Organizing Data, Information, and Knowledge in Big Data Environments

Jiangping Chen
Department of Information Science
University of North Texas
Denton, Texas, USA
Jiangping.chen@unt.edu

Wei Lu
School of Information Management
Wuhan University
Wuhan, China
WeiLu@whu.edu.cn

Oksana Zavalina
Department of Information Science
University of North Texas
Denton, Texas, USA
Oksana.Zavalina@unt.edu

ABSTRACT

The explosion of information and the availability of big data provide new significant challenges for digital libraries. For example, libraries face the "cyberinfrastructural challenge" and the need to develop better understanding of research data to support curation, sharing and reuse of data generated by data-intensive science [9][6]. The synergy between information science and data science is emerging to address these Big Data environment challenges. The most fruitful collaboration areas for this synergy include those related to information organization: "big metadata, smart metadata" the ways to leverage the "metadata capital"[4]. The proposed one-day workshop will seek to support the interdisciplinary collaboration in this important area. It will focus on the challenges and opportunities provided by Big Data environment for information and computing professionals to explore innovative strategies and solutions for organizing data, information, and knowledge.

CCS CONCEPTS

• **Information systems** → Information integration; • **Computer systems organization** → Embedded systems; Redundancy; Robotics; • **Networks** → Network reliability; • **General and reference** → Evaluation.

KEYWORDS

Information organization, knowledge organization, data management, big data analytics

1 INTRODUCTION

The objective of this workshop is to allow researchers and professionals working in different areas related to organizing data, information, and knowledge to get together and interact. Together we will explore innovation solutions to access data, information, and knowledge. Also, we hope this workshop will foster collaborations among information professionals and computing professionals.

2 PROPOSED WORKSHOP FORMAT

We plan a full day workshop with the following components:

- Short introduction by the organizers.
- Presentations of papers submitted and accepted in response to an open call for research and position papers. Papers will be presented in the short oral format. The number of oral presentations will be limited to at most 12 and allow for the maximum of time for interactive activities.
- Working groups focusing on relevant topics on data, information, and knowledge organization. Solicited topics are listed in Section 3.

3 POSSIBLE TOPICS

Semantic role labeling of Scholarly big data. Semantic role labeling (SRL) is a natural language processing (NLP) task that aims to find all arguments for a given predicate in a sentence and label them with semantic roles, which is important for Information Retrieval (IR), Information Extraction (IE), Question Answering (QA), and Machine Translation (MT).Currently, most work in SRL has focused on documents or literature from the newswire domain or medical domain[1]. With the increasing availability of digital data on scholarly activities, semantic analysis has become very important in order to make the machine understand the concepts behind academic language and mine valuable knowledge. However, very few SRL studies have been conducted in scholarly big data, and there is a lack of large-scale annotated corpora of scholarly data. The application of SRL in scholarly big data will not only promote the development of SRL technologies, but also provide broad prospects for academic text mining.

New organization framework for digital libraries. The development of information technology has brought out digital libraries and continuously affects the development and changes of them. The organization framework is the basis of the construction of digital libraries, which helps avoid divergence of development efforts and problems with interoperability. However, traditional frameworks cannot satisfy the new demands of future digital libraries, such as the processing and use of audio and video collections, the application of artificial intelligence in digital libraries, the integration of digital libraries in different regions, intelligent service, etc. In this context, new organization frameworks with innovation and disruption will become a hotspot and focus of future digital library research. For example, an organization framework

IEEE computer society

of personalized information service system of digital library based on artificial intelligence technology.

Knowledge graph quality evaluation and applications. Recently, knowledge graph has been used in different domains such as medical[3], academic[8], legal[2], etc. However, most of the existing studies did not examine the quality of knowledge graph. While high quality knowledge graph helps produce valuable discoveries, poor quality knowledge graph can lead to erroneous conclusions. The quality of knowledge graph can be evaluated from intrinsic and extrinsic perspectives[5]. Participants can propose effective quality evaluation frameworks for knowledge graph or develop innovative methods to evaluate knowledge graph from different perspectives. Also, we welcome papers to apply existing knowledge graphs such as DBPedia and Freebase to information retrieval, question answering and other tasks.

