SKG4J 2020: 1st International Workshop on Semantic and Knowledge Graph Advances for Journalism

Tareq Al-Moslmi
tareq.al-moslmi@uib.no
University of Bergen
Bergen, Norway

Raphaël Troncy
raphael.troncy@eurecom.fr
Eurecom
Sophia Antipolis, France

André Freitas
andre.freitas@manchester.ac.uk
The University of Manchester
Manchester, United Kingdom

Davide Ceolin
davide.ceolin@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands

Abdullatif Abolohom
abdullatif.abolohom@ubi.pt
University of Beira Interior
Covilhã, Portugal

ABSTRACT

SKG4J targeted contributions at the interface between Artificial Intelligence, Data Management and its implications for journalistic practice. The first version of the workshop accepted four submissions with topics emphasising the complementary requirements for delivering realistic journalistic knowledge extraction/management platforms.

1 INTRODUCTION

Journalism is undergoing a major disruption for the past ten years. The increasing importance of the Web in shaping public discourse is forcing journalistic institutions and professionals to a profound reflection on their role and practice. Despite the institutional crisis which emerged on its wake, the Web presents new challenges and opportunities which are tailored towards journalists: fact checking (and fake news), sense making at scale, interpreting large digital archives, data-driven story-telling are some of the functions which are redefining the contemporary newsroom.

These new challenges require the close coordination between journalists and emerging algorithmic and data infrastructures. Recent advances in the Artificial Intelligence (AI) space are building the foundations to support data interpretation at scale, augmenting the ability of journalists to read, verify, integrate and write stories.

Knowledge graphs, text embeddings, and neural nets based classifiers are emerging as universal infrastructure components which enable the interpretation of meaning at scale. The *1st International Workshop on Semantic and Knowledge Graph Advances for Journalism* (SKG4J[1]) was organised with the aim of serving as a forum to discuss the emerging principles and applications of this new medium in the context of journalism. While news and media organisations have been early adopters of Knowledge Graphs and AI

[1] https://almoslmi.github.io/SemanticJournalism/

methods, the potential of these technologies are not fully demonstrated.

2 SUMMARY OF THE CONTRIBUTIONS

The workshop accepted four contributions covering the topics of open information extraction and semantic representation, data privacy, technical challenges and opportunities:

- Mahouachi and Suchanek - Extracting Complex Information from Natural Language Text: A Survey
- Al-Moslmi et al. - Lifting News into a Journalistic Knowledge Platform
- Gallofre and Opdahl - Digitalisation of Newsrooms: Challenges and Opportunities for Journalistic Knowledge Platforms
- Gallofre et al. - Data Privacy in Journalistic Knowledge Platforms

In *Extracting Complex Information from Natural Language Text: A Survey* Mahouachi & Suchanek provide a critical survey on open information extraction, which currently emphasise the extraction of binary relations. The authors structure their critique by analysing a set of OpenIE platforms against semantic representation properties such as anaphora resolution, inter-sentence relations, ability to capture n-ary relations, and the representation of belief, negation, causality, anteriority and contrast.

The extraction of knowledge graphs from news corpora using more canonical information extraction and linking components is described in Lifting News into a Journalistic Knowledge Platform. Al-Molsmi et al. formalise a generic NLP lifting architecture for the extraction of journalistic knowledge graphs, which can be later integrated to pipelines demanding more specialised requirements.

While there is clarity that language technologies will become an integral part of future newsrooms, there is a gap in understanding the strengths and weaknesses of existing pipelines and the associated challenges and opportunities. In *Digitalisation of Newsrooms: Challenges and Opportunities for Journalistic Knowledge Platforms*, Ocana & Opdahl provide a systematic review of existing Journalistic Knowledge Platforms (JKP), comparing and contrasting the architectural features of these platforms with the help of six reference categories.

When operating over real world data, journalistic knowledge platforms need to take into account cross-cutting concerns sch as

data privacy. In *Data Privacy in Journalistic Knowledge Platforms* Ocana et al. describe a framework for classifying personal data in journalistic knowledge graphs, identifying scenarios and sources of personal data which can potentially can conflict with data privacy.

3 CONCLUSION

The contributions in SKG4J 2020 emphasised a broader spectrum of technical aspects required to deliver knowledge-based / NLP infrastructures on real-world journalistic settings (from complex knowledge extraction to data privacy). While there are components which are mature and fully integrated in the newsroom (such as Topic Models, Named Entity Recognition and Linking, Semantic Search), the role and value of emerging and more sophisticated explicit and latent NLP/AI models still needs to be established. Future versions of this workshop will promote the discussion of these emerging components and their ability to deliver value in the newsroom.