

LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY



Answering Scientific Questions by Organizing Research Contributions in the Open Research Knowledge Graph

Sören Auer





H

ttt

45555

LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY

11

TIB



LLMs Confabuate

DALL-E 3

12

You are a journalist. Write a summary of the developments around the widely publicized scandal involving Prof. Auer.

< 7/7 >

Send a message

> 5

Û

 \mathbf{V}

ChatGPT may produce inaccurate information about people, places, or facts. <u>ChatGPT September 25</u> <u>Version</u> In recent weeks, the academic community and the general public have been rocked by revelations surrounding the controversial Prof. Auer scandal. The scandal initially came to light when several anonymous sources leaked internal documents, pointing to potential misconduct in Prof. Auer's research. The allegations primarily revolved around data manipulation, plagiarism, and misappropriation of research funds.

Soon after the leak, multiple academic institutions distanced themselves from Prof. Auer, with some retracting his published works. Notably, three co-authors on various papers penned by Prof. Auer issued statements of their own, clarifying their roles and expressing their shock at the unfolding situation.

The academic community has since been divided, with some defending Prof. Auer, citing his previously unblemished record and contributions to the field. Others have called for a thorough and transparent investigation into the matter. Several students and postdocs under Prof. Auer's mentorship have also come forward, sharing their experiences – some in support and others detailing instances of alleged unethical behavior.

While Prof. Auer initially remained silent, he recently held a press conference vehemently denying all allegations. He stated that he has been the target of a coordinated attack aimed at tarnishing his reputation and vowed to take legal actions against those responsible for the leaks.

Currently, an independent committee is reviewing all of Prof. Auer's research. Their findings are eagerly anticipated and are expected to play a crucial role in shaping public opinion and determining the future of Prof. Auer's academic career. The scandal serves as a sobering reminder of the importance of integrity and transparency in academia.



How do LLMs work - Word Embeddings

By learning learning embeddings - the model latent relationships (e.g. male-female, junior-senior, nobel-profane)

Vocabulary: Man, woman, boy, girl, prince, princess, queen, king, monarch

	1	2	3	4	5	6	7	8	9		A
man	1	0	0	0	0	0	0	0	0		
woman	0	1	0	0	0	0	0	0	0		King
boy	0	0	1	0	0	0	0	0	0	Each word gets	
girl	0	0	0	1	0	0	0	0	0	a 1x9 vector	
prince	0	0	0	0	1	0	0	0	0	representation	
rincess	0	0	0	0	0	1	0	0	0		
queen	0	0	0	0	0	0	1	0	0		
king	0	0	0	0	0	0	0	1	0		/ ∕ ™
monarch	0	0	0	0	0	0	0	0	1		V - V



Multidimensional Semantic Feature Space



Page 6



LLMs are trained to predict next words

 \Rightarrow absolutely no representation of factual information



THE NEW YORK TIMES BESTSELLER

THINKING,



Generative AI thinks fast, but can not think slow

FAST AND SLOW

DANIEL

the Barth ?

KAHNEMAN





Connecting KG graphlets

with ML models

Both tend to hallucinate

To mitigate human brain deficiencies we have:

Peer-reviewed books, articles, curated databases, experiements, simulations,

. . .



KG graphlet authoring, curation, validation

Neuro-Symbolic AI with Knowledge Graphs





Vogt, D'Souza, Stocker, Auer: *Toward Representing Research Contributions in Scholarly Knowledge Graphs Using Knowledge Graph Cells.* JCDL 2020: 107-116





Knowledge Graphs – A definition

- Fabric of concept, class, property, relationships, entity descriptions
- Uses a knowledge representation formalism (typically RDF, RDF-Schema, OWL)
- Holistic knowledge (multi-domain, source, granularity):
 - instance data (ground truth),
 - open (e.g. DBpedia, WikiData), private (e.g. supply chain data), closed data (product models),
 - derived, aggregated data,
 - schema data (vocabularies, ontologies)
 - meta-data (e.g. provenance, versioning, documentation licensing)
 - comprehensive taxonomies to categorize entities
 - links between internal and external data
 - mappings to data stored in other systems and databases







LLMs vs. KGs

Large Language Models	Knowledge Graphs		
Very good at text processing: extraction, summarization, translation, generation	Very good at knowledge representation, integration		
Absolutely no concept of facts - still get often/sometimes factual information right	Human and machine readable - can be scrutinized/validated by humans		
Black box: might contain hidden biases, without domain specific training and fine- tuning often superficial	Transparent - information and their provenance can be traced		
Computationally very expensive to train, fine-tune and query (inference)	Scale well, but require manual integration/curation effort		