Data and information organization in different domains and for different purposes. Big data makes brings challenges for humans to organize information. Even for non-digital materials, many of them need to be organized in digital environment. For example, digital humanities emerged in the cultural heritage domain; more and more researchers are trying to build different models to organize information in this field[7]. Submissions can include discussions of theoretical developments to support organization of data, information, and knowledge in the Big Data environment, or comparison of different datasets on ways they are organized. We are also interested in the following types of submissions under this category:

- literature reviews, results of bibliometric analysis, or large-scale comparative analysis of methods and technologies (e.g., algorithms, Semantic Web tools, etc.) for information organization, as well as for organizing data and knowledge, in Big Data environment;
- discussions of large-scale automatic metadata extraction and data or information modeling approaches: domain-specific or domain-independent;
- discussions of employing Big Data analytics approaches in metadata quality evaluation and quality control, digital library use analysis, information architecture and user experience evaluation.

4 AUDIENCE

The expected audience of this workshop includes researchers in the fields of library and information science, data science, and computer science who work in the areas of information organization, information retrieval, knowledge management, data mining, and Big Data analytics. Workshop organizers also expect to attract participation from academic and corporate information professionals. The expected number of attendees will be 30.

A call for papers will be distributed immediately after the workshop proposal is accepted to faculty members who teach and/or conduct research in the areas listed above worldwide, as well as to authors who have published on the related topics in the recent three years. We will also use social media to announce the workshop to attract wider participation. The potential participants will be identified from the Internet and Web of Science database. This full-day workshop will be held on June 6th, 2019.

A double-blind review process will be applied to evaluate all submissions. Accepted papers will be published in the Electronic Library (https://www.emeraldinsight.com/journal/el), a peer-reviewed SSCI journal on Information Organization after the workshop.

5 ABOUT THE ORGANIZERS

Jiangping Chen, Ph.D. is a Professor and Chair of the Department of Information Science in the College of Information at the University of North Texas. She is also the Editor-in-Chief of the Electronic Library.

Wei Lu, Ph.D. is a Professor and Associate Dean at the School of Information Management at Wuhan University. He directs the iSchool's Information Retrieval and Data Mining Lab.

Oksana L.Zavalina, Ph.D. is an Associate Professor at the Department of Information Science at the University of North Texas.

REFERENCES

[1] Daniel Dahlmeier and Hwee Tou Ng. 2010. Domain adaptation for semantic role labeling in the biomedical domain. *Bioinformatics* 26 8 (2010), 1098–104.
[2] Joshua Herring Damir Cavar and Anthony Meyer. 2018. Case Law Analysis using Deep NLP and Knowledge Graphs. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (7-12), Georg Rehm, VÃηctor RodrÃηguez-Doncel, and JuliÃạn Moreno-Schneider (Eds.). European Language Resources Association (ELRA), Paris, France.
[3] Yang Deng, Yaliang Li, Nan Du, Wei Fan, Ying Shen, and Kai Lei. 2018. When Truth Discovery Meets Medical Knowledge Graph: Estimating Trustworthiness Degree for Medical Knowledge Condition. *CoRR* abs/1809.10404 (2018). arXiv:1809.10404 http://arxiv.org/abs/1809.10404
[4] Jane Greenberg. 2017. Big Metadata, Smart Metadata, and Metadata Capital: Toward Greater Synergy Between Data Science and Metadata. *Journal of Data and Information Science* 2, 3 (2017).
[5] Heiko Paulheim. 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8 (2017), 489–508.
[6] Dorothea Salo. 2010. Retooling libraries for the data challenge. *Ariadne* 64 (01 2010).
[7] Francesca Tomasi. 2018. Modelling in the Digital Humanities: Conceptual Data Models and Knowledge Organization in the Cultural Heritage Domain. *Historical Social Research / Historische Sozialforschung. Supplement* 31 (2018), 170–179. https://www.jstor.org/stable/26533637
[8] Ruijie Wang, Yuchen Yan, Jialu Wang, Yuting Jia, Ye Zhang, Weinan Zhang, and Xinbing Wang. 2018. AceKG: A Large-scale Knowledge Graph for Academic Data Mining. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018.* 1487–1490. https://doi.org/10.1145/3269206.3269252
[9] Zhiwu Xie and Edward Fox. 2017. Advancing library cyberinfrastructure for big data sharing and reuse. *Information Services Use* 37 (10 2017), 319–323. https://doi.org/10.3233/ISU-170853