

LREC 2016

Language Resource and Evaluation Conference

May 23-28, 2016

Portorož, Slovenia

Patrick Holthaus

patrick.holthaus@uni-bielefeld.de

the date of receipt and acceptance should be inserted later

The 10th edition of the bi-annual Language Resources and Evaluation Conference (LREC) was held from May 23rd to 28th 2016 under patronage of the president of Slovenia, Borut Pahor, chaired by Nicoletta Calzolari. The beautiful resort Portorož is located at the Adriatic Sea, in the country's coastal area between Croatia and Italy. A welcome reception close to the conference location at an enjoyable square with a ruined church introduced the around 1200 participants to the nice scenery. At latest during the gala dinner, which took place at sunset in the open air directly at the seaside, the pleasant ambience of the scenery became apparent for everyone.

In total, 744 papers have been accepted for presentation at LREC 2016, distributed over 47 oral and 65 poster sessions, of which four to five were each held in parallel. The acceptance rate was at 59 percent, which is slightly lower than earlier iterations of the conference. The proceedings are provided on-line by the European Language Resources Association (ELRA) as open access under a creative commons license (CC-BY-NC).

Along with the presentations of peer-reviewed submissions, the program also consisted of two keynotes, a panel discussion, one invited local talk, and the Antonio Zampolli prize talk. In addition to the main conference program of three days, 36 satellite workshops and tutorials happened at two days beforehand and one day afterwards.

As the conference focus lies on language resources and their evaluation, most of the submitted research presented the acquisition of a (mostly textual) corpus or methods of corpus processing such as text mining or tagging. Authors were encouraged to also describe any presented corpus in the LRE map2, a tool to mon-

itor, share, discuss, and analyze language resources for natural language processing. Furthermore, efforts were taken to formalize the citation of language resources by pushing the usage of a new unique identifier, the "International Standard Language Resource Number" (ISLRN) as a reference, for example as a BibTeX standard.

Besides the provision of corpora, major topics of the conference included machine translation and evaluation, language feature recognition (e.g., emotions, named entities, OCR), and formal aspects such as language resource policies, infrastructures, architectures, and large projects. Major cross-cutting topics of LREC 2016 were the application of crowd sourcing to support the creation of massive amounts of manual annotation, and social media mining for corpus generation - especially Twitter. Also (deep) learning played a substantial role in many presentations, for example in the context of document classification or prediction. Although addressed in one of two keynotes and in the prize talk as well, spoken dialog and multimodal interaction were represented less than written textual resources. One session, for example, covered corpora featuring conversational agents, while there were also sessions on multimodality, speech processing as well as synthesis.

In the first keynote, Roberto Pieraccini gave a historical insight on the evolution of dialog systems starting from the simple speech recognition, which was limited to understand the numbers one to ten only. He then described the rising complexity with the example of automatic help desks and later described autonomous interactive agents such as robots as the most complex dialog systems today. The second keynote from Ryan McDonald focused on the application of language resource techniques at Google research. Alongside the example of their language translation tool, he highlighted

advantages of machine learning in contrast to manual annotation in certain applications, i.e., if enough data is available. In cases of under-resourced problems for example, annotation labor instead might be very useful as a starting point for structuring and analyzing a data set.

While the panel discussion was centered on industrial applications of language resources, especially regarding the topic of the “Single European Market”, the Antonio Zampolli prize laureate Roger K. Moore again elaborated on the generation of social agents with interactive capabilities. He ordered different applications by their degree of in real-time utilization and how interactive they are. Obviously, intelligent agents have to be placed highly in both dimensions. After stating the most urgent technical requirements for such systems, he focused on usability issues and ethical questions that have to be addressed in order to compile robots and other intelligent agents in a socially acceptable way.