Computerm 2016

5th International Workshop on Computational Terminology

Proceedings of the Workshop

December 12, 2016
Osaka, Japan

Copyright of each paper stays with the respective authors (or their employers).

Introduction

Computational Terminology covers an increasingly important aspect in Natural Language Processing areas such as text mining, information retrieval, information extraction, summarisation, textual entailment, document management systems, question-answering systems, ontology building, etc. Terminological information is paramount for knowledge mining from texts for scientific discovery and competitive intelligence. Scientific needs in fast growing domains (such as biomedicine, chemistry and ecology) and the overwhelming amount of textual data published daily demand that terminology is acquired and managed systematically and automatically; while in well established domains (such as law, economy, banking and music) the demand is on fine-grained analyses of documents for knowledge description and acquisition. Moreover, capturing new concepts leads to the acquisition and management of new knowledge.

The aim of this fifth Computerm workshop is to bring together Natural Language Processing to discuss recent advances in computational terminology and its impact in many NLP applications. The topics addressed in this workshop are wide ranging:

- term extraction, recognition and filtering, which is the core of the terminological activity that lays basis for other terminological topics and tasks;

- event recognition and extraction, that extends the notion of the terminological entity from terms meaning static units up to terms meaning procedural and dynamic processes;

- acquisition of semantic relations among terms, which is also an important research topic as the acquisition of semantic relationships between terms finds applications such as the population and update of existing knowledge bases, definition of domain specific templates in information extraction and disambiguation of terms;

- term variation management, that helps to deal with the dynamic nature of terms, their acquisition from heterogeneous sources, their integration, standardisation and representation for a large range of applications and resources, is also increasingly important, as one has to address this research problem when working with various controlled vocabularies, thesauri, ontologies and textual data. Term variation is also related to their paraphrases and reformulations, due to historical, regional, local or personal issues. Besides, the discovery of synonym terms or term clusters is equally beneficial to many NLP applications;

- definition acquisition, that covers important research and aims to provide precise and non-ambiguous description of terminological entities. Such definitions may contain elements necessary for the formal description of terms and concepts within ontologies;

- consideration of the user expertise, that is becoming a new issue in the terminological activity, takes into account the fact that specialized domains contain notions and terms often non-understandable to non-experts or to laymen (such as patients within the medical area, or bank clients within banking and economy areas). This aspect, although related to specialized areas, provides direct link between specialized languages and general language;

- systematic terminology management and updating domain specific dictionaries and thesauri, that are important aspects for maintaining the existing terminological resources. These aspects become crucial because the amount of the existing terminological resources is constantly increasing and because their perennial and efficient use depends on their maintenance and updating, while their re-acquisition is costly and often non-reproducible;

- monolingual and multilingual resources, that open the possibility for developing cross-lingual and multi-lingual applications, requires specific corpora, methods and tools which design and evaluation are challenging issues;

- robustness and portability of methods, which allows to apply methods developed in one given context to other contexts (corpora, domains, languages, etc.) and to share the research expertise among them;

- social netwoks and modern media processing, that attracts an increasing number of researchers and that provides challenging material to be processed;

- utilization of terminologies in various NLP applications, as they are a necessary component of any NLP system dealing with domain-specific literature, is another novel and challenging research direction.

In the call for paper, we encouraged authors to submit their research work related to various aspects of computational terminology related approaches, ranging from term extraction in various languages (using verb co-occurrence, information theoretic approaches, machine learning, etc.), translation pairs extracting from bilingual corpora based on terminology, up to semantic oriented approaches and theoretical aspects of terminology.

Besides, experiments on the evaluation of terminological methods and tools are also encouraged since they provide interesting and useful proof about the utility of terminological resources:

- direct evaluation may concern the efficiency of the terminological methods and tools to capture the terminological entities and relations, as well as various kinds of related information;

- indirect evaluation may concern the use of terminological resources in various NLP applications and the impact these resources have on the performance of the automatic systems. In this case, research and competition tracks (such as TREC, BioCreative, CLEF, CLEF-eHealth, I2B2, *SEM, and other shared tasks), provide particularly fruitful evaluation contexts and proved very successful in identifying key problems in terminology such as term variation and ambiguity.

This workshop is a continuation of previous Computerm workshops. The last Computerm was joined to the previous COLING conference in 2014.

