International Journal on Digital Libraries
https://doi.org/10.1007/s00799-021-00306-x

Analysing the requirements for an Open Research Knowledge Graph: use cases, quality requirements, and construction strategies

Arthur Brack[1,2] · Anett Hoppe[1] · Markus Stocker[1] · Sören Auer[1,2] · Ralph Ewerth[1,2]

Received: 4 February 2021 / Revised: 8 July 2021 / Accepted: 12 July 2021

Abstract
Current science communication has a number of drawbacks and bottlenecks which have been subject of discussion lately: Among others, the rising number of published articles makes it nearly impossible to get a full overview of the state of the art in a certain field, or reproducibility is hampered by fixed-length, document-based publications which normally cannot cover all details of a research work. Recently, several initiatives have proposed knowledge graphs (KG) for organising scientific information as a solution to many of the current issues. The focus of these proposals is, however, usually restricted to very specific use cases. In this paper, we aim to transcend this limited perspective and present a comprehensive analysis of requirements for an Open Research Knowledge Graph (ORKG) by (a) collecting and reviewing daily core tasks of a scientist, (b) establishing their consequential requirements for a KG-based system, (c) identifying overlaps and specificities, and their coverage in current solutions. As a result, we map necessary and desirable requirements for successful KG-based science communication, derive implications, and outline possible solutions.

Keywords Scholarly communication · Research knowledge graph · Design science research · Requirements analysis

1 Introduction

Today's scholarly communication is a document-centred process and as such, rather inefficient. Scientists spend considerable time in finding, reading, and reproducing research results from PDF files consisting of static text, tables, and figures. The explosion in the number of published articles [14] aggravates this situation further: It gets harder and harder to stay on top of current research, that is to find relevant works, compare and reproduce them, and later on, to make one's own contribution known for its quality.

Some of the available infrastructures in the research ecosystem already use *knowledge graphs* (KG)[1] to enhance their services. Academic search engines, for instance, such as *Microsoft Academic Knowledge Graph* [38] or *Literature Graph* [1] utilise metadata-based graph structures which link research articles based on citations, shared authors, venues, and keywords.

Recently, initiatives have promoted the usage of KGs in science communication, but on a deeper, semantic level [3,50,56,73,78,84,115]. They envision the transformation of the dominant document-centred knowledge exchange to knowledge-based information flows by representing and expressing knowledge through semantically rich, interlinked KGs. Indeed, they argue that a shared structured representation of scientific knowledge has the potential to alleviate

✉ Arthur Brack
arthur.brack@tib.eu

Anett Hoppe
anett.hoppe@tib.eu

Markus Stocker
markus.stocker@tib.eu

Sören Auer
auer@tib.eu

Ralph Ewerth
ralph.ewerth@tib.eu

1 TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany

2 L3S Research Center, Leibniz University Hannover, Hannover, Germany

[1] Acknowledging that knowledge graph is vaguely defined, we adopt the following definition: A *knowledge graph* (KG) consists of (1) an *ontology* describing a conceptual model (e.g. with classes, relation types, and axioms), and (2) the corresponding *instance data* (e.g. objects, literals, and <subject, predicate, object>-triplets) following the constraints posed by the ontology (e.g. instance of relations, axioms, etc.). The construction of a KG involves *ontology design* and *population* with instances.

Published online: 28 July 2021

some of the science communication's current issues: Relevant research could be easier to find, comparison tables automatically compiled, own insights rapidly placed in the current ecosystem. Such a powerful data structure could, more than the current document-based system, also encourage the interconnection of research artefacts such as datasets and source code much more than current approaches (like Digital Object Identifier (DOI) references etc.), allowing for easier reproducibility and comparison. To come closer to the vision of knowledge-based information flows, research articles should be enriched and interconnected through machine-interpretable semantic content. The usage of Papers With Code [80] in the machine learning community and Jaradeh et al.'s study [56] indicate that authors are also willing to contribute structured descriptions of their research articles.

