Open Access

Open Journal of Web Technologies (OJWT)
Volume 5, Issue 1, 2018
http://www.ronpub.com/ojwt
ISSN 2199-188X

The First International Workshop on *Web Data Processing & Reasoning* (WDPAR 2018)

Sven Groppe [A], Christophe Cruz [B]

[A]Institute of Information Systems (IFIS), University of Lübeck,
Ratzeburger Allee 160, D-23562 Lübeck, Germany, groppe@ifis.uni-luebeck.de

[B]Laboratoire LE2I, Université de Bourgogne Franche-Comté, 67, rue Sully, 21078, Dijon Cedex, France,
christophe.cruz@u-bourgogne.fr

Abstract

The first Web Data Processing & Reasoning (WDPAR) workshop in conjunction with the 41st German Conference on Artificial Intelligence (KI) taking place in Berlin, Germany in 2018 is a forum for all researchers especially interested in processing of and reasoning on web data. The proceedings of WDPAR@KI 2018 are published in the Open Journal of Web Technologies (OJWT) (www.ronpub.com/ojwt) as special issue and the publisher of OJWT is RonPub. This editorial provides an overview over the aims and scope of the workshop and the review procedure. Furthermore, we introduce the accepted papers and their topics in the editorial.

Type of Paper and Keywords

Editorial: *Web data, reasoning, artificial intelligence, web security, anonymous shopping, search engines, public transportation, recommender systems, editorial, open access, WDPAR, KI, OJWT, RonPub*

1 Aims of the Workshop

The workshop on Web Data Processing and Reasoning (WDPAR) aims to provide a forum for scientists, researchers and practitioners to discuss and exchange ideas, new techniques, approaches and applications related to the web, e.g. approaches and applications utilizing web data (i.e., data collected from the web, available in the web or being about the web), execution environments in the web and advanced web applications. A special focus of the workshop is on intelligent web data processing via reasoning for data exchange, data integration and advanced applications.

The WDPAR workshop is associated with the 41st German Conference on Artificial Intelligence (KI) taking place in Berlin, Germany from 24-28th, September, 2018. The papers accepted in the WDPAR@KI 2018 are published as special issue in the Open Journal of Web Technologies (OJWT)[1] that is issued by RonPub[2].

This paper is accepted at the *International Workshop on Web Data Processing & Reasoning (WDPAR 2018)* in conjunction with the 41st German Conference on Artificial Intelligence (KI) in Berlin, Germany. The proceedings of WDPAR@KI 2018 are published in the Open Journal of Web Technologies (OJWT) as special issue.

[1] OJWT: `www.ronpub.com/ojwt`
[2] RonPub: `www.ronpub.com`

1.1 Types of Papers

The WDPAR workshop accept papers, which reports the latest research work in and make significant contributions to this field of Web data processing & Reasoning. Four types of papers are solicited:

- **Research papers:** which propose new approaches, theories or techniques related to Web Data Processing and Reasoning including new data structures, algorithms and whole systems. They should make substantial theoretical and empirical contributions to the research field.
- **Experiments and analysis papers:** which focus on the experimental evaluation of existing approaches including data structures and algorithms and bring new insights through the analysis of these experiments. Results of Experiments and Analysis Papers can be, for example, showing benefits of well-known approaches in new settings and environments, opening new research problems by demonstrating unexpected behavior or phenomena, or comparing a set of traditional approaches in an experimental survey.
- **Application papers:** which report new applications and practical experiences on applications related to the web. Application Papers might describe how to apply Web technologies to specific application domains with web-scale data demands like social networks, web search, e-business, collaborative environments, e-learning, medical informatics, bioinformatics and geographic information system. We especially welcome Application Papers, which develop applications utilizing reasoning about web data in a new way.
- **Vision papers:** which identify emerging or future research issues and directions, and describe new research visions for web data processing and reasoning. The new visions will potentially have great impacts on society.

