Communications in Computer and Information Science 1459

More information about this series at http://www.springer.com/series/7899

Boris Villazón-Terrazas ·
Fernando Ortiz-Rodríguez ·
Sanju Tiwari · Ayush Goyal ·
MA Jabbar (Eds.)

Knowledge Graphs and Semantic Web

Third Iberoamerican Conference and Second Indo-American Conference, KGSWC 2021 Kingsville, Texas, USA, November 22–24, 2021 Proceedings

Springer

Editors
Boris Villazón-Terrazas
Tinamica/Universidad Internacional de la Rioja
Madrid, Spain

Sanju Tiwari
Tamaulipas Autonomous University
Ciudad Victoria, Mexico

MA Jabbar
Vardhaman College of Engineering
Hyderabad, Telangana, India

Fernando Ortiz-Rodríguez
Tamaulipas Autonomous University
Ciudad Victoria, Mexico

Ayush Goyal
Texas A&M University – Kingsville
Kingsville, TX, USA

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-030-91304-5 ISBN 978-3-030-91305-2 (eBook)
https://doi.org/10.1007/978-3-030-91305-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains the main proceedings of the Third Iberoamerican Knowledge Graph and Semantic Web Conference and the Second Indo-American Knowledge Graphs and Semantic Web Conference (KGSWC 2021), which were held jointly during November 22–24, 2021. KGSWC is established as a yearly venue for discussing the latest scientific results and technology innovations related to knowledge graphs and the Semantic Web. At KGSWC, international scientists, industry specialists, and practitioners meet to discuss knowledge representation, natural language processing/text mining, and machine/deep learning research. The conference's goals are (a) to provide a forum for the AI community, bringing together researchers and practitioners in the industry to share ideas about research and development projects, and (b) to increase the adoption of AI technologies in these regions.

KGSWC 2021 took place virtually from Kingsville, Texas, USA, building on the success of past events in 2019 and 2020. It was also a venue for broadening the focus of the Semantic Web community to span other relevant research areas in which semantics and web technology play an important role and for experimenting with innovative practices and topics that deliver extra value to the community.

The main scientific program of the conference comprised 24 papers: 22 full research papers and two short research papers selected out of 85 reviewed submissions, which corresponds to an acceptance rate of 28.2%. The conference was completed with four workshops, a hackathon, and a winter school where researchers could present their latest results and advances and learn from experts. The program also included five exciting invited keynotes (James Hendler, Amit Sheth, Soren Auer, Pascal Hitzler, and Paco Nathan), with novel Semantic Web topics.

The General and Program Committee chairs would like to thank the many people involved in making KGSWC 2021 a success. First, our thanks go to the local chairs of the main event and to the Program Committee for ensuring a rigorous review process that led to an excellent scientific program with an average of three reviews per paper.

We also had a great selection of workshops and tutors from the winter school. Thanks to our Cuban peers for the continued support of this event and for organizing IWSW 2021: the 4th International Workshop on Semantic Web. We had the opportunity to initiate two workshops in Europe with the support of our colleagues in ENGIE-France (PGMOnto2021: the First International Workshop on Joint Use of Probabilistic Graphical Models and Ontology and IWDLQ2021: the First International Workshop on Deep Learning for Question Answering). Thanks to Patience Usoro Usip for the organization of IWMSW-2021: the First International Workshop on Multilingual Semantic Web. We are also grateful to Amit Sheth's team at the Artificial Intelligence Institute of the University of South Carolina, USA, for the work and commitment involved in organizing the hackathon for the second year in a row.

Further, we thank the kind support of Springer. We also thank Antonio Cardona and Gerardo Haces, who administered the website. We finally thank our sponsors and our

community for their vital support of this edition of KGSWC. The editors would like to close the preface with warm thanks to our supporting keynote speakers and our enthusiastic authors who made this event truly international.

