then discussed and demonstrated the process to that lead to create a knowledge base of musical content starting from real data coming from musical libraries and be transformed to be compliant to Schema.org for various consumption scenarios. The entire DOREMUS tools chain were presented (e.g. tools for reconciling large multilingual knowledge graphs); the workshop covered also how the DOREMUS data can be consumed through various applications including an exploratory search engine and music recommender systems.

Hybrid techniques for knowledge-based NLP. Knowledge graphs meet machine learning and all their friends⁶ Many different artificial intelligence techniques can be used to explore and exploit large document corpora that are available inside organizations and on the Web. While natural language is symbolic in nature and first approaches were based on symbolic and rule-based methods (e.g., ontologies and knowledge bases), most widely used methods have been based on statistical approaches (e.g., linear methods such as support vectors machines, probabilistic topic models, and non-linear methods such as neural networks). These two approaches, knowledge-based and statistical methods, have their limitations

⁴http://www.cs.put.poznan.pl/alawrynowicz/wordpress/?page_id=662

⁵https://doremus-anr.github.io/kcap17_tutorial

⁶<http://expertsystemlab.com/kcap2017>

and strengths; there is an increasing trend that seeks to combine them to get the best of both worlds. This tutorial covered the foundations and modern practical applications of knowledge-based and statistical methods, techniques and models and their combination for exploiting large document corpora. This tutorial firstly focused on the foundations of many of the techniques that can be used for this purpose, including knowledge graphs, word embeddings, neural network methods, probabilistic topic models, and then demonstrated how a combination of these techniques is being used in practical applications and commercial projects where the instructors are currently involved.

These five co-located events attracted a large audience, who shared insights and fostered discussions with instructors and organizers. The overall take home message was in line with the conference scope, i.e. better understanding and framing the research of knowledge-based approaches to created autonomous and intelligent systems.