1.2 Topics of Interest

Topics relevant to this workshop include, but are NOT limited to:

- Optimization of web data processing
- Applications of web data
- Reasoning on web data
- Advanced and new forms of web applications
- Web streams
  - Continuous queries and reasoning
  - Stream pipelines
  - Fault-tolerant stream processing
  - Optimized processing of a web-scale number of queries
- Semantic web
  - Applications
  - Semantic big data
  - Reasoning
  - Semantic data and query processing
- Linked data
  - Integration of heterogeneous linked data
  - Real-world applications
  - Statistics and visualizations
  - Quality
  - Ranking techniques
  - Provenance
  - Mining and consuming linked data
- Distributed processing of data in the web
  - Distributed reasoning
  - Distributed rule processing
  - Distributed query processing
  - Fault-tolerant processing
- Processing and reasoning of web services
  - Orchestration
  - New protocols for web services
- Rule processing of web data
- Evaluation of different Learning Approaches in the context of Web Data and Web Applications: unsupervised learning, supervised or reinforced learning, transfer learning, zero-shot learning, adversarial networks, and deep probabilistic models
- Knowledge representation and retrieval of web data
- Web Data exploration and visualization
- Web data mining
- Learning for web database tuning and web query optimization
- Case studies of AI-accelerated web workloads
- Web protocols and standards
- AI-enabled web data integration strategies
- Web security and privacy
- Web trust
- Natural language processing for web applications
  - Queries and chatbot interfaces
  - Result summarization
- Evaluating quality of approximate results from AI-enabled web queries

2 SUBMISSIONS AND REVIEWS

2.1 General Co-Chairs and Technical Program Committee

Two general co-chairs (see Appendix A) and ten technical program committee members (listed in Appendix B) have served as experts on the topics of interest of the workshop. They are researchers and practitioners from academia (83%) and industry (17%) from six nations.

2.2 Submission Statistics and Review Procedure

The workshop received eight paper submissions, five of which were accepted for presentation and publication in the workshop proceedings (62,5%). Each paper was typically reviewed by at least two members from the program committee (by three members in most of the cases), who remained anonymous to the authors *(single blind review)*. The complete review process of papers co-authored by a workshop chair or one of their colleagues was handled by the other workshop chair in order to avoid a conflict of interest. Additionally these papers were reviewed in a double blind review, where the reviewers are not aware of the authors. The reviewers evaluated the papers according to the following aspects:

- Relevance to the workshop
- Novelty and practical impacts
- Technical soundness
- Appropriateness and adequacy of:
  - Literature review
  - Background discussion
  - Analysis of issues
- Presentation, including:
  - Overall organization
  - English
  - Readability

3 WORKSHOP PROCEEDINGS

In this part, we first discuss in Section 3.1 why we have decided to publish the workshop proceedings in the Open Journal of Web Technologies (OJWT). We then provide a short overview of the accepted papers in Section 3.2.

3.1 Publication of the Workshop Proceedings

The open access model guarantees high visibility in the research community and usage of published results – hence leading to a potentially high impact. We have therefore chosen a journal applying the open access model for publishing the workshop proceedings of WDPAR 2018 as special issue. Furthermore, we attach importance to a journal not asking for transferring copyright, but applying the license model. Hence, we finally decided to choose the Open Journal of Web Technologies (OJWT)[3], which is an open access, peer-reviewed, academic journal published by RonPub[4]. OJWT distributes its articles under Creative Commons Attribution License[5], which permits unrestricted use, distribution, and reproduction free of charge in any medium, provided that the original work is properly cited. Furthermore, by publishing in OJWT, the papers will be indexed in major scientific indexes and the long-term preservation of the articles is ensured by the German National Library[6].

3.2 Overview of Accepted Papers

For each of our types of papers we have at least one out of five papers from this category: one regular research paper [5], one experiments and analysis paper [2], one application paper [1] and two visionary papers [3, 4].

The authors of the regular research paper [5] investigate insecure deployments of authoritative DNS servers which allow for hijacking of subdomains without the domain owner's consent. By using security holes of insecure deployments, attackers are able to perform effective man-in-the-middle attacks on the victim's online services, including TLS-secured connections, without having to touch the victim's DNS zone or leaving a trace on the machine providing the compromised service, such as the web or mail server. The authors also provide examples of such insecure deployments at real-world DNS operators and suggest remedies for the problem.

