LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY



The Open Research Knowledge Graph and its Use for Plasma Physics

Int. Workshop on FAIR Data in Plasma Science May 18



Digitalization in everyday life

Navigation to TIB in Hannover

50 years ago



Sehenswürdigkeiten Rivel Cardina of Henrich Startpunkt des Roten Foders Starting point of the Red Thread tour or State Manes over Manuers of His Withein Busch Museur allepunkt der Stadmundtel Excist internation office Tautet Information Ernet-August-Purz II Tel: +49:511 12345-111 E-mail: Info@hancover-Iourismus.de



- + New Features:
 - \rightarrow Zoom in
 - \rightarrow Traffic jam warning
 - \rightarrow Opening hours
 - \rightarrow Interesting places around



Similar in other domains....



Who still remembers?



Mail order catalogs



Encyclopedias



Phone books

Whole industries got disrupted and our lives were significantly changed

What about Science?

Over 300 years ago

ACTA FRIIDITORIIM ANNO M DC L XXXIV publicata,

100 years ago



20 years ago

A Relational Model of Data for Large Shared Data Banks E. F. CODD IBM Research Laboratory, San Jose, California

Information Retrieva

The relational view (or model) of data described in Section 1 appears to be superior in several respects to the graph or network model [3, 4] presently in vogue for noninferential systems. It provides a means of describing data with its natural structure only-that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representa-

P. BAXENDALE, Editor



Today

BIBLIOTHEK - Forschung und Praxis 2020: 44(3): 516-529

DE GRUYTER

Textmining

Sören Auer*, Allard Oelen, Muhammad Haris, Markus Stocker, Jennifer D'Souza, Kheir Eddine Farfar, Lars Vogt, Manuel Prinz, Vitalis Wiens and Mohamad Yaser Jaradeh

Improving Access to Scientific Literature with **Knowledge Graphs**

Science does not harvest the full potential of digitalization

dicata. Cum S.Cafarea Majeltatis & Potentisfimi Ele-Eloris Saxonia Privilegiis, LIPSIE. Proflant apud J. GROSSIUM & J. F. GLETITSCHIUM, Typis CHRISTOPHORI GENTHERL Anno M DCLXXXIV.

are made, it enables no to prediet the exact course of all motions resulting from gravitation. In this book, which is a popular exposition written for the average reader, Professor Einstein explains his famous theory which has so excited the scientific world. This volume is intended primarily for those readers who, though interested in the trend of modern theory, are not conversant with the mathematical analysis used in theoretical physics. The author's aim has been to give an exact insight into the theory of relativity, and to present the main ideas in the clearest and simplest form. He has succeeded admirably, and these who desire an authoritative and understandable explanation of the Einstein theory will find it between the covers of this book.

HENRY HOLT AND COMPANY

calculus, security, data integrity CR CATEGORES: 3.70, 3.73, 3.75, 4.20, 4.22, 4.29

Volume 13 / Number 6 / June, 1970

This paper is concerned with the application of ele-mentary relation theory to systems which provide shared

access to large banks of formatted data. Except for a paper

by Childs [1], the principal application of relations to data

systems has been to deductive question-answering systems.

Levein and Maron [2] provide numerous references to work

In contrast, the problems treated here are those of data

inconsistency which are expected to become troublesome even in nondeductive systems.

1.1. INTRODUCTION

in this area.

without logically impairing some application programs is still quite limited. Further, the model of data with which users interact is still cluttered with representational prop-erties, particularly in regard to the representation of collections of data (as opposed to individual items). Three of the principal kinds of data dependencies which still need 1. Relational Model and Normal Form to be removed are: ordering dependence, indexing depend-ence, and access path dependence. In some systems these

dependencies are not clearly separable from one another. 1.2.1. Ordering Dependence. Elements of data in a data bank may be stored in a variety of ways, some involving no concern for ordering, some permitting each element to participate in one ordering only, others permitting each element to participate in several orderings. Let us consider those existing systems which either require or permit data elements to be stored in at least one total ordering which is closely associated with the hardware-determined ordering independence—the independence of application programs of address. For example, the records of all is concerning and terminal activities from growth in data types and parts might be stored in ascending order by part serial changes in data representation—and certain kinds of data number. Such systems normally permit application programs to assume that the order of presentation of records from such a file is identical to (or is a subordering of) the