[1] Babaei, D'Souza, Auer: LLMs4OL: Large Language Models for Ontology Learning. ISWC 2023



KG-LLM Integration Opportunities

The fast take-up of KGs was hindered by complexity and manual effort \Rightarrow can be significantly reduced with the help of LLMs

Knov	vledge Graphs	Large Language Models		
• •	Represent a base of validated, trustworthy knowledge Help organize and integrate enterprise knowledge from various sources Provide input for enterprise and domain- specific training and fine-tuning of LLMs	 Help curating knowledge in the KG by suggesting and recommending Create mappings, queries etc. for the KG Can become a frontend for human interaction with the KG 		

[1] Ekaputra, Llugiqi, Sabou, Ekelhart, Paulheim, Breit, Revenko, Waltersdorfer, Farfar, Auer: Describing and Organizing Semantic Web and Machine Learning Systems in the SWeMLS-KG. ESWC 2023



Pathway to Artificial General Intelligence



Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, Hannaneh Hajishirzi https://arxiv.org/abs/2310.11511



How did information flows change in the digital era?



How does it work today?











The World of Publishing & Communication has profundely changed

- New means adapted to the new possibilities were developed, e.g. "zooming", dynamics
- Business models changed completely
- More focus on data, interlinking of data / services and search in the data
- Integration, **crowdsourcing, data curation** play an important role



What about Scholarly Communication?

TIB

Resources for Research and Development

Global spending on R&D has reached a record high of almost US\$ 1.7 trillion:

\$1,700,000,000,000

A large share of this money is not efficiently used.

Sources:

http://uis.unesco.org/apps/visualisations/research-anddevelopment-spending/ https://en.wikipedia.org/wiki/List_of_countries_by_rese

arch and development spending

¢	Country/Region	Expenditures on R&D (billions of US\$, PPP)	% of GDP PPP \$	Expenditures on R&D per capita (US\$ PPP)	Year +
	United States	612.714	3.1	1,866	2019
2	China China	514.798	2.2	368	2019
	• Japan	172.614	3.2	1,375	2019
	Germany	131.932	3.2	1,586	2019
;	South Korea	100.055	4.6	1,935	2019
;	France	63.658	2.2	944	2019
'	💶 India	58.691	0.65	43	2018
;	🗮 United Kingdom	51.702	1.8	762	2019
,	Taiwan	42.945	3.5	1,822	2019
0	Russia	38.549	1.0	263	2019
1	📀 Brazil	38.0	1.3	181	2017
2	Italy	33.840	1.4	561	2019
3	et Canada	26.636	1.5	700	2019
4	c. Turkey	24.827	1.1	297	2019
5	Mustralia	23.3	2.1	987	2014
6	C Spain	22.468	1.3	473	2019
7	Netherlands	20.167	2.2	1154	2019
8	Sweden	17.722	3.4	1,707	2019
9	• Israel	16.888	4.9	1,810	2019
0	Poland	16.086	1.3	420	2019
	 <td> Country/Region United States China Japan Japan Germany South Korea South Korea India India India Taiwan Taiwan Brazil Brazil Italy Italy Italy Italy Spain Spain Spain Sweden Sweden Israel Israel </td><td>Country/Region Expenditures on R&D (billions of US\$, PPP) Image: United States 612.714 Image: China 514.798 Image: China 514.798 Image: China 172.614 Image: China 100.055 Image: China 100.055 Image: China 58.691 Image: China 58.691 Image: China 58.691 Image: China 51.702 Image: China 24.827</td><td>Country/Region Expenditures on R&D (billions of US\$, PPP) % of GDP PPP + Image: United States 612.714 3.1 Image: China 514.798 2.2 Image: China 514.798 2.2 Image: China 514.798 2.2 Image: China 514.798 2.2 Image: China 172.614 3.2 Image: Germany 131.932 3.2 Image: South Korea 100.055 4.6 Image: South Korea 3.650 1.3 Image: South Korea 3.8549 1.0 Image: South Korea 3.8.0 1.3 <t< td=""><td>Country/Region Expenditures on R&D (billions of US\$, PPP) % of GDP PPP Expenditures on R&D per capita (US\$ PPP) Image: United States 612.714 3.1 1,866 Image: China 514.798 2.2 368 Image: China 514.798 2.2 368 Image: China 172.614 3.2 1,375 Image: Germany 131.932 3.2 1,586 Image: Germany 51.702 1.8 762 Image: Germany 3.8.691 0.655 1.822 Image: Gerazil <</td></t<></td>	 Country/Region United States China Japan Japan Germany South Korea South Korea India India India Taiwan Taiwan Brazil Brazil Italy Italy Italy Italy Spain Spain Spain Sweden Sweden Israel Israel 	Country/Region Expenditures on R&D (billions of US\$, PPP) Image: United States 612.714 Image: China 514.798 Image: China 514.798 Image: China 172.614 Image: China 100.055 Image: China 100.055 Image: China 58.691 Image: China 58.691 Image: China 58.691 Image: China 51.702 Image: China 24.827	Country/Region Expenditures on R&D (billions of US\$, PPP) % of GDP PPP + Image: United States 612.714 3.1 Image: China 514.798 2.2 Image: China 514.798 2.2 Image: China 514.798 2.2 Image: China 514.798 2.2 Image: China 172.614 3.2 Image: Germany 131.932 3.2 Image: South Korea 100.055 4.6 Image: South Korea 3.650 1.3 Image: South Korea 3.8549 1.0 Image: South Korea 3.8.0 1.3 <t< td=""><td>Country/Region Expenditures on R&D (billions of US\$, PPP) % of GDP PPP Expenditures on R&D per capita (US\$ PPP) Image: United States 612.714 3.1 1,866 Image: China 514.798 2.2 368 Image: China 514.798 2.2 368 Image: China 172.614 3.2 1,375 Image: Germany 131.932 3.2 1,586 Image: Germany 51.702 1.8 762 Image: Germany 3.8.691 0.655 1.822 Image: Gerazil <</td></t<>	Country/Region Expenditures on R&D (billions of US\$, PPP) % of GDP PPP Expenditures on R&D per capita (US\$ PPP) Image: United States 612.714 3.1 1,866 Image: China 514.798 2.2 368 Image: China 514.798 2.2 368 Image: China 172.614 3.2 1,375 Image: Germany 131.932 3.2 1,586 Image: Germany 51.702 1.8 762 Image: Germany 3.8.691 0.655 1.822 Image: Gerazil <