In order to attract many visitors to their own website, it is extremely important for website developers that their webpage is one of the best ranked webpages of search engines. As a rule, search engine operators do not disclose their exact ranking algorithm, so that website developers usually have only vague ideas about which measures have particularly positive influences on the webpage ranking. Conversely, the authors of the experiments and analysis paper [2] ask the question: "What are the properties of the best ranked webpages?" For this purpose, the authors perform a detailed analysis, in which they compare the properties of the best ranked webpages by widely used search engines with the worse

[3] https://www.ronpub.com/ojwt
[4] https://www.ronpub.com
[5] http://creativecommons.org/licenses/by/4.0/
[6] http://www.dnb.de/EN/Home/home_node.html

ranked webpages. Furthermore, they compare country-specific differences.

The application paper [1] deals with the complexity of urban public bus networks in big cities and aim to simplify their use. The authors present pervasive services for mobility called Notify.me that employs open data from the public bus network in Madrid. Their solution covers both a guiding service to assist users traveling by bus and a notifying service that informs them when a relevant point (i.e., the final destination or transfers) on their route has been reached. Notify.me computes a route including bus lines, transfers and pedestrian routes from a starting point (e.g., the user's current location), a destination and the user's preferences. The authors also provide the detail of a comprehensive user evaluation of their developed services.

The two visionary papers [3, 4] describe systems and approaches, which have not been implemented yet, but seems to have great potential for future work and having a high impact on (future) daily live.

Word embeddings makes it possible to find equivalent terms between experts and non-experts, by approach the similarity between words or by revealing hidden semantic relations. The authors of [3] propose hence an approach to use the most advanced word embeddings techniques to bridge the gap between the discourses of experts and non-experts and more specifically the terminologies used by the two communities. Controlled vocabularies with these new semantic enrichments are exploited in a hybrid recommendation system incorporating content-based ontology and keyword-based ontology to obtain relevant wines recommendations regardless of the level of expertise of the end user. The major aim is to find a non-expert vocabulary from semantic rules to enrich the knowledge of the ontology and improve the indexing of the items (i.e. wine) and the recommendation process.

Whenever clients shop in the Internet, they provide identifying data of themselves to parties like the webshop, shipper and payment system. These identifying data merged with their shopping history might be misused for targeted advertisement up to possible manipulations of the clients. The data also contains credit card or bank account numbers, which may be used for unauthorized money transactions by the involved parties or by criminals hacking the parties' computing infrastructure. In order to minimize these risks, the authors of [4] propose an approach for anonymous shopping by separation of data. The authors argue for the feasibility of their approach by discussing important operations like simple reclamation cases and criminal investigations.

4 Summary

The WDPAR workshop aims to attract researchers and practitioners discussing on problems occurring during processing of and reasoning on web data. This year's edition covers many different aspects especially web security (domain hijacking and anonymous shopping), public transportation apps, recommender systems and ranking of webpages in common search engines.

We hope our readers enjoy this selection of papers addressing research about processing of and reasoning on web data.

References

[1] M. Ángel Garrido Blázquez, P. Cáceres, B. Vela, C. E. Cuesta, J. M. C. Barca, and A. Sierra-Alonso, "Consuming web data in a guiding app for public bus users," *Open Journal of Web Technologies (OJWT)*, vol. 5, no. 1, pp. 31–43, 2018, Special Issue: Proceedings of the International Workshop on Web Data Processing & Reasoning (WDPAR 2018) in conjunction with the 41st German Conference on Artificial Intelligence (KI) in Berlin, Germany. [Online]. Available: https://www.ronpub.com/ojwt/OJWT_2018v5i1n05_Blazquez.html

[2] S. Babayigit and S. Groppe, "Webpage ranking analysis of various search engines with special focus on country-specific search," *Open Journal of Web Technologies (OJWT)*, vol. 5, no. 1, pp. 44–64, 2018, Special Issue: Proceedings of the International Workshop on Web Data Processing & Reasoning (WDPAR 2018) in conjunction with the 41st German Conference on Artificial Intelligence (KI) in Berlin, Germany. [Online]. Available: https://www.ronpub.com/ojwt/OJWT_2018v5i1n06_Babayigit.html

[3] C. Cruz, C. N. Van, and L. Gautier, "Word embeddings for wine recommender systems using vocabularies of experts and consumers," *Open Journal of Web Technologies (OJWT)*, vol. 5, no. 1, pp. 23–30, 2018, Special Issue: Proceedings of the International Workshop on Web Data Processing & Reasoning (WDPAR 2018) in conjunction with the 41st German Conference on Artificial Intelligence (KI) in Berlin, Germany. [Online]. Available: https://www.ronpub.com/ojwt/OJWT_2018v5i1n04_Cruz.html

