ACL-IJCNLP 2015

Proceedings of the First Workshop on Computing News Storylines (CNewsStory 2015)

July 31, 2015
Beijing, China

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-941643-71-6

Introduction

This volume contains the proceedings of the 1st Workshop on Computing News Storylines (CNewsStory 2015) held in conjunction with the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015) at the China National Convention Center in Beijing, on July 31st 2015.

Narratives are at the heart of information sharing. Ever since people began to share their experiences, they have connected them to form narratives. The study of storytelling and the field of literary theory called narratology have developed complex frameworks and models related to various aspects of narrative such as plots structures, narrative embeddings, characters' perspectives, reader response, point of view, narrative voice, narrative goals, and many others. These notions from narratology have been applied mainly in Artificial Intelligence and to model formal semantic approaches to narratives (e.g. Plot Units developed by Lehnert (1981)). In recent years, computational narratology has qualified as an autonomous field of study and research. Narrative has been the focus of a number of workshops and conferences (AAAI Symposia, Interactive Storytelling Conference (ICIDS), Computational Models of Narrative). Furthermore, reference annotation schemes for narratives have been proposed (NarrativeML by Mani (2013)).

The majority of the previous work on narratives and narrative structures have mainly focused on the analysis of fictitious texts. However, modern day news reports still reflect this narrative structure, but they have proven difficult for automatic tools to summarise, structure, or connect to other reports. This difficulty is partly rooted in the fact that most text processing tools focus on extracting relatively simple structures from the local lexical environment, and concentrate on the document as a unit or on even smaller units such as sentences or phrases, rather than cross-document connections. However, current information needs demand a move towards multidimensional and distributed representations which take into account the connections between all relevant elements involved in a "story". Additionally, most work on cross-document temporal processing focuses on linear timelines, i.e. representations of chronologically ordered events in time (for instance, the Event Narrative Event Chains by Chambers (2011), or the SemEval 2015 Task 4: Cross Document TimeLines by Minard et al. (2014)). Storylines, though, are more complex, and must take into account temporal, causal and subjective dimensions. How storylines should be represented and annotated, how they can be extracted automatically, and how they can be evaluated are open research questions in the NLP and AI communities.

The workshop aimed to bring together researchers from different communities working on representing and extracting narrative structures in news, a text genre which is highly used in NLP but which has received little attention with respect to narrative structure, representation and analysis. Currently, advances in NLP technology have made it feasible to look beyond scenario-driven, atomic extraction of events from single documents and work towards extracting story structures from multiple documents, while these documents are published over time as news streams. Policy makers, NGOs, information specialists (such as journalists and librarians) and others are increasingly in need of tools that support them in finding salient stories in large amounts of information to more effectively implement policies, monitor actions of "big players" in the society and check facts. Their tasks often revolve around reconstructing cases either with respect to specific entities (e.g. person or organizations) or events (e.g. hurricane Katrina). Storylines represent explanatory schemas that enable us to make better selections of relevant information but also projections to the future. They form a valuable potential for exploiting news data in an innovative way.

Albeit small in number, the contributions that are published in this volume do indeed cover the topics we intended to touch upon. We received 12 submissions and accepted 9. Two papers focus on tracking and representing emergent news topics (Tadashi) and develop personalised news aggregation systems (Fedorovsky et al.). Events, the primary source of information and blocks for storylines, are the targets of three papers which tackles different issues such as improving event type detection (Li et al.), the analysis of the properties of sequences of events (Simonson and Davis), and the automatic extraction of

news agendas as the ability of storylines to direct action (Stalpouskaya and Baden). Notions such as relevance and importance are at the core of two papers: one paper which describes a formal model and a preliminary implementation for automatically extracting storylines from news stream (Vossen et al.), and one paper which proposes a post-retrieval snippet clustering based on pattern structures (Makhalova et al.). Finally, a proposal for storyline representation and evaluation (Laparra et al.) and the adaptation of approaches and methods from the domain of fiction to the news data (Miller et al.) are reported.

We would like to thank the members of the Program Committee for their timely reviews. We would also like to thank the authors for their contributions.

Organizers:

Tommaso Caselli, VU University Amsterdam
Marieke van Erp, VU University Amsterdam
Anne-Lyse Minard, Fondazione Bruno Kessler
Mark Finlayson, Florida International University
Ben Miller, Georgia State University
Jordi Atserias, Yahoo! Barcelona
Alexandra Balahur, European Commission's Joint Research Centre
Piek Vossen, VU University Amsterdam

Program Committee:

Sabine Bergler, Concordia University, Canada
Matje van de Camp, De Taalmonsters, The Netherlands
Erik van der Goot, EC Joint Research Centre, Italy
Nathanael Chambers, United States Naval Academy, USA
Leon Derczynski, University of Sheffield, UK
Anthony Jameson, DFKI, Germany
Bernardo Magnini, Fondazione Bruno Kessler, Italy
Jarred McGinnis, Logomachy.org, UK
Roser Morante, VU University Amsterdam, The Netherlands
Silvia Pareti, Google Inc. & University of Edinburgh
Ellen Riloff, University of Utah, USA
Roser Saurí, Pompeu Fabra University, Spain
Hristo Tanev, EC Joint Research Centre, Italy
Xavier Tannier, Université Paris-Sud & LIMSI-CNRS, France
Naushad UzZanam, Nuance Communications, Inc.
Laure Vieu, IRIT-CNRS - Université Toulouse III, France
Marc Verhagen, Brandeis University, USA

Table of Contents

Conference Program

31 July 2015

09:15–09:25 ***Opening Remarks***

09:25–09:50 *Interactions between Narrative Schemas and Document Categories*
Dan Simonson and Anthony Davis

09:50–10:10 *Improving Event Detection with Abstract Meaning Representation*
Xiang Li, Thien Huu Nguyen, Kai Cao and Ralph Grishman

10:10–10:30 *News clustering approach based on discourse text structure*
Tatyana Makhalova, Dmitry Ilvovsky and Boris Galitsky

10:30–11:00 ***Coffee Break***

11:00–11:25 *To Do or Not to Do: the Role of Agendas for Action in Analyzing News Coverage of Violent Conflict*
Katsiaryna Stalpouskaya and Christian Baden

11:25–11:45 *MediaMeter: A Global Monitor for Online News Coverage*
Tadashi Nomoto

11:45–12:05 *Expanding the horizons: adding a new language to the news personalization system*
Andrey Fedorovsky, Maxim Ionov, Varvara Litvinova, Tatyana Olenina and Darya Trofimova

12:05–12:30 *Storylines for structuring massive streams of news*
Piek Vossen, Tommaso Caselli and Yiota Kontzopoulou

12:30–14:00 ***Lunch Break***

14:00–14.20 *From TimeLines to StoryLines: A preliminary proposal for evaluating narratives*
Egoitz Laparra, Itziar Aldabe and German Rigau

14:20–14:40 *Cross-Document Non-Fiction Narrative Alignment*
Ben Miller, Jennifer Olive, Shakthidhar Gopavaram and Ayush Shrestha

14:40–15:30 ***Discussion and Conclusions***