The Computerm 2016 workshop received 28 submissions from 12 countries and 4 continents addressing issues on 9 languages. Further to a double-blind peer-reviewing process, 5 papers were accepted as long oral presentations, 1 as short oral presentation and 8 as posters. The acceptance rate for oral presentations is 21% and the overall acceptance rate is 50%. The selected papers tackle various terminology related problems such terminological resource evaluation, semantic relation acquisition and, above all, extraction, recognition and filtering of terms, with classical ruled-based methods or statistical approaches as well as word embeddings. The domain of application of the proposed approaches is varied, going from Mathematics and Art to Environment. We believe this workshop will be a great place for fruitful research discussions, and the emergence of new research topics and collaborations. The objective of the combined oral and poster presentations is to strengthen this point.

Acknowledgements

Organisers

Patrick Drouin, Observatoire de linguistique Sens-Texte, Université de Montréal, Canada
Natalia Grabar, CNRS UMR 8163 STL, France
Thierry Hamon, LIMSI-CNRS & Université Paris 13, France
Kyo Kageura, Library and Information Science Laboratory, University of Tokyo, Japan
Koichi Takeuchi, Graduate School of Natural Science and Technology, Okayama University, Japan

Programme Committee

Lynne Bowker, University of Ottawa, Canada
Béatrice Daille, University of Nantes, France
Louise Deléger, INRA, France
Gregory Grefenstette, INRIA, University Paris Sud, France
Yoshihiko Hayashi, Waseda University, Japan
Olga Kanishcheva, Kharkiv Polytechnic Institute, Ukraine
Georgios Kontonatsios, NaCTeM, University of Manchester, UK
Marie-Claude L'Homme, University of Montréal, Canada
Philippe Langlais, RALI, Canada
Veronique Malaise, Elsevier BV, the Netherlands
Elizabeth Marshman, University of Ottawa, Canada
Fleur Mougin, University Bordeaux, France
Agnieszka Mykowiecka, IPIPAN, Poland
Rogelio Nazar, University Pompeu Fabra, Spain
Goran Nenadic, University of Manchester, UK
Fabio Rinaldi, University of Zurich, Switzerland
Selja Seppälä, University of Florida, USA
Karine Verspoor, University of Melbourne, Australia
Jorge Vivaldi Palatresi, University Pompeu Fabra, Spain
Pierre Zweigenbaum, LIMSI, France

Invited Speaker

Min Song, Text and Social Media Mining Lab, Yonsei University, Seoul, Republic of Korea

Table of Contents

Workshop Program

Monday, December 12, 2016

9:00–9:15   **Opening Remarks**

9:15–10:45   **Invited Speaker: Professor Min Song**

9:15–10:45   *Analyzing Impact, Trend, and Diffusion of Knowledge associated with Neoplasms Research*
Min Song

10:45–11:10   *Coffee Break*

11:10–12:00   **Session 1**

11:10–11:35   *Local-Global Vectors to Improve Unigram Terminology Extraction*
Ehsan Amjadian, Diana Inkpen, Tahereh Paribakht and Farahnaz Faez

11:35–12:00   *Recognition of non-domain phrases in automatically extracted lists of terms*
Agnieszka Mykowiecka, Malgorzata Marciniak and Piotr Rychlik

12:00–14:00   *Lunch Break*

Monday, December 12, 2016 (continued)

14:00–15:15   **Session 2**

14:00–14:25   *Contextual term equivalent search using domain-driven disambiguation*
Caroline Barriere, Pierre André Ménard and Daphnée Azoulay

14:25–14:50   *A Method of Augmenting Bilingual Terminology by Taking Advantage of the Conceptual Systematicity of Terminologies*
Miki Iwai, Koichi Takeuchi, Kyo Kageura and Kazuya Ishibashi

14:50–15:15   *Acquisition of semantic relations between terms: how far can we get with standard NLP tools?*
Ina Roesiger, Julia Bettinger, Johannes Schäfer, Michael Dorna and Ulrich Heid

15:15–15:30   **Boosters**

15:30–15:50   *Coffee Break*

15:50–17:00   **Poster Session**

Evaluation of distributional semantic models: a holistic approach
Gabriel Bernier-Colborne and Patrick Drouin

A Study on the Interplay Between the Corpus Size and Parameters of a Distributional Model for Term Classification
Behrang QasemiZadeh

Pattern-based Word Sketches for the Extraction of Semantic Relations
Pilar León-Araúz, Antonio San Martín and Pamela Faber

Constructing and Evaluating Controlled Bilingual Terminologies
Rei Miyata and Kyo Kageura

Providing and Analyzing NLP Terms for our Community
Gil Francopoulo, Joseph Mariani, Patrick Paroubek and Frédéric Vernier

Evaluating a dictionary of human phenotype terms focusing on rare diseases
Simon Kocbek, Toyofumi Fujiwara, Jin-Dong Kim, Toshihisa Takagi and Tudor Groza

Monday, December 12, 2016 (continued)

17:00   **Closing Remarks**