

Erschließung wissenschaftlicher Literatur mit dem Open Research Knowledge Graph

Sören Auer, Irina Sens, Markus Stocker

Wissenschaftliche Bibliotheken müssen sich auf die geänderten Anforderungen der Wissenschaft einstellen. Die Digitalisierung der wissenschaftlichen Arbeitsweisen, Prozesse und Publikationsformen ist eine zentrale Herausforderung. Die Methoden des wissenschaftlichen Informationsaustausches (engl. Scholarly Communication) sind seit Jahrhunderten mehr oder weniger unverändert statische Textartikel. Diese können zwar heute elektronisch als PDF oder HTML beliebig vervielfältigt und schnell über das Internet abgerufen werden, aber an der grundsätzlichen Repräsentation als unstrukturierte, statische Artikel hat sich nichts Grundlegendes geändert.

Digitale Enzyklopädien und Landkarten

Andere Informationsdomänen haben sich dagegen grundlegend gewandelt und völlig neue digitale Repräsentationsformen entwickelt. Die einzige verbliebene Enzyklopädie zum Beispiel ist Wikipedia, die nicht einfach eine digitale PDF-Kopie einer analogen Enzyklopädie ist, sondern gänzlich neue Erarbeitungs-, Repräsentations- und Organisationsformen für enzyklopädische Inhalte realisiert hat und damit zum Beispiel die Realisierung von Enzyklopädieversionen für hunderte Sprachen und verschiedenste Zielgruppen in einer ganz neuen Tiefe und Breite ermöglicht. Weitere Beispiele für gänzlich digitale Informationsdienste sind unter anderem

- digitale Kartenanwendungen (wie Google Maps oder OpenStreetMaps), die physische Straßen- und Landkarten inzwischen fast ganz ersetzt haben,
- Online-Shops und E-Commerce-Anwendungen mit gänzlich neuen Such-, Bewertungs- und Datenvernetzungsfunktionen statt der klassischen Versandhauskataloge,
- digitale Kommunikationsanwendungen, welche zum Beispiel Telefonbücher obsolet gemacht haben.

Alle diese Beispiele verdeutlichen, dass analoge Repräsentationsformen (Bücher und Dokumente) nicht einfach nur als PDF „pseudodigitalisiert“ wurden, sondern komplett neu als digital-geborene Anwendungen realisiert wurden. Solche „digital-first“ Anwendungen basieren auf einer grundlegend neuen,

Die Weitergabe von Wissen hat sich seit vielen hundert Jahren nicht grundlegend verändert: Sie erfolgt in der Regel dokumentenbasiert – früher als klassischer Aufsatz auf Papier gedruckt, heute als PDF. Bei jährlich rund 2,5 Millionen neuen Forschungsbeiträgen ertrinken die Forschenden in einer Flut pseudodigitalisierter PDF-Publikationen. Die Folge: Die Forschung wird ernsthaft geschwächt. Denn viele Forschungsergebnisse können durch andere nicht reproduziert werden, es gibt mehr und mehr Redundanzen und das Meer von Publikationen ist unübersichtlich. Deshalb denkt die TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften Wissenskommunikation neu: Statt auf statische PDF-Artikel setzt die TIB auf Wissensgraphen. Sie arbeitet daran, Wissen unterschiedlichster Form – Texte, Bilder, Grafiken, Audio- und Video-Dateien, 3D-Modelle und vieles mehr – intuitiv mithilfe dynamischer Wissensgraphen zu vernetzen. Der Wissensgraph soll verschiedene Forschungsideen, -ansätze, -methoden und -ergebnisse maschinenlesbar darstellen, sodass völlig neue Zusammenhänge von Wissen zutage treten und zur Lösung globaler Probleme beitragen könnten. Die großen gesellschaftlichen Herausforderungen verlangen Interdisziplinarität und das Zusammenfügen von Erkenntnis-Einzelteilen. Mit dem Wissensgraphen kann das gelingen und der Fluss wissenschaftlicher Erkenntnisse revolutioniert werden.

strukturierten und datenorientierten Informationsorganisation und ermöglichen damit gänzlich neue Unterstützung durch intelligente Such- und Filterfunktionen, die Integration vielfältiger Zusatzinformationen und Dienste, Crowdsourcing usw.

Pseudodigitalisierter Wissensaustausch erschwert wissenschaftliche Arbeit

Der nur pseudodigitalisierte wissenschaftliche Informationsaustausch dagegen ist in etwa damit vergleichbar, dass wir Produkte aus per E-Mail versendeten PDF-Katalogen herausuchen oder den Weg in den Urlaub auf einer PDF-Straßenkarte finden müssten. Dies verursacht für die wissenschaftliche Arbeit große Probleme:

- Wir sind mit einer stetig wachsenden Fülle wissenschaftlicher Publikationen konfrontiert, die natürlich mit digitalen Werkzeugen schneller erstellt werden können. Im Bereich Technik und Naturwissenschaften beispielsweise hat sich die Zahl der Publikationen pro Jahr innerhalb einer Dekade nahezu verdoppelt.