Scholarly Communication has not changed (much) 17th century



19th century

THE INTELLECTUAL OBSERVER.

JANUARY, 1865.

CELESTIAL CHEMISTRY, AND THE PHYSICAL CONSTITUTION OF THE STARS AND NEBULÆ.

BY THOMAS W. BURE, F.E.A.S., F.C.S.

(With a Coloured Plate.)

Few things are more remarkable in the present aspect of science than the manner in which its various departments come into contact one with another, thus aiding the student in a way quite unlooked for, and throwing light upon the subject of research from a quarter whence it was least expected. As when stones are thrown into water, so the circle of each science at first seems to be totally distinct from all the others. but gradually these separate circles enlarge and widen, until they intersect and produce larger circles and wider generalizations in the increasing domain of human knowledge. Thus, chemistry was, in the time of Davy, furnished with a new and powerful analytical agent in the shape of voltaic electricity, and the same agency, which is itself evoked by chemical action, has given us the long series of discoveries in electro magnetism, culminating in the splendid practical application of the electric telegraph. So, too, photography, which is essentially chemical in its nature, has been of the greatest service to the physicist in furnishing him with a constant and unerring record of the indications of his barometer, thermometer, and magnetic instruments, and has even come to the assistance of the astronomer and depicted for him the changing appearances of the moon's surface, the spots on the sun, and the fleeting phenomena of a solar eclinse

Quite recently the application of some of the phenomena of light to the discrimination of the chemical constitution of bodies, or spectrum analysis as it is called, is a discovery of the highest order and most extraordinary utility to the chemist, while its extension to the discovery of the cause of the dark lines VOL. VI.-NO. VI. C C

20th century

P. BAXENDALE, Editor

A Relational Model of Data for Large Shared Data Banks

E E Conn IBM Research Laboratory, San Jose, California

Information Retrieval

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural arowth in the types of stored information. Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on n-ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

KEY WORDS AND PHRASES: date bank, data base, data structure, data organization, hierarchies of data, networks of data, relations, derivability, redundancy, consistency, composition, join, retrieval language, predicate cokulus, security, data integrity CR CATEGORES: 3,70, 3,73, 3,75, 4,20, 4,22, 4,29

I. Relational Model and Normal Form

1.1. INTRODUCTION

This paper is concerned with the application of elementary relation theory to systems which provide shared access to large banks of formatted data. Except for a naper by Childs [1], the principal application of relations to data to participate in one ordering only, others permitting each

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 superim posing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence he tween programs on the one hand and machine represents tion and organization of data on the other.