November 2021

Boris Villazón-Terrazas
Fernando Ortiz-Rodríguez
Sanju Tiwari
Ayush Goyal
MA Jabbar

Organization

Chairs

Boris Villazón-Terrazas	Tinámica/Universidad Internacional de La Rioja, Spain
Fernando Ortiz-Rodriguez	Universidad Autonoma de Tamaulipas, Mexico
Sanju Tiwari	Universidad Autónoma de Tamaulipas, Mexico

Local Chairs

Ayush Goyal	Texas A&M University – Kingsville, USA
MA Jabbar	Vardhaman College of Engineering, India

Workshops

Shishir Shandilya	VIT Bhopal University, India

Winter School

Sanju Tiwari	Universidad Autónoma de Tamaulipas, Mexico
Fernando Ortiz-Rodriguez	Universidad Autonoma de Tamaulipas, Mexico
Boris Villazón-Terrazas	Tinámica/Universidad Internacional de La Rioja, Spain

Hackathon

Kaushik Roy	University of South Carolina, USA
Manas Gaur	University of South Carolina, USA

Program Committee Chairs

Sanju Tiwari	Universidad Autónoma de Tamaulipas, Mexico
Fernando Ortiz-Rodriguez	Universidad Autonoma de Tamaulipas, Mexico
Boris Villazón-Terrazas	Tinámica/Universidad Internacional de La Rioja, Spain

Program Committee

Aidan Hogan	Universidad de Chile, Chile
Alberto Fernandez	University Rey Juan Carlos, Spain
Alejandro Rodriguez	Universidad Politécnica de Madrid
Amed Abel Leiva Mederos	Universidad Central de las Villas, Cuba
Ana B. Rios-Alvarado	Universidad Autónoma de Tamaulipas, Mexico
Anastasija Ņikiforova	University of Latvia, Latvia
Anna Fensel	University of Innsbruck, Austria

Armin Haller	Australian National University, Australia
Ayush Goyal	Texas A&M University – Kingsville, USA
Boris Villazón-Terrazas	Tinámica/Universidad Internacional de La Rioja, Spain
Boyan Brodaric	Geological Survey of Canada, Canada
Claudio Gutierrez	Universidad de Chile, Chile
Cogan Shimizu	Kansas State University, USA
Demian Abrego Almazán	Universidad Autónoma de Tamaulipas, Mexico
Diego Collarana	Fraunhofer IAIS, Germany
Edelweis Rohrer	Universidad de la República, Uruguay
Edgard Marx	Leipzig University of Applied Sciences, Germany
Edgar Tello Leal	Universidad Autónoma de Tamaulipas, Mexico
Erick Antezana	Norwegian University of Science and Technology, Norway
Eric Pardede	La Trobe University, Australia
Fatima N. Al-Aswadi	Universiti Sains Malaysia, Malaysia
Fernando Ortiz-Rodriguez	Universidad Autónoma de Tamaulipas, Mexico
Gerardo Haces-Atondo	Universidad Autónoma de Tamaulipas, Mexico
Gustavo De Assis Costa	Federal Institute of Education, Science and Technology, Goiás, Brazil
Harry Halpin	Inria Paris, France
Harshvardhan J. Pandit	Trinity College Dublin, Ireland
Irlan Grangel	Bosch Corporate Research, Germany
Ismael Navas-Delgado	University of Malaga, Spain
Ivan Lopez-Arevalo	Cinvestav – Tamaulipas, Mexico
Jean-Paul Calbimonte	University of Applied Sciences and Arts of Western Switzerland, Switzerland
Jose L. Martinez-Rodriguez	Universidad Autónoma de Tamaulipas, Mexico
José Luis Redondo García	Amazon Research, UK
Jose María Alvarez Rodríguez	Carlos III University of Madrid, Spain
Jose Melchor Medina-Quintero	Universidad Autónoma de Tamaulipas, Mexico
Juan Antonio Lossio-Ventura	National Institutes of Health, USA
Juan Carlos Olivares Rojas	Instituto Tecnológico de Morelia, Mexico
Lu Zhou	Kansas State University, USA
MA Jabbar	Vardhaman College of Engineering, India
Manuel Enrique	Universidad de las Ciencias Informáticas, Cuba
Maria Hallo	Escuela Politécnica Nacional, Ecuador
María Poveda-Villalón	Universidad Politécnica de Madrid, Spain
Maricela Bravo	Universidad Autónoma Metropolitana, Mexico
Marlene Goncalves	Universidad Simón Bolívar, Venezuela
Mayank Kejriwal	University of Southern California, USA
Md Kamruzzaman Sarker	Kansas State University, USA
Miguel-Angel Sicilia	University of Alcala, Spain