- Durch das dramatische Wachstum an Publikationen erfolgt oft keine qualitativ ausreichend gute Begutachtung (Peer-Review): Einerseits gibt es zu wenige qualifizierte Gutachterinnen und Gutachter mit ausreichend Zeit, andererseits ist die Begutachtung, also der Beitrag zum Stand der Forschung, immer schwieriger zu ermitteln.
- Die Mehrheit wissenschaftlicher Publikationen ist durch Fachkolleginnen und -kollegen (und oft sogar durch die Autorinnen und Autoren selbst) nicht reproduzierbar. Ein zentraler Grund dafür ist die unstrukturierte Darstellung in statischen PDF-Artikeln, wo unter Umständen wichtige Informationen fehlen.
- Verschiedene Forschungsansätze sind aufgrund der unstrukturierten Darstellung kaum mehr vergleichbar, was die Ermittlung des Stands der Forschung extrem erschwert, insbesondere für jüngere oder interdisziplinär Forschende.
- Die unstrukturierte Darstellung von Forschungsergebnissen ermöglicht keine oder nur sehr unzureichende maschinelle Unterstützung. Forschungsbeiträge können nicht effektiv durchsucht, gefiltert oder visualisiert werden. Assistenzsysteme, wie sie mit Google Now oder Alexa bereits für Alltagssituationen verfügbar sind, sind für die Bewältigung der wissenschaftlichen Informationsflut derzeit undenkbar.

Natürlich gibt es eine Vielzahl von Initiativen, welche sich zum Ziel gesetzt haben, diese Probleme zu adressieren. Wir haben allerdings den Eindruck, dass oft eher an den Symptomen gearbeitet wird, anstatt die fundamentalen Probleme zu adressieren. Teilweise beruhen sie auf Fehlannahmen.

Die erste Fehleinschätzung ist: Text- und Datamining kann das Problem der Erschließung und Exploration wissenschaftlicher Artikel lösen. Voll-automatisiertes Text-Mining und Methoden der natürlichen Sprachverarbeitung allein werden nicht die ausreichende Genauigkeit liefern, damit die extrahierten Informationen tatsächlich hilfreich sind. Solche Methoden erreichen für die Erkennung von Entitäten (Named Entity Recognition) oft nur 50 bis 60 Prozent Genauigkeit und Abdeckung. Für die Extraktion von Beziehungen (Relation Extraction), die essenziell für verbesserte maschinelle Unterstützung ist, liegen die Ergebnisse kaum bei einem Drittel gefundener oder korrekter Aussagen, was deren Anwendung im Prinzip ausschließt.

Eine zweite Fehleinschätzung ist, dass die Fülle wissenschaftlicher Informationen nur mit voll-automatisierten Verfahren und eventuell Maschinellen Lernen zu bewältigen ist. Maschinelle Lernverfahren

funktionieren nur dort, wo ausreichend Trainingsdaten vorhanden sind. Dies ist bei der strukturierten Extraktion wissenschaftlicher Ergebnisse aus unstrukturierten Artikeln nicht der Fall und auch auf absehbare Zeit nicht möglich. Mit Crowd- oder besser Expert sourcing könnten wir die strukturierte Organisation wissenschaftlicher Informationen – eventuell unterstützt durch maschinelle Verfahren – hingegen bewältigen. Dafür ist initial weder eine vollständige Verarbeitung der wissenschaftlichen Literatur noch eine Beteiligung eines großen Prozentsatzes der Wissenschaftlerinnen und Wissenschaftler notwendig. Es würde zunächst ausreichen, wenn sich nur wenige Wissenschaftlerinnen und Wissenschaftler an der Kuratierung und Organisation beteiligen und unter Umständen auch nur wenige Forschungsprobleme und Spezialgebiete abgedeckt sind. Das würde für potenzialträchtige Anwendungen in diesen Gebieten ausreichen und könnte dazu führen, später einen Netzwerkeffekt zu etablieren, der breitere Wissenschaftsbereiche erfasst. Ein gutes Beispiel, wie so etwas funktionieren kann, ist OpenStreetMaps: Eine Crowdsourcing-Anwendung, mit der wenige tausend Mitstreiterinnen und Mitstreiter eine offene Weltkarte geschaffen haben, die in vielen Bereichen wesentlich detaillierter als kommerzielle Angebote ist und durch die innovative Datenorganisation für vielfältigste Anwendungen vom Katastrophenschutz über Navigation bis hin zu behindertengerechter Mobilität und Fahrradkarten genutzt wird.

Zusammenfassend lässt sich also sagen, dass wir eine neue Form der Repräsentation wissenschaftlicher Erkenntnisse benötigen, die hochgradig semantisch strukturiert ist und die großskalige Zusammenarbeit von Fachwissenschaftlern, Wissensingenieuren, Informationswissenschaftlern und Bibliothekaren sowie Nutzerinnen und Nutzern ermöglicht und gleichzeitig einen evolutionären Übergang von den bisherigen wissenschaftlichen Publikations- und Incentivierungssystemen ermöglicht. Genau auf diese Anforderungen zielt der von der TIB und Partnern entwickelte Open Research Knowledge Graph, den wir in diesem Beitrag vorstellen.