The work of a researcher is manifold, but current proposals usually focus on a specific use case (e.g. the aforementioned examples focus on enhancing academic search). In this paper, we present a detailed analysis of common literature-related tasks in a scientist's daily life and analyse (a) how they could be supported by an ORKG, (b) what requirements result for the design of (b1) the KG and (b2) the surrounding system, (c) how different use cases overlap in their requirements and can benefit from each other. Our analysis is led by the following research questions:

1. Which use cases should be supported by an ORKG?
   (a) Which user interfaces are necessary?
   (b) Which machine interfaces are necessary?
2. What requirements can be defined for the underlying ontologies to support these use cases?
   (a) Which granularity of information is needed?
   (b) To what degree is domain specialisation needed?
3. What requirements can be defined for the instance data in context of the respective use cases?
   (a) Which completeness is sufficient for the instance data?
   (b) Which correctness is sufficient for the instance data?
   (c) Which approaches (human vs. machine) are suitable to populate the ORKG?

Our analysis concentrates on eliciting use cases, defining quality requirements for the underlying KG to support these use cases, and elaborating construction strategies for the KG. We follow the design science research (DSR) methodology [51]. In this study, we focus on the first phase of DSR and conduct a requirements analysis. The objective is to chart necessary (and desirable) requirements for successful KG-based science communication, and, consequently, provide a map for future research.

Compared to our paper at the 24th International Conference on Theory and Practice of Digital Libraries 2020 [16], this journal paper has been modified and extended as follows: The related work section is updated and extended with the new sections *Quality of knowledge graphs* and *Systematic literature reviews*. The new "Appendix 1" section contains comparative overviews of datasets for research knowledge graph population tasks such as sentence classification, relation extraction, and concept extraction. These comparisons are intended to give a sense of what kind of information can be automatically extracted from scientific texts with what accuracy using current state-of-the-art methods. This is important to suggest appropriate construction strategies (i.e. manual, semi-automatic, automatic) for the respective use cases based on their data quality requirements. To be consistent with the terminology in related work, we use the term "completeness" instead of "coverage" and "correctness" instead of "quality". The requirements analysis in Sect. 3 is revised and contains more details with more justifications for the posed requirements and approaches.

The remainder of the paper is organised as follows. Section 2 summarises related work on research KGs, scientific ontologies, KG construction, data quality requirements, and systematic literature reviews. The requirements analysis is presented in Sect. 3, while Sect. 4 discusses implications and possible approaches for ORKG construction. Finally, Sect. 5 concludes the requirements analysis and outlines areas of future work. "Appendix 1" section contains comparative overviews for the tasks of sentence classification, relation extraction, and concept extraction.

2 Related work

This section gives a brief overview of (a) existing research KGs, (b) ontologies for scholarly knowledge, (c) approaches for KG construction, (d) quality dimensions of KGs, and (e) processes in systematic literature reviews.

2.1 Research knowledge graphs

Academic search engines (e.g. Google Scholar, Microsoft Academic, Semantic Scholar) exploit graph structures such as the Microsoft Academic Knowledge Graph [38], SciGraph [113], the Literature Graph [1], or the Semantic Scholar Open Research Corpus (S2ORC) [70]. These graphs interlink research articles through metadata, e.g. citations, authors, affiliations, grants, journals, or keywords.

To help reproduce research results, initiatives such as Research Graph [2], Research Objects [7], and OpenAIRE [73] interlink research articles with research artefacts such

as datasets, source code, software, and video presentations. Scholarly Link Exchange (Scholix) [20] aims to create a standardised ecosystem to collect and exchange links between research artefacts and literature.

Some approaches connect articles at a more semantic level: Papers With Code [80] is a community-driven effort to supplement machine learning articles with tasks, source code, and evaluation results to construct leaderboards. Ammar et al. [1] link entity mentions in abstracts with DBpedia [66] and Unified Medical Language System (UMLS) [11], and Cohan et al. [23] extend the citation graph with citation intents (e.g. citation as background or used method).

Various scholarly applications benefit from semantic content representation, e.g. academic search engines by exploiting general-purpose KGs [112], and graph-based research paper recommendation systems [8] that utilise citation graphs and mentioned entities. However, the coverage of science-specific concepts in general-purpose KGs is rather low [1], e.g. the task "geolocation estimation of photos" from Computer Vision is neither present in Wikipedia nor in the Computer Science Ontology (CSO) [95].

2.2 Scientific ontologies

Various ontologies have been proposed to model metadata such as bibliographic resources and citations [83]. Iniesta and Corcho [93] reviewed ontologies to describe scholarly articles. In the following, we describe some ontologies that conceptualise the semantic content in research articles.