A further advantage of the relational view is that it forms a sound basis for treating derivability, redundancy and consistency of relations-these are discussed in Section 2. The network model, on the other hand, has spawned a number of confusions, not the least of which is mistaking the derivation of connections for the derivation of rela tions (see remarks in Section 2 on the "connection trap") Finally, the relational view permits a clearer evaluation of the scope and logical limitations of present formatted data systems, and also the relative merits (from a logical standpoint) of competing representations of data within a single system. Examples of this clearer perspective are cited in various parts of this paper. Implementations of systems to support the relational model are not discussed

1.2. DATA DEPENDENCIES IN PRIMERY SYSTEMS The provision of data description tables in recently developed information systems represents a major advance toward the goal of data independence [5, 6, 7]. Such tables facilitate changing certain characteristics of the data representation stored in a data bank. However, the variety of data representation characteristics which can be changed without logically impairing some application programs is still quite limited. Further, the model of data with which users interact is still cluttered with representational properties, 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 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

21th century

AGDISTIS - Graph-Based Disambiguation of Named Entities using Linked Data

Ricardo Usbeck^{1,2}, Axel-Cyrille Ngonga Ngomo¹, Michael Röder^{1,2}, Daniel Gerber¹, Sandro Athaide Coelho³, Sören Auer⁴, and Andreas Both²

¹ University of Leipzig, Germany , ² R & D, Unister GmbH, Germany, ³ Federal University of Juiz de Fora, Brazil, ⁴ University of Bonn & Fraunhofer IAIS, Germany email: {usbeck|ngonga}@informatik.uni-leipzig.de

Abstract. Over the last decades, several billion Web pages have been made available on the Web. The ongoing transition from the current Web of unstructured data to the Web of Data yet requires scalable and accurate approaches for the extraction of structured data in RDF (Resource Description Framework) from these websites. One of the key steps towards extracting RDF from text is the disambiguation of named entities. While several approaches aim to tackle this problem, they still achieve poor accuracy. We address this drawback by presenting AGDIS-TIS, a novel knowledge-base-agnostic approach for named entity disambiguation. Our approach combines the Hypertext-Induced Topic Search (HITS) algorithm with label expansion strategies and string similarity measures. Based on this combination, AGDISTIS can efficiently detect the correct URIs for a given set of named entities within an input text. We evaluate our approach on eight different datasets against state-of-theart named entity disambiguation frameworks. Our results indicate that we outperform the state-of-the-art approach by up to 29% F-measure.

1 Introduction

The vision behind the Web of Data is to provide a new machine-readable layer to the Web where the content of Web pages is annotated with structured data (e.g., RDFa [1]). However, the Web in its current form is made up of at least 15 billion Web pages.¹ Most of these websites are unstructured in nature. Realizing the vision of a usable and up-to-date Web of Data thus requires scalable and accurate natural-language-processing approaches that allow extracting RDF from such unstructured data. Three tasks play a central role when extracting RDF from unstructured data: named entity recognition (NER), named entity disambiguation (NED), also known as entity linking [16], and relation extraction (RE). For the first sentence of Example 1, an accurate named entity recognition approach would return the strings Barack Obama and Washington, D.C., A high-quality DBpedia-based named entity disambiguation (NED) approach would use these already recognized named entities and map the strings

Data gathered from http://www.worldwidewebsize.com/ on January 4th, 2014.



The scientific method is at stake!



- Static, unstructured documents
- Potential of AI unused



ambiguation of Act Data 2. Research is drowning in a flood of publications

>50M overall

- >3M published every year
- ~70% not cited



DATA LAKE

3. Peer-review & reproducibility are severely impaired

- Major share of scientific publications not reproducible
- Quality of peer-review deteriorates



4. Despite recent attention to FAIR research data methods ...

- do not scale,
- are insufficiently automatized and lack user integration,
- lack explainability, security and support for causal inference



We need to rethink the way how research is represented and communicated

Challenges we are facing:

Digitalisation of Science	Monopolisation by commercial actors	Reproducibility Crisis	Proliferation of publications	Deficiency of Peer Review
 Data integration and analysis Digital collaboration 	 Publisher look-in effects Maximization of profits ^[1] 	 Majority of experiments are hard or not reproducible ^[2] 	 Publication output doubled within a decade continues to rise ^[3] 	 Deteriorating quality ^[4] Predatory publishing

[1] http://thecostofknowledge.com, https://www.projekt-deal.de

- [2] M. Baker: 1,500 scientists lift the lid on reproducibility, Nature, 2016.
- [3] <u>Science and Engineering Publication Output Trends</u>, National Science Foundation, 2018.
- [4] J. Couzin-Frankel: Secretive and Subjective, Peer Review Proves Resistant to Study. Science, 2013.