Strukturierte Beschreibung von Forschungsbeiträgen mit dem Open Research Knowledge Graph

Der Open Research Knowledge Graph (ORKG) ermöglicht es, die traditionell in wissenschaftlichen Artikeln beschriebenen Forschungsbeiträge strukturiert und semantisch zu beschreiben. Dafür werden Beiträge dem ORKG hinzugefügt, indem per DOI von CrossRef zentrale Metadaten des Beitrages abgerufen (oder

manuell ergänzt) werden und dann die inhaltliche Beschreibung der Forschungsbeiträge mittels spezialisierter Eingabemasken erfolgt. Solche strukturierten inhaltlichen Beschreibungen wissenschaftlicher Beiträge sollten das adressierte Forschungsproblem, die verwendeten Materialien und Methoden sowie das erzielte Ergebnis auf eine solche Weise beschreiben, dass der Beitrag mit anderen Arbeiten, die dasselbe Forschungsproblem adressieren, vergleichbar wird. Die semantische Beschreibung folgt dem RDF-Subjekt-Prädikat-Objekt-Paradigma und kann jederzeit flexibel von den Nutzerinnen und Nutzern um eigene weitere Prädikate (Properties oder Attribute) erweitert werden. Eine Vorschlagsfunktion ermöglicht das leichte Finden und Wiederverwenden existierender Prädikate und Entitäten. Abbildung 1 veranschaulicht die strukturierte Eingabe eines Forschungsbeitrages über die Wirkung von Ozon auf das Pflanzenwachstum.

The screenshot shows a web form for entering research data. At the top, there's a title 'Plant height statistical test' with a plus icon. Below it, a 'Use template' section shows 'Student's t-test' and 'Research Problem' as selected templates. The main form has several fields: 'Has research problem' with the value 'Effect of ozone on plant growth', 'Dataset' with the URL 'http://datarepository.org/mydataset', 'Study design dependent variable' with the value 'Plant height', and 'P-value' with the value '0.0023'. There are plus icons next to each field and a 'Template' dropdown menu. At the bottom, there are two '+ Add property' buttons.

Abbildung 1: Strukturierte Eingabe eines Forschungsbeitrages im ORKG

Vorlagen

Die strukturierte Beschreibung von Forschungsbeiträgen ist oft keine leichte Aufgabe, denn die Beschreibung wissenschaftlicher Erkenntnisse ist komplex und beruht auf Fachwissen. Einerseits muss entschieden werden, in welcher Granularität ein Forschungsbeitrag beschrieben werden soll. Andererseits sollten Forschungsbeiträge, die dasselbe Problem behandeln, vergleichbar sein. Aus diesem Grund unterstützt der ORKG die Möglichkeit, Vorlagen zu erstellen, die die Struktur wissenschaftlicher Information spezifizieren. Vorlagen können dann bei der

Beschreibung von Forschungsbeiträgen wiederverwendet werden, um die Dateneingabe zu erleichtern und die Vergleichbarkeit zu sichern.

