1st Conference on Language, Data and Knowledge – Galway, 2017

On June 19-20, the first conference on Language, Data and Knowledge (LDK 2017) took place in Galway, Ireland, which was attended by over 100 participants from 27 countries. This new biennial conference series brings together researchers from across disciplines concerned with the acquisition, curation and use of language data in the context of data science and knowledge-based applications. Language datasets, such as corpora, typological resources and, of course, dictionaries, are of increasing importance to machine learning-based approaches in Natural Language Processing (NLP), linked data and Semantic Web research and applications that depend on linguistic and semantic annotation with lexical, terminological and ontological resources, manual alignment across languages or other human-assigned labels. The acquisition, provenance, representation, maintenance, usability, quality as well as legal, organizational and infrastructure aspects of language data are therefore rapidly becoming major areas of research that were at the focus of the conference. A further focus was the combined use and exploitation of language data and knowledge graphs in data science-based approaches to use cases in industry, including biomedical applications, as well as use cases in humanities and social sciences.

The LDK conference has been initiated by a consortium of researchers from the Insight Centre for Data Analytics (National University of Ireland Galway), InfAI (Leipzig University, Germany) and Goethe University Frankfurt, Germany, and a Scientific Committee of leading researchers in NLP, linked data and the Semantic Web, language resources and digital humanities. LDK 2017 was endorsed by several international organisations, namely DBpedia, the ACL Special Interest Group for Annotation (SIGANN), Global WordNet Association, CLARIN, Big Data Value Association (BDVA) and DARIAH Ireland.

The conference stands at the intersection of data science and NLP and brought together researchers from across computer science as well as linguistics and the humanities. The conference series is concerned with the creation of data resources as well as the metadata, development, evaluation, quality and legal aspects of publishing such data as well as knowledge graphs as a key focus for providing knowledge about the world in terms of ontologies, terminologies and linked data. Direct applications in NLP are of particular importance in certain semantic technologies, question answering, multilinguality and big data. Finally, applications of these technologies to domains as varied as digital humanities, enterprise data analytics and text mining of biomedical literature are of importance.

The conference featured invited talks from both industry and academia. Zoltan Szlavik from IBM Benelux gave a talk about cognitive computing and the Watson systems. Kathleen McKeown from the Data Science Institute of the Columbia University talked about the intersection of data science and NLP. Antal van den Bosch of Radboud University in the Netherlands and director of the Meertens Institute related data science technologies with new breakthroughs in digital humanities. Finally, Isaac Graham from NUI Galway talked about relevant features of the Irish and Welsh languages.

In addition to the main conference there were a number of workshops and tutorials. The 1st OntoLex Workshop discussed the development of the OntoLex model, a recent vocabulary from the World Wide Web Consortium, for representing dictionaries relative to ontologies as linked data on the Web. The Translation Inference Across Dictionaries (TIAD) shared task explored the automatic generation of bilingual dictionaries from existing resources and was co-organized by K Dictionaries and the Polytechnic University of Madrid. A further workshop was organized on Challenges for Wordnets, concerning the construction of dictionaries in the wordnet format, organized by the Global WordNet Association. Importantly, also a tutorial on text analytics for Digital Humanities and Social Sciences with CLARIN was organized jointly by CLARIN and DARIAH Ireland. Finally, a meeting of the DBpedia Association took place after the main conference.

The next edition of the LDK conference is planned to take place in Leipzig, Germany in 2019. More details about the conference can be found at http://ldk2017.org.

John P. McCrae, Paul Buitelaar, Christian Chiarcos and Sebastian Hellmann