Nandana Mihindukulasooriya	IBM Research, USA
Nelson Piedra	Universidad Técnica Particular de Loja, Ecuador
Nicholas Beliz-Osorio	Universidad Tecnológica de Panamá, Panama
Panos Alexopoulos	Textkernel B.V., The Netherlands
Patience Usoro Usip	University of Uyo, Nigeria
Regina Motz	Universidad de la República, Uruguay
Rim Hantach	ENGIE, France
Roberto Navigli	Sapienza University of Rome, Italy
Rosa M. Rodríguez	University of Jaén, Spain
Russa Biswas	FIZ Karlsruhe, Germany
Samir Sellami	ENSET-Skikda, Algeria
Sanju Tiwari	Universidad Autónoma de Tamaulipas, Mexico
Shikha Mehta	Jaypee Institute, India
Takanori Ugai	Fujitsu Laboratories Ltd., Japan
Tassilo Pellegrini	St. Pölten University of Applied Sciences, Austria
Terunobu Kume	Fujitsu Laboratories Ltd., Japan
Tommaso Soru	University of Leipzig, Germany
Victor Saquicela	Universidad de Cuenca, Ecuador
Yusniel Hidalgo Delgado	Universidad de las Ciencias Informáticas, Cuba
Zarour Nacer Eddine	University of Constantine 2, Algeria

Keynote Abstracts

Knowledge Graph Semantics

James Hendler

Institute for Data Exploration and Applications, Rensselaer Polytechnic Institute

Oh dear, there's that word again – "semantics!" Isn't that what doomed that Semantic Web thing and led to knowledge graphs instead? In fact, many of the same problems, and particularly problems with interoperability, arise again for KGs, and thus we must explore the old problem in this new area. This is even more important when we start to explore the "personal knowledge graph (PKG)," that is, the ability to have private and public information combined in KG technology. In this talk, I discuss how knowledge graphs, PJGs, linked data and, yes, semantics are all critically linked and why the latter is still relevant to the growth and scaling of knowledge graphs into the future - and specifically to the ability to extract better data from them.

Don't Handicap AI without Explicit Knowledge

Amit Sheth

AI Institute at the University of South Carolina

Knowledge representation as expert system rules or using frames and variety of logics, played a key role in capturing explicit knowledge during the hay days of AI in the past century. Such knowledge, aligned with planning and reasoning are part of what we refer to as Symbolic AI. The resurgent AI of this century in the form of Statistical AI has benefitted from massive data and computing. On some tasks, deep learning methods have even exceeded human performance levels. This gave the false sense that data alone is enough, and explicit knowledge is not needed. But as we start chasing machine intelligence that is comparable with human intelligence, there is an increasing realization that we cannot do without explicit knowledge. Neuroscience (role of long-term memory, strong interactions between different specialized regions of data on tasks such as multimodal sensing), cognitive science (bottom brain versus top brain, perception versus cognition), brain-inspired computing, behavioral economics (system 1 versus system 2), and other disciplines point to need for furthering AI to neuro-symbolic AI (i.e., hybrid of Statistical AI and Symbolic AI, also referred to as the third wave of AI). As we make this progress, the role of explicit knowledge becomes more evident. I will specifically look at our endeavor to support higher-level intelligence than what current AI systems support, our desire for AI systems to interact with humans naturally, and our need to explain the path and reasons for AI systems' workings. Nevertheless, the variety of knowledge needed to support understanding and intelligence is varied and complex. Using the example of progressing from NLP to NLU, I will demonstrate varieties of explicit knowledge (represented as knowledge graphs), which may include, linguistic, language syntax, common sense, general (world model), specialized (e.g., geographic), and domain-specific (e.g., mental health) knowledge. I will also argue that despite this complexity, such knowledge can be scalability created and maintained (even dynamically or continually). Finally, I will describe our work on knowledge-infused learning as an example strategy for fusing statistical and symbolic AI in a variety of ways.