Vergleiche

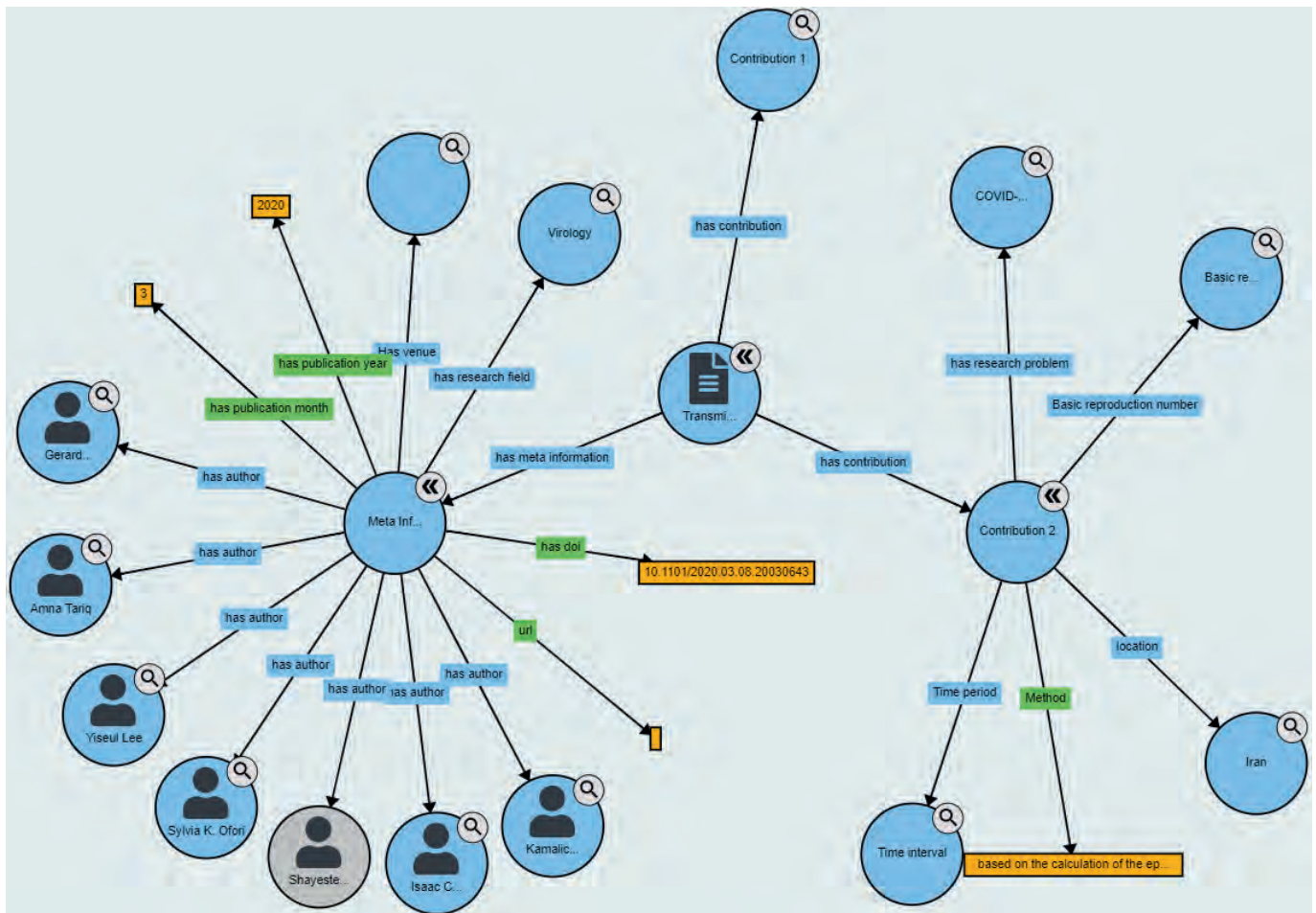
Der ORKG ermöglicht es, Forschungsbeiträge, die sich mit einem bestimmten Problem befassen, automatisiert zu vergleichen. Vergleiche unterstützen Nutzerinnen und Nutzer bei der Gewinnung einer State-of-the-Art-Übersicht. Ein klassisches Beispiel in der Informatik ist ein Vergleich der Best-/Worst-Case-Leistung von Sortieralgorithmen oder die Präzision und Rückruf von Algorithmen zur Fahrzeugerkennung in Bildern. Für Forschende in Virologie und Epidemiologie ist es interessant, die Reproduktionszahl verschiedener Viren vergleichen zu können. Solche Vergleiche geben einen Überblick über Schlüsselinformationen zu einem Forschungsproblem über Dutzende oder Hunderte von Beiträgen hinweg und sind somit ein wertvolles Instrument, um eine Übersicht über den Stand der Forschung in einem Gebiet zu erlangen.

Graph-Visualisierung

Da es sich beim ORKG um einen Wissensgraphen handelt, können Forschungsbeiträge auch als Graphen visualisiert werden. Die Graph-Visualisierung ist eine hoch entwickelte Benutzeroberfläche zur visuellen Exploration von Graphen und ist somit eine Möglichkeit, mit ORKG-Inhalten zu interagieren. Innovative Funktionen unterstützen und gestalten die Exploration der hochstrukturierten Graphen intuitiv. Der Graph wird automatisch optimal auf dem Bildschirm angeordnet. Knoten lassen sich leicht erweitern, kollabieren oder entfernen. Benutzerinnen und Benutzer können in dem Graphen nach Informationen suchen.

Observatorien

Der ORKG verlässt sich stark auf fachkundige Inhaltskuration und Wissensorganisation. Um disziplinäres Fachwissen zu bündeln, implementiert der ORKG das Konzept von Observatorien. Observatorien bündeln Expertinnen und Experten verschiedener Institutionen, die ORKG-Inhalte für eine bestimmte Disziplin oder ein bestimmtes Forschungsproblem kuratieren und organisieren. Observatorien und ihre Expertinnen und Experten können auf vielfältige Weise Beiträge leisten. Neben dem Hinzufügen und Beschreiben von Beiträgen oder dem Kuratieren bestehender Beiträge spielen Observatorien eine entscheidende Rolle bei der Wissensorganisation in einem Forschungsbereich. Observatorien können beispielsweise für ein Fachgebiet relevante Vorlagen erstellen. Auf diese



Weise tragen Observatorien dazu bei, die Schaffung von qualitativ hochwertigem und vergleichbar strukturiertem wissenschaftlichem Wissen für ihr Gebiet sicherzustellen. Da Wissenskuration und -organisation ressourcenintensiv ist, werden Expertinnen und Experten sowie entsprechende Observatorien und Institutionen im ORKG prominent sichtbar gemacht.

