© 2014 The Authors
The papers in this volume are licensed by the authors under a Creative Commons Attribution 4.0 International License.

ISBN 978-1-873769-31-7
Introduction

Welcome to the First AHA!-Workshop on Information Discovery in Text!

In this workshop, we are bringing together leading researchers in the emerging field of Information Discovery to discuss approaches for Information Extraction that are not bound by a pre-specified schema of information, but rather discover relational or categorial structure automatically from given unstructured data.

This includes approaches that are based on unsupervised machine-learning over models of distributional semantics, as well as OpenIE methods that relax the definition of semantic relations in order to more openly extract structured information. Other approaches focus on inexpensively training information extractors to be used across different domains with minimal supervision, or on adapting existing IE systems to new domains and relations.

As different approaches on Information Discovery are gaining momentum, many fundamental questions arise that merit discussion: How do these approaches compare and what are their relative strengths and weaknesses? What are the desiderata and applications for Information Discovery methods? How can such methods be evaluated and compared? And most importantly, what is the potential of Information Discovery methods and where can current research lead?

We received 19 paper submissions of which the programme committee has accepted ten - six of which were chosen for oral presentation and four as posters.

We look forward to a workshop full of interesting paper presentations, invited talks and lively discussion.

Sincerely,

Alan Akbik and Larysa Visengeriyeva

AHA! Chairs
Workshop Chairs and Organizers
   Alan Akbik, Technische Universität Berlin
   Larysa Visengeriyeva, Technische Universität Berlin

Programme Committee
   Isabelle Augenstein, University of Sheffield
   Christoph Boden, Technische Universität Berlin
   Danushka Bollegala, University of Liverpool
   Leon Derczynski, University of Sheffield
   Jonathan Gordon, University of Rochester
   Max Heimel, Technische Universität Berlin
   Holmer Hemsen, Technische Universität Berlin
   Johannes Kirschnick, Technische Universität Berlin
   Xiao Ling, University of Washington
   Steffen Remus, Technische Universität Darmstadt
   Martin Riedl, Technische Universität Darmstadt
   Tim Rocktäschel, University College London
   Vivek Srikumar, Stanford University
   Mark Steedman, University of Edinburgh
   Andreas Vlachos, University College London
   Gerhard Weikum, Max Plank Institute Saarbrücken
   Limin Yao, University of Massachusetts

Steering Committee
   Hans Uszkoreit, German Center for Artificial Intelligence
   Volker Markl, Technische Universität Berlin
# Table of Contents

*Application-Driven Relation Extraction with Limited Distant Supervision*
Andreas Vlachos and Stephen Clark ................................................................. 1

*A Comparative Study of Conversion Aided Methods for WordNet Sentence Textual Similarity*
Muhidin Mohamed and Mourad Oussalah ......................................................... 7

*Word Clustering Based on Un-LP Algorithm*
Jiguang Liang, Xiaofei Zhou, Yue Hu, Li Guo and Shuo Bai .......................... 13

*Mining temporal footprints from Wikipedia*
Michele Filannino and Goran Nenadic ............................................................ 19

*Automatic Detection and Analysis of Impressive Japanese Sentences Using Supervised Machine Learning*
Daiki Hazure, Masaki Murata and Masato Tokuhisa ....................................... 26

*Proposition Knowledge Graphs*
Gabriel Stanovsky, Omer Levy and Ido Dagan .................................................. 32

*Using Distributional Semantics to Trace Influence and Imitation in Romantic Orientalist Poetry*
Nitish Aggarwal, Justin Tonra and Paul Buitelaar .......................................... 38

*Unsupervised Approach to Extracting Problem Phrases from User Reviews of Products*
Elena Tutubalina and Vladimir Ivanov ............................................................ 43

*Extracting a Repository of Events and Event References from News Clusters*
Silvia Julinda, Christoph Boden and Alan Akbik ............................................ 49

*Towards Social Event Detection and Contextualisation for Journalists*
Prashant Khare and Bahareh Heravi ................................................................. 54
Conference Program

Application-Driven Relation Extraction with Limited Distant Supervision
Andreas Vlachos and Stephen Clark

A Comparative Study of Conversion Aided Methods for WordNet Sentence Textual Similarity
Muhidin Mohamed and Mourad Oussalah

Word Clustering Based on Un-LP Algorithm
Jiguang Liang, Xiaofei Zhou, Yue Hu, Li Guo and Shuo Bai

Mining temporal footprints from Wikipedia
Michele Filannino and Goran Nenadic

Automatic Detection and Analysis of Impressive Japanese Sentences Using Supervised Machine Learning
Daiki Hazure, Masaki Murata and Masato Tokuhisa

Proposition Knowledge Graphs
Gabriel Stanovsky, Omer Levy and Ido Dagan

Using Distributional Semantics to Trace Influence and Imitation in Romantic Orientalist Poetry
Nitish Aggarwal, Justin Tonra and Paul Buitelaar

Unsupervised Approach to Extracting Problem Phrases from User Reviews of Products
Elena Tutubalina and Vladimir Ivanov

Extracting a Repository of Events and Event References from News Clusters
Silvia Julinda, Christoph Boden and Alan Akbik

Towards Social Event Detection and Contextualisation for Journalists
Prashant Khare and Bahareh Heravi