Graph Thinking

Paco Nathan

Managing Partner at Derwen, Inc.

In an effort to bridge between current research and industry practices for graph technologies, a few observations help. First, recent innovations such as graph neural networks provide immediate solutions in business use cases for inference which have perplexed the semantic web community for decades. Second, given more stringent reproducibility requirements for researchers to publish source code and datasets alongside peer-reviewed papers, a skills gap emerges in the crucial area of open source integration. Third, while currently there are more than 30 vendors in the graph database space, and data management is of course quite important, lessons from the arc of Big Data circa 2006–2015 indicate that demands for scalable graph computation outpace the supply focused on database-centric approaches – as industry use cases confirm. These factors are disruptive for current industry narratives, as we shall explore in the talk. They also represent substantial opportunities for the research community.

Since the late–2010s, business use cases for graph technologies have become more widespread. Along with that, a practice of graph data science has emerged, blending graph capabilities into existing data science teams and industry training curricula. However, these nascent practices tend to face common hurdles: (1) enterprise IT staff who are familiar with relational databases yet unfamiliar with graph use cases, so they tend to discourage the latter (turf wars); (2) the fact that graph practices in industry at generally must cut across company divisions and lines of executive responsibility – e.g., where an enterprise knowledge graph integrates data from logistics, research, and market analysis to provide a "business 360" for top execs – so this tends to alarm mid-level executives (more turf wars); and (3) graph database vendors tenaciously concerned with database features and corresponding lock-in for licensing strategies that don't scale effectively, while generally less concerned with horizontal scale-out for graph computing (the turf wars to come).

Taking inventory of these observations and hurdles commonly encountered, enterprise graph practices are being pushed toward open source integration to obtain effective technology solutions. That said, other challenges remain for graph technologies which can be described in terms of behavioral economics. The academic community should become more fully aware about these points in particular. This talk explores current solutions and near-term outlook including: (1) graph thinking as a cognitive framework for approaching complex problem spaces; (2) open source integration as the key both for viable industry practices as well as an essential pathway for validation of academic research; and (3) a survey of industry use cases aside from the "usual suspects" of advertising, e-commerce, social networks, and financialization.

One overall theme emerges from these arguments: whereas data science use cases from a decade ago tended to focus on extractive business models and "low-hanging fruit" such as online marketing or financialization, the graph data science uses have veered away from the tech giants, focusing instead on more complex industry problems: waste mining and circular economy in the manufacturing, network medicine and personalized drug discovery in pharma, identifying obfuscated cyber attacks in security, tracking carbon footprint across thousands of vendors in global supply chain, and so on. This talk will explore the inherent connections between graph technologies and complexity, along with what learnings can be applied to AI from more established work in pedagogy, behavioral economics, and leadership. NB: core parts of this material were prepared during a 2021 independent assessment of the graph technologies space, conducted on behalf of a large EU-based manufacturing firm, and the approaches described here have been developed working with business units in the context of their production practices.

Building Knowledge Graphs Leveraging Expert, Crowd and Machine Intelligence

Soren Auer and Allard Oelen

Director TIB, Head of Research Group Data Science and Digital Libraries

The Need for a Scholarly Knowledge Graph. The amount of published scholarly articles remains to grow steadily every year. Therefore, new methods are needed to organize this growing number of publications. Traditionally, scholarly communication is largely document-based, hindering machine-actionability of the presented knowledge. If scholarly knowledge is provided in a structured format, via knowledge graphs, the knowledge becomes readable for machines. This addresses, and largely resolves, the issues scholarly communication currently faces. However, creating such a knowledge graph is a complicated endeavor. Automated techniques, possibly powered by machine learning, are currently not sufficiently accurate to autonomously create a high-quality knowledge graph. The quality aspect of the graph is crucial for it to become a valuable tool for researchers. On the other hand, a fully manual approach, in which humans create structured paper descriptions, does not scale well to large quantities of papers. A hybrid approach can solve the quality issues of the automated approach while also addressing the scalability issues of the manual approach. In this hybrid approach, human intelligence is supported by machine intelligence to create a scalable high-quality knowledge graph.