Root Cause – Deficiency of Scholarly Communication?

Li	аск от		
	Transparency	Integratability	Machine assistance
i	information is hidden in text	fitting different research results together	unstructured content is hard to process
	Identifyability	Collaboration	Overview
	of concepts beyond metadata	one brain barrier	Scientists look for the needle in the haystack

≡	Google Scholar	CRISPR Q
٠	Artikel	Untefähr 238.000 Ergebnisse (0,04 Sek.)
	Beliebige Zeit Seit 2019 Seit 2018 Seit 2015 Zeitraum wä	Multiplex genome engineering using CRISPR/Cas systems <u>Cong (A Ran, D Cox, S Lin, R Barretto, 2013 - science.sciencemag.org</u> <u>Functional elucidation of causal genetic variants and elements requires precise genome</u> ype II prokaryotic CRISPR (cluster h for CRISPR:
	> 238 Nach Releva sortieren Nach Datum sortieren	Abnliche Artikel Alle 45 Version How good is CRISPR (wrt. precision, safety, cost)? Clustered regularly interspaced short palindromic repeats (CP
	Beliebige Sprache Seiten auf Deutsch	of the genomes of most Bacteria and Archaea and are though bacteriophages. We found that, after viral challenge, bacteria ☆ ワワ Zitiert von: 3418 Ähnliche Artikel Alle 29 Version What specifics has genome editing with insects?
	 Patente einschließen Zitate einschließen Alert erstellen 	Genome engineering using the CRISPR-Cas9 <u>FA Ran, PD Hsu, J Wright</u> , V Agarwala, DA Scott Nature Targeted nucleases are powerful tools for mediating genome a The RNA-guided Cas9 nuclease from the microbial clustered palindromic repeats (CRISPR) adaptive immune system can ☆ 99 Zitiert von: 3650 Ähnliche Artikel Alle 19 Version



How can we fix it?



A SCIENTIST OF THE FUTURE RECORDS EXPERIMENTS WITH A TINY CAMERA FITTED WITH UNIVERSAL-FOCUS LENS. THE SMALL SQUARE IN THE EYEGLASS AT THE LEFT SIGHTS THE OBJEC

AS WE MAY THINK

A TOP U.S. SCIENTIST FORESEES A POSSIBLE FUTURE WORLD IN WHICH MAN-MADE MACHINES WILL START TO THINK









Memex (Memory Extender)

MEMEX in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference.

Concepts

Overarching Concepts

- Research problems
- Definitions
- Research approaches
- Methods

Artefacts

- Publications
- Data
- Software
- Image/Audio/Video
- Knowledge Graphs / Ontologies

Domain specific Concepts

Mathematics

- Definitions
- Theorems
- Proofs
- Methods
-

• Experiments

Physics

- Data
- Models

• ...

Chemistry

- Substances
- Structures
- Reactions

•

Computer Science

- Concepts
- Implementations
- Evaluations

• ...

Technology

- Standards
- Processes
- Elements
- Units, Sensor data

Architecture

- Regulations
- Elements
- Models

•



The Open Research Knowledge Graph



TIB

Page 30

Leibniz Universität Hannover

	🖄 orkg 🕛	apers Tools 🗸 About 🗗	Search	. Q Add	paper Sign in	
Properties	Estimation of the epidemic properties of the 2019 novel coronavirus: A mathematical modeling study Contribution 1 - 2020	Estimation of the epidemic properties of the 2019 novel coronavirus: A mathematical modeling study Contribution 2 - 2020	Estimation of the epidemic properties of the 2019 novel coronavirus: A mathematical modeling study Contribution 3 - 2020	Transmission potential of COVID-19 in Iran COVID-19 in Iran Contribution 1 - 2020	Transmission potential of COVID-19 in Iran CovID-19 in Iran Contribution 2 - 2020	Estimating the generation interval for COVID-19 based on symptom onset data Contribution 1 - 2020
Has research problem	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number
Location	Wuhan City, China	Wuhan City, China	Wuhan City, China	Iran	Iran	Singapore
Study date	2020-01-10/2020-01-23	2020-01-23/2020-02-08	2020-01-10/2020-02-08	2020-02-19/2020-02-29	2020-02-19/2020-02-29	2020-01-21/2020-02-26
R0 estimates (average)	4.38	3.41	3.39	3.6	3.58	1.27
95% confidence interval	3.63-5.13	3.16-3.65	3.09-3.70	3.2-4.2	1.29-8.46	1.19-1.36
Method	a weighted average of Exponential growth, Maximum likelihood, Sequential Bayesian, Time-dependent reproduction numbers, and SEIR model basic reproduction numbers by calculating weights from a Poisson loss function	a weighted average of Exponential growth, Maximum likelihood, Sequential Bayesian, Time-dependent reproduction numbers, and SEIR model basic reproduction numbers by calculating weights from a Poisson loss function	a weighted average of Exponential growth, Maximum likelihood, Sequential Bayesian, Time-dependent reproduction numbers, and SEIR model basic reproduction numbers by calculating weights from a Poisson loss function	generalized growth model	based on the calculation of the epidemic's doubling times: estimated epidemic doubling time of 1.20 (95% Cl, 1.05, 1.44) days	generation interval