Wissensnachnutzung

Die strukturierte und semantische Beschreibung des Wissens ermöglicht die vereinfachte Wissensnachnutzung. Die oben beschriebenen Vergleiche sind nur eine Art der Wissensnachnutzung. Tatsächlich wird in der Wissenschaft das Wissen in der Literatur auf unzählige Arten nachgenutzt. Um diese Vielfalt zu unterstützen, implementiert der ORKG eine webbasierte Schnittstelle (REST API), welche zum Beispiel mit der Python-Programmiersprache genutzt werden kann. Diese ermöglicht es, ORKG-Inhalte (einzelne Beitragsbeschreibungen sowie Vergleiche) in eine Datenanalyseumgebung wie zum Beispiel Jupyter Notebooks zu laden, weiterzuverarbeiten und domänenspezifische Anwendungen und Visualisierungen zu erstellen. Es lassen sich so einfach Datenvisualisierungen erstellen, aber auch darüber hinaus komplexe datenwissen-

schaftliche Aktivitäten implementieren, die ORKG-Daten mit anderen Daten integrieren, Daten interpolieren, modellieren, simulieren etc.

Weitere Funktionen

Neben den vorgestellten Beispielen umfasst das ORKG-Konzept eine Reihe weiterer bereits implementierter beziehungsweise in naher Zukunft geplanter Funktionen:

- Tabellarische Vergleiche des Stands der Forschung zu einem Forschungsproblem können eigenständig mit DOI publiziert und in verschiedenen Formaten exportiert werden.
- Übersichtstabellen zum Stand der Forschung können semi-automatisch aus PDF-Dokumenten in den ORKG importiert werden.
- Alle Beiträge des ORKG werden versioniert, sodass alle Änderungen nachverfolgt und von der Fachgemeinschaft diskutiert werden können.
- Automatische Extraktionsfunktionen unterstützen die Befüllung des Wissensgraphen.
- Forschungsprobleme können eigenständig beschrieben, mit relevanten Quellen versehen und einer Taxonomie von Forschungsgebieten zugeordnet werden.

Abbildung 2: Dynamische Graph-Visualisierung im ORKG

Abbildung 3:
Vergleich von
Frage-Antwort-
Systemen und
deren Evalua-
tionsergebnisse
im unteren Teil

Properties	LIMS participation at QALD 5@CLEF Contributed 1 – 2015	Cross-Lingual Question Answering Using Common Semantic Space Contributed 1 – 2016	CASIA@ V2: a MLN-based question answering system over linked data Contributed 1 – 2014
Has research problem	Question answering systems	Question answering systems	Question answering systems
Implementation	SemGraphQA	UTQA	CASIA
Disambiguation task	Local disambiguation	Local disambiguation	Local disambiguation
Query construction task	Using info. from the QA	Empty	Using machine learning
Question analysis task	Dependency parser NE n-gram strategy	POS learned	Dependency parser NE n-gram strategy POS learned
On	QALD-6	QALD-6	QALD-6
Dataset	DBpedia 2015 DBpedia 2015 with abstracts LinkedSpending	DBpedia 2015 DBpedia 2015 with abstracts LinkedSpending	DBpedia 2015 DBpedia 2015 with abstracts LinkedSpending
Evaluation	SemGraphQA	UTQA	UTQA
Language	Farsi	English	Spanish
F-measure	0.37	0.65	0.68
Precision	0.70	0.70	0.76
Recall	0.25	0.61	0.62

- Fachwissenschaftlerinnen und Fachwissenschaftler können eigenständig domänenspezifische Visualisierungen auf Basis des ORKG realisieren

Anwendungsfälle

In diesem Abschnitt zeigen wir, wie der ORKG zur Strukturierung von Informationen für drei verschiedene Bereiche verwendet werden kann. Für jede Domäne wird ein ORKG-Vergleich generiert, um einen Überblick über den Stand der Technik für dieses spezielle Beispiel zu geben.

Informatik

In der Informatik ist eine Reihe von wiederkehrenden Eigenschaften wissenschaftlicher Beiträge für den ORKG von Interesse. Zu diesen Eigenschaften gehören unter anderem Evaluationsergebnisse (F-Maß, Präzision, Recall), Datensätze, Benchmarks, Modelleigenschaften und Implementierungsansätze. Für eine beträchtliche Anzahl von Informatikarbeiten sind diese Eigenschaften anwendbar und für viele Forscherinnen und Forscher von Interesse. In diesem Beispiel zeigen wir, wie diese wiederkehrenden Eigenschaften im Bereich der Fragebeantwortung (Question Answering, QA) verwendet werden. Das Ziel der QA-Aufgabe besteht darin, automatisch Antworten auf Fragen in natürlicher Sprache zu geben. Um QA-bezogene Informationen im ORKG zu organisieren, werden methodische Ansätze und Evaluationsergebnisse verschiedener Arbeiten und der imple-