Communications of the ACM 377

owiedge graph such as the ORKG can be used to give a manuellen (crowd/expert sourcing) und (halb-)automat condensed overview on the state-of-the-art addressing a signer Techniken ein. Nur mit einer solchen Kombination particular research quest, for example as a tabular com- aus menschlicher und maschineller Intelligenz können parison of contributions according to various characteris- wir die erforderliche Qualität der Darstellung erreichen, tics of the approaches. Further possible intuitive access um neuartige Explorations- und Unterstützungsdienste für interfaces to such scholarly knowledge graphs include Forscher zu ermöglichen. Im Ergebnis kann ein Wissensdomain-specific (chart) visualizations or answering of nat-ural language questions. graph wie der ORKG verwendet werden, um einen kompri-mierten Überblick über den Stand der Technik in Bezug auf

ohamad Yaser Jaradeh, yaser. jaradeh@tib.eu

*Corresponding author: Prof. Dr. Sören Auer, auer@tib.eu Allard Oelen, allard.oelen@tib.eu Muhammad Haris, muhammad.haris@tib.eu Dr. Markus Stocker, markus.stocker@tib.eu tels Question Answering. Dr. Jennifer D'Souza, jennifer. dsouza@tib.eu Kheir Eddine Farfar, kheir.farfar@tib.eu Lars Vogt, Lars.vogt@tib.eu Manuel Prinz, manuel.prinz@tib.eu Vitalis Wiens, vitalis.wiens@tib.eu

eine bestimmte Forschungsaufgabe zu geben, z.B. als ta-bellarischer Vergleich der Beiträge nach verschiedenen Merkmalen der Ansätze. Weitere mögliche intuitive Nutzungsschnittstellen zu solchen wissenschaftlichen Wissensgraphen sind domänenspezifische Visualisierungen oder die Beantwortung natürlichsprachlicher Fragen mit-

Keywords: Subject classification: knowledge graph: se

Schlüsselwörter: Sacherschließung; Wissensgraph; Se mantic Web; Crowdsourcing; Text Mining

Not much has changed!



A Consequence of Document Centered Information Flows: The Publication Flood

- ~ 2.5 Mio new publications per year
- Researchers lack overview, even in small fields
- Loss of knowledge
- Answering questions is like looking for a needle in the haystack





An Example – CRISPR



ES Lander - Cell, 2016 - Elsevier

... for CRISPR-based resistance, they set out to create the first artificial CRISPR arrays—programming CRISPR ... As they predicted, the strains carrying the new CRISPR sequence showed ... ☆ Speichern ワワ Zitieren Zitiert von: 538 Ähnliche Artikel Alle 20 Versionen

A CRISPR view of development

MM Harrison, BV Jenkins... - Genes & ..., 2014 - genesdev.cshlp.org ... as "spacers" between repetitive sequences in the CRISPR locus of the host genome. The CRISPR locus is transcribed and processed into short CRISPR RNAs (crRNAs) that guide the ... ☆ Speichern ワワ Zitieren Zitiert von: 272 Ähnliche Artikel Alle 10 Versionen

[HTML] CRISPR-based diagnostics

MM Kaminski, <u>OO Abudayyeh</u>, <u>JS Gootenberg</u>... - Nature Biomedical ..., 2021 - nature.com ... with the **CRISPR**-associated (Cas) enzyme. Although there are diverse **CRISPR**-Cas ... these systems are connected by their dependence on **CRISPR** RNA (crRNA), which guides ... ☆ Speichern ワワ Zitieren Zitiert von: 59 Ähnliche Artikel Alle 10 Versionen



Specific research questions:

- Who applied CRISPR to butterflies?
- How to apply CRISPR with minimal costs?
- How do different genome editing techniques compare?

Further Challenges of Document-Orientation



Reproducibility Crisis



ELSEVIER Monopolization of commercial actors



Deficiency of Peer-Review



Lack of machine assistance



Predatory Publishing

Time to Rethink Scholarly Communication!



The solution is not "better pdfs"...



"The lightbulb was **not** invented by improving the candle." **Oren Harari**

Digitalization is **more** than just Digitization! Current and future scientific challenges can not be tackled with an outdated communication system.

Digitalize Knowledge, Not Documents!

The Open Research Knowledge Graph





As the name already suggests, ORKG is a **knowledge graph**.

Knowledge Graphs are widely used in industry...





Why not use them for (open) science as well?

https://www.slideshare.net/Frank.van.Harmelen/adoption-of-knowledge-graphs-late-2019



Advantages of a Graph-Based Approach

- Machine-actionable
- Automated finding and linking of research contributions towards a specific problem
- Natural language question answering possible e.g. "How do different genome editing techniques compare?"