[4] S. Groppe, F. Kuhr, and M. A. Coskun, "Anonymous shopping in the internet by separation of data," *Open Journal of Web Technologies (OJWT)*, vol. 5, no. 1, pp. 14–22, 2018, Special Issue: Proceedings of the International Workshop on Web

Data Processing & Reasoning (WDPAR 2018) in conjunction with the 41st German Conference on Artificial Intelligence (KI) in Berlin, Germany. [Online]. Available: https://www.ronpub.com/ojwt/OJWT_2018v5i1n03_Groppe.html

[5] P. Thomassen, J. Benninger, and M. Margraf, "Hijacking dns subdomains via subzone registration: A case for signed zones," *Open Journal of Web Technologies (OJWT)*, vol. 5, no. 1, pp. 6–13, 2018, Special Issue: Proceedings of the International Workshop on Web Data Processing & Reasoning (WDPAR 2018) in conjunction with the 41st German Conference on Artificial Intelligence (KI) in Berlin, Germany. [Online]. Available: https://www.ronpub.com/ojwt/OJWT_2018v5i1n02_Thomassen.html

APPENDIX A - WORKSHOP CHAIRS

The workshop is organized by:

- Sven Groppe, University of Lübeck, Germany
- Christophe Cruz, Université de Bourgogne Franche-Comté, France

APPENDIX B - PROGRAM COMMITTEE

Our Program Committee members include the following experts:

- Mithun Balakrishna, Lymba Corporation, USA
- Tanya Braun, University of Lübeck, Germany
- Josue Balandrano Coronel, The University of Texas at Austin, USA
- Thomas Eisenbarth, University of Lübeck, Germany
- Marcel Gehrke, University of Lübeck, Germany
- Hasan Ali Khattak, Comsats Institute of Information Technology, Islamabad, Pakistan
- Hariharan Krishnaswamy, DELL Technologies, USA
- Gianfranco E. Modoni, National Research Council of Italy - Bari, Italy
- Özgür L. Özçep, University of Lübeck, Germany
- Nuno Silva, Instituto Superior de Engenharia do Porto, Portugal

AUTHOR BIOGRAPHIES

Sven Groppe earned his diploma degree in Informatik (Computer Science) in 2002 and his Doctor degree in 2005 from the University of Paderborn. He earned his habilitation degree in 2011 from the University of Lübeck. He worked in the European projects B2B-ECOM, MEMPHIS, ASG and TripCom. He was a member of the DAWG W3C Working Group, which developed SPARQL. He was the project leader of the DFG project LUPOSDATE, an open-source Semantic Web database, and one of the project leaders of two research projects, which research on FPGA acceleration of relational and Semantic Web databases. He is also the chair of the Semantic Big Data workshop series, which is affiliated with the ACM SIGMOD conference (so far 2016 to 2018), and of the Very Large Internet of Things workshop in conjunction with the VLDB conference in 2017 and 2018. His research interests include databases, Semantic Web, query and rule processing and optimization, Cloud Computing, peer-to-peer (P2P) networks, Internet of Things, data visualization and visual query languages.

Christophe Cruz is an Associate Professor at the laboratory Le2i at the University of Bourgogne, and Vice-President in charge of Digital Policy and Information Systems of the University UBFC (Université Bourgogne-Franche-Comté). He leads the cluster team "Smart Environment" of the laboratory Le2i. Previously, he was a postdoctoral fellow at the i3mainz - Institute for Spatial Information and Surveying Technology working with Prof. Dr.-Ing. Frank Boochs. Christophe Cruz obtained a Ph.D. in Computer Science in 2005 and a habilitation degree in 2012 at the University of Bourgogne. He leads both academic and industrial projects ranging from knowledge modeling to Big Data processing. (HiGeoMes ANR-10-FRAL-0003, TexTelSem ANR-13-FRAL-0008, First Eco). His research interests include Description Logics, Semantic Web, Data Science, Recommender Systems, Big Data, Spatio-temporal Data, Text processing, Knowledge Modeling and user-centric systems.