mentierten Ansätze repräsentiert und verglichen. Ein interessanter Aspekt dieser Probleme besteht darin, dass die verglichenen QA-Systeme und die Evaluationsergebnisse dieser Systeme ursprünglich weder in denselben Artikeln noch von denselben Autorinnen und Autoren veröffentlicht werden. Aufgrund der dynamischen Aspekte des ORKG ist es jedoch möglich, im Nachhinein Verknüpfungen zwischen den Systemen und Auswertungen herzustellen. In Abbildung 3 ist ein Vergleich verschiedener QA-Systeme und ihrer Evaluationsergebnisse in Bezug auf Präzision, Recall und F-Maß im QALD-Benchmark dargestellt. Jedes QA-System wird in einem eigenen in unterschiedlichen Journalen publizierten Artikel vorgestellt, kann aber im Vergleich direkt mit anderen Ansätzen verglichen werden. Es ist dadurch möglich, sich schnell einen Überblick über den Stand der Technik und die Leistungsfähigkeit der Systeme zu verschaffen. Darüber hinaus gibt der Vergleich dieser Evaluationsergebnisse im Zeitverlauf Aufschluss über die Fortschritte für einen bestimmten Forschungsbereich.

Epidemiologie und COVID-19

Die COVID-19-Pandemie führt auch in der Forschung zu einer Fülle an täglich neu veröffentlichten Artikeln. Um die COVID-19-Forschung zu unterstützen, haben viele Verlage beschlossen, Artikel mit Bezug zu COVID-19 Open Access zu veröffentlichen. Zwar ist der Zugang von entscheidender Bedeutung, aber die Organisation der in Artikeln veröffentlichten Informa-

Properties	The early phase of the COVID-19 outbreak in Lombardy, Italy Contribution 1 - 2020	Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia Contribution 1 - 2020	Estimation of the Transmission Risk of 2019-nCoV and Its Implication for Public Health Interventions Contribution 1 - 2020	Pattern of early human-to-human transmission of Wuhan 2019-nCoV Contribution 1 - 2020
Has research problem	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number
Location	Lombardy, Italy	China	China	China and overseas
Study date	2020-02-20	2020-01-22	2020-01-22	2020-01-18
R0 estimates (average)	3.1	2.2	6.47	2.2
95% confidence interval	2.9-3.2	1.4-3.9	5.71-7.23	Empty

tionen ist für eine effektive Forschung unerlässlich. Leider ist dies jedoch extrem zeitaufwändig und damit ein großes Hindernis. Ein Beispiel für eine fundamentale Forschungserkenntnis, die über zahlreiche veröffentlichte (Preprint-)Artikel verteilt ist, ist die COVID-19-Basisreproduktionszahl R_0 , das jeweilige 95%-Konfidenzintervall, der Standort der untersuchten Population und der Beobachtungszeitraum. Im Gegensatz zur herkömmlichen dokumentengestützten Veröffentlichung solcher Informationen in natürlichsprachlichen Texten, Tabellen oder Abbildungen können wir mit dem ORKG solche Informationen in strukturierter, semantischer Form veröffentlichen. Informationen sind somit maschinell und publikationsübergreifend auswertbar. Maschinell verwertbares wissenschaftliches Wissen eröffnet eine Reihe von sehr interessanten Möglichkeiten. Wie bereits erwähnt, ist es durch die strukturierte semantische Darstellung wissenschaftlicher Erkenntnisse im Wissensgraphen möglich, Literaturvergleiche automatisch zu erstellen. Abbildung 4 veranschaulicht dies für unseren Anwendungsfall der COVID-19-Basisreproduktionszahl. Darüber hinaus können sich ORKG-Vergleiche im Gegensatz zu statischen PDF-Übersichtsartikeln (Surveys oder Reviews) weiterentwickeln. Wenn neue Literatur zur Basisreproduktionszahl veröffentlicht wird, ist es einfach, einen solchen Vergleich kontinuierlich zu erweitern, der somit weiterhin den aktuellen Wissensstand in vergleichbarer Weise widerspiegelt.

Die wahre Stärke des ORKG zeigt sich jedoch, wenn der Wissensgraph für weitere übergreifende Datenanalysen genutzt wird. Wir demonstrieren dies, indem wir Jupyter Notebooks mit dem ORKG verbinden, um die Flexibilität von datenwissenschaftlichen Umgebungen und Programmiersprachen wie Python zu nutzen, um die COVID-19-Vergleichsdaten zu visualisieren. Abbildung 5 zeigt eine mögliche Visualisierung der R_0 -Werte und des 95%-Konfidenzintervalls über die Zeit. Natürlich können bei solchen Analysen Daten aus weiteren Quellen herangezogen werden,

zum Beispiel in diesem Fall statistische Daten zur Todesfallrate.