ORKG Papers Tools V	About 🗹	Search	Q Add	paper Sign in
View paper			View paper	Graph view 🖍 Edit
Transmission interval estim	nates suggest pre-symptomat en Tindale Michelle Coombe Jessic Lee Louxin Zhang Dongxuan Chen	tic spread of COVI	D-19 lock Wing Y	'in Venus Lau
			DOI: 1	0.1101/2020.03.03.20029983
Contribution 1 Contribution 2 Research problems COVID-19 reproductive number Contribution data		Add to comparison		Share this paper: 🕜 У
95% Confidence interval	1.45-2.48			
Location	Singapore			
Mean incubation period	7.1 (6.13, 8.25) days			
Mean serial interval	4.56 (2.69, 6.42) days			
R0 estimates (average)	1.97			
Study date	2020-01-19/2020-02-26			

ORKG Papers Tools V	About 🗹	Search	٩	Add paper	Sign in
Edit mode Every change you make is auto	matically saved		/iew paper	T Graph view	× Stop editing
 March 2020 E Virology Lau Manu Saraswat Yen-Hsiang Bria Edit data 	uren Tindale 🔹 Michelle Coombe n Lee 🔒 Louxin Zhang 🔹 Don	💄 Jessica Stockdale 🔹 Em agxuan Chen 🔹 Jacco Wallinga	ma Garlock	Wing Yin Venus Lau e Colijn DOI: 10.1101/2020.0	1
Contribution 1 🎤 📋 Contribu	tion 2 +			Share this	paper:
Research problems		Add to compariso	on		
Contribution data					
95% Confidence interval	1.45-2.48 +				
Location	Singapore +				
Mean incubation period	7.1 (6.13, 8.25) days + Edit value				
Mean serial interval	4.56 (2.69, 6.42) days 🖉 💼				











Brief introduction

ORKG Ask

Find research you are actually looking for







Co-funded by the Horizon 2020 programme of the European Union



IILeibniz02Universität04Hannover

- **Mission**: empower researchers to help them **find** and **explore** research articles
- Architecture of ORKG Ask



- **Mission**: empower researchers to help them **find** and **explore** research articles
- Architecture of ORKG Ask



- **Mission**: empower researchers to help them **find** and **explore** research articles
- Architecture of ORKG Ask



Why ORKG Ask is different from *Google Scholar, Semantic Scholar*, but also from *Typeset.io, Elicit,* etc.

Knowledge graph integration

- Stores LLM responses as structured knowledge in the graph
- Have more targeted search (e.g. population > 20)
- Let's users validate the knowledge

and **explore** research

Knowledge Graphs

(KGs)

Some more information

- Generated response tables from PaperScout can be exported to the ORKG easily → Starting point of new comparisons
- **BibTeX is the main format** of PaperScout: existing bibliographies can be easily imported to get started, and generated data can be exported
- Easy to get started, possibility to **transfer** users to the ORKG



Application overview

- **1.** Homepage \rightarrow Start a search
- **2. Results** \rightarrow Tabular overview of results
- **3. Paper view** \rightarrow More detailed view of papers
- **4.** My library \rightarrow Organized saved searches and papers



Find research you are actually looking for

Type your question...

Getting started

- How does climate change impact biodiversity?
- · Why are aging Covid patients more susceptible to severe complications?
- How does social media affect the college selection process?
- What are the interesting theories about dark matter and dark energy?
- What is the significance of higher-dimensional algebra?

PaperScout is a scholarly search and exploration system powered by LLMs and Knowledge Graphs

Q

255.275.131 Papers

Contact • Data Protection • Accessibility • Imprint • Sitemap



Filters
Year
Author (1)
Research field
Туре (1)
Language (3)
Publisher (3)
Search
 'Springer Science and Business Media LLC' (25545) 'Baishideng Publishing Group Inc.'
(20413)
 'Wiley' (14238) 'Elsevier RV' (37179)
MDPI AG' (35246)
Universidad Cesar Vallejo' (25696)

'Frontiers Media SA' (11468)

Reset

Filters

- Filters are an important aspects of more fine-grained search
- ORKG ASK relies on available metadata from CORE, in the future this should **come from** the graph (ORKG, OpenAlex, DataCite etc.)

people, ideas, skills, images, and	on interactions between Italians and	enriching the colonial knowle
objects.		debate.