Open Research Knowledge Graph. The Open Research Knowledge Graph (ORKG) is a scholarly knowledge graph that contains high-quality manually curated scholarly knowledge. In contrast to other scholarly knowledge graphs, the data in the ORKG represents the actual knowledge presented in the paper, instead of merely metadata. This includes knowledge related to research problems, materials, methods, and results. This structured data is used to semi-automatically create overviews of the state-of-the-art. These overviews are called "comparisons" in the ORKG. Comparisons can be used for different purposes, for example, to compare results of different approaches to the same research problem (i.e., leader boards), to show an overview of the transformation of research results over time, or to provide comprehensive summaries of related work.

The ORKG leverages crowdsourcing to populate the graph. At the core, users (who are generally researchers) enter data related to their work and publications. They collaborate with knowledge engineers, which includes subject librarians, to determine what to describe and how to describe this. This includes the generation of "Graph templates", a construct to unify different contribution descriptions for papers addressing the same research problem. Research can be organized in ORKG observatories which address a specific set of research problems maintained by researchers and domain experts for the respective research field.

Human and Machine Intelligence. Within the ORKG, multiple approaches are used to populate the graph. As previously mentioned, the ORKG does not use manual or automated approaches in isolation but leverages hybrid forms instead. This includes machine-in-the-loop and human-in-the-loop approaches. In the former approach, the machine assists the user, while in the latter approach the human assists the machine. For the machine-in-the-loop approach, the human initiates the process of adding new content to the graph and intelligent machine components assist the user while doing so. The ORKG leverages this approach at several locations, including the “Abstract annotator”, which extracts scientific concepts from a user-provided abstract. The user can decide which concepts to add and which to remove. This approach is also used in the “Sentence annotator” in which the user is assisted by an intelligent user interface to annotate key sentences within a scholarly article. The intelligent components include key sentence highlighting and annotation class suggestions.

Finally, a human-in-the-loop approach is used in a novel ORKG feature that is currently under development. This feature autonomously processes scholarly article abstracts at scale using a set of different Natural Language Processing (NLP) tools. The resulting data from the NLP processing will be validated by humans in the form of microtasks. Each task presents an easy to answer, for example asking the user whether a displayed statement is correct or not. User votes are stored as provenance data on the statements, allowing others to evaluate the credibility and correctness of a statement based on the user votes. We envision that any ORKG user, whether it is an incidental visitor or a researcher, can vote on statements and that aggregating all votes provides a scalable and high-quality method for validating scholarly NLP data.

Modular Schema Development for Wikibase

Pascal Hitzler

Endowed Lloyd T. Smith Creativity in Engineering Chair and Director of the Center for Artificial Intelligence and Data Science at Kansas State University

Wikibase - the software that underlies Wikidata - provides excellent support for knowledge graph construction. Key features include an intuitive MediaWiki interface, native support for context (qualifiers) and provenance (references), and compatibility with W3C standards, in particular RDF. Originally developed for the open crowd-sourced setting of Wikidata, it can also be used for knowledge graph development in more controlled use cases. However, if the use case calls for a tightly defined graph schema in the form of a traditional ontology, it is not immediately clear how to do traditional ontology development such that it is compatible with the modeling and graph structuring choices of Wikibase. In this talk, we will report on our recent efforts to close the gap between traditional ontology engineering and Wikibase. Drawing from our experiences in the Enslaved project which uses both ontology modeling and Wikibase to realize a database on the history of the slave trade, we present a solution based on ontology design patterns that closes the gap between the two approaches and thus enables ontology design that can then be used directly with Wikibase.

Contents