Materialwissenschaft

Schließlich stellen wir ein Beispiel aus der Elektrochemie, den Material- und Ingenieurwissenschaften vor. Silizium ist ein wichtiges Element der modernen Technologie. Es wird häufig zur Herstellung von Metalllegierungen, optischen Fasern, Solarelementen, Hochleistungskeramik, Batterien, Mikrochips und zahlreichen anderen vorteilhaften Anwendungen

Abbildung 4: Automatischer Vergleich der in der Literatur veröffentlichten grundlegenden Reproduktionszahlen

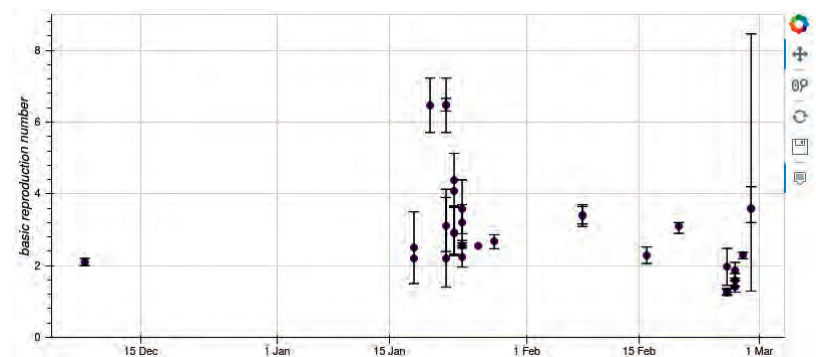


Abbildung 5: Visualisierung der R_0 -Werte und des 95%-Konfidenzintervalls über die Zeit als Datenanalyse auf dem ORKG-Wissensgraph

verwendet. Für Solar- und Elektronikgeräte besteht ein Bedarf an Solarsilizium (SoG-Si) mit einer Reinheit von 99,9999 Prozent oder Elektroniksilizium mit einer noch höheren Reinheit. Die Siliziumelektrochemie in geschmolzenen Salzen hat in letzter Zeit aufgrund ihres Potenzials zur Herstellung von SolarGrade-Silizium mit vernachlässigbarem Kohlenstoff-Fußabdruck große Aufmerksamkeit erregt. Der in Abbildung 6 dargestellte Vergleich bietet einen umfassenden Überblick über mehrere Parameter wie Siliziumdioxid-Vorläufer, Elektrolyt, Kontaktelektrode oder Temperatur der experimentellen Bedingungen der siliziumelektrochemischen Reduk-

Properties	Facile electrosynthesis of silicon carbide nanowires from silica/carbon precursors in molten salt <small>Contribution 1 – 2017</small>	Up-scalable and controllable electrolytic production of photo-responsive nanostructured silicon <small>Contribution 1 – 2013</small>	Electrochemical formation of a p-n junction of thin film silicon deposited in molten salt <small>Contribution 1 – 2017</small>	Silicon surface texturing by electro-deoxidation of a thin silica layer in molten salt <small>Contribution 1 – 2010</small>	Electrodeposition of crystalline and photoactive silicon directly from silicon dioxide nanoparticles in molten CaCl ₂ <small>Contribution 1 – 2002</small>
Has research problem	Silicon electrochemistry	Silicon electrochemistry	Silicon electrochemistry	Silicon electrochemistry	Silicon electrochemistry
Electrolyte	CaCl ₂	CaCl ₂	CaCl ₂ -CaO	CaCl ₂	CaCl ₂
Si precursor	SiO ₂ and C powder, pellet	SiO ₂ pellet	CaSiO ₃ , SiO ₂ powder	SiO ₂ layer (0.3–2.0 μm) on Si	SiO ₂ powder (5–15 nm)
Contacting electrode	Ni	Mo	graphite, p-Si	Mo	Ag, Mo, n-Si (100)
Counter electrode	graphite	graphite	graphite	graphite	glassy carbon
(pseudo)reference electrode	Pt	Ag/AgCl	graphite	graphite	Mo - Ca/Ca ₂ ⁺
Temperature	900 °C	900 °C	850 °C	850–900 °C	850 °C
Process specification	synthesis of Si-C nanowires	photoresponsive nanostructured Si	p-n junction of Si films	structuring, photoresponsive layer	Si layer formation

Abbildung 6: ORKG-Vergleich für Arbeiten zur siliziumelektrochemischen Reduktion in geschmolzenen Elektrolyten in der Materialwissenschaft

tion in geschmolzenen Elektrolyten. Auf diese Weise können Forschende leicht publikationsübergreifend relevante Parameter analysieren, die in den Prozessspezifikationen für die Erzeugung von Siliziumoberflächenstrukturen verwendet werden.

Einbindung bibliothekarischer Communities und Dienste

Bibliotheken können eine zentrale Rolle in der Weiterentwicklung und Kuratierung des ORKG haben. Wissenschaftlerinnen und Wissenschaftler müssen nicht nur durch KI- und Text-Mining-Verfahren unterstützt werden, auch die informationswissenschaftlichen Kernkompetenzen in der inhaltlichen und formalen Erschließung sowie der Wissensorganisation sind für den Erfolg des ORKG unverzichtbar. Der Open Research Knowledge Graph soll daher eine umfassende und breite Wissensinfrastruktur für die Zusammenarbeit von Fachwissenschaftlern, Informationswissenschaftlern, Bibliothekaren und Nutzern werden. Um möglichst viele Stakeholder und Multiplikatoren einzubinden, arbeiten wir derzeit an verschiedenen Strategien:

- Der ORKG ist eine radikal offene Forschungsinfrastruktur entsprechend den Open-Science-, Open-Data- und Open-Source-Prinzipien: alle Software, Informationen und Daten sind unter offenen Lizenzen verfügbar. Der ORKG stellt eine umfassende Programmierschnittstelle (API) bereit, mit der leicht spezifische Anwendungen auf Basis der ORKG-Infrastruktur realisiert werden können.
- In fachspezifische Observatorien organisieren Fachwissenschaftlerinnen und Fachwissenschaftler gemeinsam mit Fachreferentinnen und Fachreferenten die Forschungsarbeiten zu relevanten Forschungsfragen in einem klar definierten Fach-

gebiet. Durch die prominente Darstellung der beteiligten Personen und Organisationen bei den jeweiligen Einträgen im ORKG wollen wir deren Einsatz sichtbar machen und anerkennen.