TL;DR

The planetary boundary concept, created in 2009, identifies critical environmental limits. Steffen et al. update it, emphasizing three boundaries, including climate change, with profound global implications and a refined regional application.

Abstract

The planetary boundary (PB) concept, introduced in 2009, aimed to define the environmental limits within which humanity can safely operate. This approach has proved influential in global sustainability policy development. Steffen et al. provide an updated and extended analysis of the PB framework. Of the original nine proposed boundaries, they identify three (including climate change) that might push the Earth system into a new state if crossed and that also have a pervasive influence on the remaining boundaries. They also develop the PB framework so that it can be applied usefully in a regional context. Read more

Unstructured data	Structured dat	ta	
Research questions	Topics	earth system resilience	✓ 80%
 How small-scale regime shifts can propagate across scales and possibly lead to global-level transitions 		planetary boundaries holocene epoch impact	✓ 66%✓ 65%

Planetar	y boundaries: Guiding human development on a changing planet	
Will Steffen • I	Katherine Richardson • Johan Rockström • Sarah E. Cornell • Ingo Fetzer • Elena M. Bennett • Reinette Biggs • Stephen R. Carpenter • Wim de Vries • 12 m	
	PDF Bookmark 66 Cite Add	d to search
15 Januz	date DOI Journal Publisher DI ISSN	
\$ Func		
\$ Fund Stordale	Paper view	
\$ Fund Stordak	Paper view	
\$ Fund Stordak TL;DR The plan	 Paper view Shows some additional information of a paper 	
S Fund Stordals TL;DR The plat profoun	Paper viewShows some additional information of a paper	
\$ Fund Stordald TL;DR The plat profoun	 Paper view Shows some additional information of a paper Not as important as the search results, but good for 	
\$ Fund Stordale TL;DR The plat profoun Abstra The plat global s	 Paper view Shows some additional information of a paper Not as important as the search results, but good for 	ntial
\$ Fund Stordald TL;DR The plat profoun Abstra The plat global s (includir PB fram	 Paper view Shows some additional information of a paper Not as important as the search results, but good for SEO reasons 	n intial three op th
\$ Fund Stordals TL;DR The plat profoun Abstra The plat global s (includir PB fram	 Paper view Shows some additional information of a paper Not as important as the search results, but good for SEO reasons 	n ential three op th

	Search	My library Saved searches	Jane Doe 👻
Paperscout		With 4 selected	+ Add
Collections		Geological and climatic influences on mountain biodiversity 66	
My favorites	/ 1	🚢 Alexandre Antonelli • W. Daniel Kissling • Suzette G. A. Flantua • 8 more 🍵 October 20	18 🚳 10.5281/zenodo.1341999 PDF
Promising research pape Last category	r	Cascading, interactive, and indirect effects of climate change on aquatic comm Susanne Menden-Deuer • Julia C. Mullarney • Maarten Boersma • 3 more July 2023	nunities, habitats, and ecosystems 66 10.1002/Ino.12384 PDF
Add new category		Improving the forecast for biodiversity under climate change 66 M. C. Urban • G. Bocedi • A. P. Hendry • JB. Mihoub • 4 more September 2016	10.1126/science.aad8466 PDF
Sort By rele	evance 🔻	Planetary boundaries: Guiding human development on a changing planet 66 Will Steffen • Katherine Richardson • Johan Rockström • 12 more January 2015	10.1126/science.1259855 PDF
		1 2 3 4 5 56 >> 21-30 of 423432 papers	Items per page 25 👻



ASK Architecture



Prompts

Prompt A

You are an analysis-support bot that operates on scholarly documents and follows instructions. You will get as an input: the research paper content and a set of properties/criteria to look for. You will extract the values corresponding to the list of provided predicates. Limit the values to only the content without prefixing it with the paper contains or the text says. Your output should ALWAYS be like this:

<extraction><key>PROPERTY</key><value>VALUE</value></ extraction>

properties {properties}.

paper content {content}

Prompt B

You are an analysis-support bot of scholarly documents that follows instructions. You will generate a comprehensive answer to the given research question (but no more than three/four sentences) solely based on the content provided. Cite the number of the content referenced for each claim like this: [1] for a single reference or [2][3] for multiple references. Emphasize brevity, focusing on essential details and omitting unnecessary information. Avoid adding the question in the answer and Do not include any notes or comments.