• Explore knowledge in entirely new ways

An Example: COVID-19 Basic Reproduction Number



Properties		The early phase of the COVID-19 outbreak in Lombardy, Italy 2020 - Contribution 1	Transmission potential of COVID-19 in Iran 2020 - Contribution 1	Transmission potential of COVID-19 in Iran 2020 - Contribution 2	Estimating the generation interval for COVID-19 based on symptom onset data 2020 - Contribution 1
location	T	Lombardy, Italy	Iran	Iran	Singapore
<u>Time period</u>	T	Time interval	Time interval	<u>Time interval</u>	<u>Time interval</u>
<u>has beginning</u>	T	2020-01-14	2020-02-19	2020-02-19	2020-01-21
<u>has end</u>	T	2020-03-08	2020-02-29	2020-02-29	2020-02-26
Basic reproduction number	ř	Basic reproduction number estimate value specification	Basic reproduction number estimate value specification	Basic reproduction number estimate value specification	Basic reproduction number estimate value specification
Has value	T	3.1	3.6	3.58	1.27
Confidence interval (95%)	ř	Confidence interval (95%)	Confidence interval (95%)	Confidence interval (95%)	Confidence interval (95%)
Lower confidence limit	T	2.9	3.4	1.29	1.19
Upper confidence limit	T	3.2	4.2	8.46	1.36
Method*	T		generalized growth model	based on the calculation of the epidemic's doubling times: estimated epidemic doubling time of 1.20 (95% CI, 1.05, 1.44) days	generation interval ite 14

ORKG's Objectives



Provide overview over the state-of-theart for specific research problems



Tackle interdisciplinary challenges such as climate change research, disease prevention, etc. Foster collaboration

Findable Accessible Interoperable

^Ny **60**

Make research FAIR





ORKG: Lighthouse in the Publication Flood





ORKG's Current Features





We are constantly improving and developing new features. Also interesting: Describe datasets & software

So, what about Plasma Physics?



A >> Physical Sciences & Mathematics >> Physics



So, what about Plasma Physics?



- Already some content: 98 described papers
- Last year: Curation Grants
 - 1 grantee from laser & plasma physics
 - 14 Comparisons, 1 Review
 - This year: Same grantee \rightarrow Content will increase
- Could be more...

Who creates ORKG content?

The answer is **YOU**!

We follow a **crowd based approach** for the curation process

Everyone can create, edit, add, complement, reuse, etc.





TIB **Observatories: Taking the Lead in Content Curation** Template 🔞 Quantity kind Description Properties Forma Ensure high quality standard Promote ORKG Template use cases Organize research Create templates Stay in contact with in your field and simplify using development team: **ORKG** for beginners Issues & Requests will be prioritized **Build a community knowledge** graph for your discipline

How to get an Observatory started



- Collect typical research questions from your field: What type of questions must the system be able to answer to be beneficial for your discipline?
 - E.g. "By how much do the theoretical values of transition energies deviate from measurements?"
- What is needed to answer these questions? And where can it be found in literature?
 - E.g. system characteristics (Element, ionization state, transition, which experiment etc.), results, numerical values



Data Curators

- Create a data model
- Template

Templates



- Describe papers, datasets, software
- Forms with predefined properties
- Easy to fill out
- Increase interoperability

Researchers

- Enter data: focus on content, no worries about data model
- Can reuse & expand

Research problem		
Element		
Isoelectronicity		
Ionization State		
Result		

Create a Community Knowledge Graph for your discipline!





Profit from organized knowledge & machine assistance Increase FAIRness

Team

Lead



Co-Lead



Dr. Markus Stocker



Dr. Jennifer D'Souza Dr. Lars Vogt





Dr. Oliver Karras



Dr. Vinodh Ilangovan



PhD Students

Allard Oelen

Developers



Golsa Heidari



Hassan Hussein



Muhammad Haris



Yaser Jaradeh



Salomon Kabenamualu Community Building



Kheir Eddine Farfar



Manuel Prinz



Omar Arab Oghli



Dr. Anna-Lena Lorenz Seite 25

Summary



Rethink scholarly communication

Machine-actionable knowledge representation for FAIR research





Crowd-based approach

Get involved with an Observatory and build a community knowledge graph for plasma physics



Learn more: orkg.org Contact us: info@orkg.org Follow us: @orkg_org

TIB

LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY



MORE INFORMATION

<u>www.orkg.org</u> Follow us: @orkg_org

Contact info@orkg.org



Creative Commons Attribution 3.0 Germany https://creativecommons.org/licenses/by/3.0/de/deed.en