- Wir planen umfassende und gut dokumentierte Vergleiche des Stands der Forschung zu einer spezifischen Forschungsfrage zukünftig als eigenständige Publikationen in einem Open-Access-Journal zu veröffentlichen. Bereits jetzt können solche Vergleiche im ORKG mit einem DOI versehen und zitierbar veröffentlicht werden.
- (Semi-)automatische Verfahren setzen auf gute Normdaten auf, dieser Teil der intellektuellen Erschließung bleibt Aufgabe der Fachreferate. Verschiedene Ontologien und Normdaten wie zum Beispiel die gemeinsame Normdatei können dabei helfen, das Wissen im ORKG zu strukturieren.
- Gleichermaßen gibt es für Bibliothekarinnen und Bibliothekare weiterhin die Aufgabe der formalen Erschließung und der Kuratierung von Formalkriterien: Nachdem lange Zeit nur selbstständige Werke formal erschlossen wurden, ist die maschinell unterstützte Aufsatzerschließung hinzugekommen. Die Erfassung und Kuratierung formaler Erschließungselemente für den ORKG ist die logische nächste Stufe.

Wir würden uns sehr freuen, wenn wir in der Bibliothekscommunity weitere Mitstreiterinnen und Mitstreiter gewinnen, die gemeinsam mit Fachwissenschaftlerinnen und Fachwissenschaftlern Observatorien betreuen, existierende Infrastrukturen mit dem ORKG verknüpfen oder neue fachspezifische Anwendungen (zum Beispiel Elemente für Fachinformationsdienste) auf Basis der ORKG-Infrastruktur realisieren.

Fazit

Mit dem von der TIB und Partnern entwickelten Open Research Knowledge Graph werden wissenschaftliche Erkenntnisse entsprechend den FAIR-Prinzipien organisiert. Konzepte, die derzeit noch tief in den unstrukturierten Publikationen versteckt sind, bekommen im ORKG mit persistenten Identifikatoren eine eindeutige Identität und können mit anderen relevanten Konzepten oder Artefakten semantisch verknüpft werden. Über die Web-, API-, Daten-Dump- oder Anfrage-Schnittstellen und die offenen Lizenzen sind die semantisch strukturierten ORKG-Inhalte auf vielfältige Weise leicht abrufbar.

Trotz des Potenzials und Nutzens des ORKG stehen wir natürlich noch ganz am Anfang der Entwicklung und es gibt noch viele offene Fragen zu beantworten: Wie können wir Fachwissenschaftlerinnen und Fachwissenschaftler zunehmend in die Kuratierung einbeziehen? Skalieren die Techniken des ORKG für breite Fachgebiete und kann eine kontinuierliche Evolution der semantischen Repräsentation erreicht werden? Wie können Incentivierungssysteme für Wissenschaftlerinnen und Wissenschaftler sinnvoll für den ORKG adaptiert oder transformiert werden? Welche Aufgaben übernehmen Bibliotheken? **I**



Prof. Dr. Sören Auer

TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften
 · Direktor der TIB
 · Leitung Programmbereich Forschung und Entwicklung
 · Leitung Forschungsgruppe Data Science & Digital Libraries
 · Professor für „Data Science & Digital Libraries“ an der Fakultät für Elektrotechnik und Informatik der Leibniz Universität Hannover
 auer@tib.eu



Dr. Irina Sens

TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften
 · Stellvertretende Direktorin der TIB
 · Leitung Bibliotheksbetrieb
 irina.sens@tib.eu



Dr. Markus Stocker

TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften
 · Leitung der Nachwuchsforschungsgruppe Knowledge Infrastructures
 markus.stocker@tib.eu

Mit smarter Logistk für Bibliotheken in die Zukunft

Effiziente Lösung für Mediensortierung und -transport



Perfekt gerüstet für künftige Aufgaben moderner Bibliotheken

- ⬡ Schnittstelle zur Rückgabe- und Sortiertechnik flex AMH™ von bibliotheca
- ⬡ Unterstützung **24/7 Betrieb** und **Self-Service Konzept** moderner Bibliotheken
- ⬡ Automatischer Transport **individuell** auf die Bibliothek zugeschnitten
- ⬡ Einsetzbar sowohl in **Bestands-** als auch **Neubauten**



Telelift GmbH · Frauenstraße 28 · 82216 Maisach · info@telelift-logistic.com · www.telelift-logistic.com · +49 (0)8141 / 315 91-0