Research Question: {question} # Abstracts: {abstracts}

Answer with inline-citations as [#]:



LEIBNIZ-INFORMATIONSZENTRUM TECHNIK UND NATURWISSENSCHAFTEN UNIVERSITÄTSBIBLIOTHEK

HybridAI: Establish true Human-Machine Collaboration

[1] Allard Oelen, Sören Auer: Leveraging Large Language Models for Realizing Truly Intelligent User Interfaces. CHI 2024.

ORKG | Knowledge transformation

To create a scholarly knowledge graph, a **transformation** from unstructured to structured knowledge should happen



Unstructured knowledge

Structured knowledge

Can we use Natural Language Processing (NLP) for the transformation process?

ORKG | Knowledge transformation

Can we use Natural Language Processing (NLP) for the transformation process?

 NLP techniques are not sufficiently accurate to perform this task autonomously

• But we can **intertwine machine intelligence with human intelligence** to get a synergy → the best of both worlds!

Gradations of automation



Better scalable

Gradations of automation



Machine-in-the-loop | Add paper wizard

General	1 General Research field		3 Contributions	The third step is the description of
pecify research cont	ibutions 📀		Abstract annotator	contribution data
Contribution 1 🇪 (+)	60000	Help ☐ Preferences	Research problem	Machine-in-the-
+ Add property			A Fast Learning Algorithm Method	Іоор
+ Research problem		Problem Research Material	Complementary priors	
+ Method + Result		Methods Results		

Machine-in-the-loop | Smart suggestions



Launch an LLM for assistance directly from the curation interface

Abstract annotator

Info: we automatically annotated the abstract for you. Please remove any incorrect annotations

Annotation labels 🔞

Material 🔽



We show how to use complementary priors to eliminate the explaining-away effects that make inference difficult in **densely connected belief nets** that have **many hidden layers**. Using complementary priors, we derive a fast, greedy algorithm that can learn **deep**, **directed belief networks one layer at a time**, provided the **top two layers** form an undirected associative memory. The fast, greedy algorithm is used to initialize a slower learning procedure that fine-tunes the weights using a **contrastive version of the wake-sleep algorithm**. After fine-tuning, a **network** with three hidden layers forms a very good generative model of the joint distribution of **handwritten digit images** and their labels. This generative model gives better digit classification than the best discriminative learning algorithms. The **low-dimensional manifolds** on which the digits lie are modeled by long ravines in the free-energy landscape of the top-level associative memory, and it is easy to explore these ravines by using the directed connections to display what the **associative memory** has in mind.



×



- Annotations can be added or removed
- A confidence slider hides suggestions with a low score

The Team

Head

Auer

Group Leaders





Dr. Jennifer Dr. Lars Vogt D'Souza

PostDocs



Dr. Ildar **Baimuratov**



erc

Dr. Gollam Dr. Oliver Karras Rabby

Dr. Anna-Lena

Lorenz





TIB

Dr. Vinodh llangovan

Lena John

Doctoral Researchers

Stocker



Software Development

Zhiyin Tan



Farhana Keya Golsa Heidari Mahsa Shamsabadi



Muhammad Haris









Vlad Nechakhin Hassan Hussein Hamed Babaei Salomon Kabongo Tim Wittenborg **Administration & Support**





Dr. Allard Manuel Prinz Oelen



Dr. Yaser Eddine Farfar Jaradeh

Aftab



Akhilesh Vyas

Marcel Konrad

Regina Hulm

Irmtraut Schröder Simone Matern Seite 63



European

Research

Community Management

Council

Eccenca Corporate Memory – Neuro-Symbolic AI for Enterprises



mastering complexity



Conclusions

- We need to reinvent scholarly communication
- Knowledge Graphs are perfectly suited to capture research contributions in a structured and semantic way making them human and machine interpretable
- With our Open Research Knowledge Graph .
 initiative we aim to establish a registry for research contributions
- Curation and synergistic combination of human, expert and machine intelligence is a challenge

Stay tuned

- https://tib.eu
- Consider creating an ORKG observatory for your domain
- Mailinglist/group: <u>https://groups.google.com/forum/#!forum/orkg</u>
- Open Research Knowledge Graph: <u>https://orkg.org</u>
- ERC Consolidator Grant ScienceGRAPH on the topic



European Research Council



LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY





https://de.linkedin.com/in/soerenauer

https://mstdn.social/@soeren_auer

Prof. Dr. Sören Auer

in

0

l l Leibniz *i o 2* Universität *i o 4* Hannover

TIB & Leibniz University of Hannover auer@tib.eu