Several ontologies focus on rhetorical [27,49,109] (e.g. Background, Methods, Results, Conclusion), argumentative [69,105] (e.g. claims, contrastive, and comparative statements about other work) or activity-based structure [84] (e.g. sequence of research activities) of research articles. Others describe scholarly knowledge with linked entities such as problem, method, theory, statement [19,50], or focus on the main research findings and characteristics of research articles described in surveys with concepts such as problems, approaches, implementations, and evaluations [40,106].

Various domain-specific ontologies exist, for instance, mathematics [65] (e.g. definitions, assertions, proofs), machine learning [62,74] (e.g. dataset, metric, model, experiment), and physics [96] (e.g. formation, model, observation). The EXPeriments Ontology (EXPO) is a core ontology for scientific experiments that conceptualise experimental design, methodology, and results [99], while the Scientific Observation Model (CRMsci) is an ontology of metadata about scientific observations, processed data, and measurements in descriptive and empirical sciences (e.g. biodiversity, geology, geography, archaeology) [34]. Various repositories provide access to several ontologies such as Open Biological and Biomedical Ontologies (OBO) Foundry [98] for the domain of life sciences or Linked Open Vocabularies (LOV) [107] for web data.

Taxonomies for domain-specific research areas support the characterisation and exploration of a research field. Salatino et al. [95] give an overview, e.g. Medical Subject Heading (MeSH), Physics Subject Headings (PhySH), Computer Science Ontology (CSO). Gene Ontology [26] and Chemical Entities of Biological Interest (CheBi) [30] are KGs for genes and molecular entities.

2.3 Construction of knowledge graphs

Nickel et al. [77] classify KG construction methods into four groups: (1) curated approaches, i.e. triples created manually by a closed group of experts, (2) collaborative approaches, i.e. triples created manually by an open group of volunteers, (3) automated semi-structured approaches, i.e. triples extracted automatically from semi-structured text via handcrafted rules, and (4) automated unstructured approaches, i.e. triples are extracted automatically from unstructured text.

2.3.1 Manual approaches

WikiData [108] is one of the most popular KGs with semantically structured, encyclopaedic knowledge curated manually by a community. As of January 2021, WikiData comprises 92M entities curated by almost 27.000 active contributors. The community also maintains a taxonomy of categories and "infoboxes" which define common properties of certain entity types. Furthermore, Papers With Code [80] is a community-driven effort to interlink machine learning articles with tasks, source code, and evaluation results. KGs such as Gene Ontology [26] or Wordnet [41] are curated by domain experts. Research article submission portals such as EasyChair (https://wwww.easychair.org/) enforce the authors to provide machine-readable metadata. Librarians and publishers tag new articles with keywords and subjects [113]. Virtual research environments enable the execution of data analysis on interoperable infrastructure and store the data and results in KGs [101].

2.3.2 Automated approaches

Automatic KG construction from text: Petasis et al. [85] present a review on *ontology learning*, that is ontology creation from text, while Lubani et al. [72] review *ontology population systems*. Pajura and Singh [88] give an overview of the involved tasks for *KG population*: (a) *information extraction* to extract a graph from text with *entity extraction* and *relation extraction*, and (b) *graph construction* to clean and complete the extracted graph, as it is usually ambiguous, incomplete and inconsistent. *Coreference resolution* [17,71] clusters different mentions of the same entity in text and *entity*

linking [63] maps mentions in text to entities in the KG. *Entity resolution* [104] identifies objects in the KG that refer to the same underlying entity. For *taxonomy population*, Salatino et al. [95] provide an overview of methods based on rule-based natural language processing (NLP), clustering and statistical methods.

The Computer Science Ontology (CSO) has been automatically populated from research articles [95]. The AI-KG was automatically generated from 333,000 research papers in the artificial intelligence (AI) domain [32]. It contains five entity types (tasks, methods, metrics, materials, others) linked by 27 relations types. Kannan et al. [58] create a multimodal KG for deep learning papers from text and images and the corresponding source code. Brack et al. [17] generate a KG for 10 different science domains with the concept types material, method, process, and data. Zhang et al. [115] suggest a rule-based approach to mine research problems and proposed solutions from research papers.

Information extraction from scientific text: Information extraction is the first step in the automatic KG population pipeline. Nasar et al. [75] survey methods on information extraction from scientific text. Beltagy et al. [9] present benchmarks for several scientific datasets and Peng et al. [82] especially for the biomedical domain. "Appendix 1" section presents comparative overviews of datasets for the tasks sentence classification, relation extraction, and concept extraction, respectively, in research papers.

There are datasets which are annotated at *sentence level* for several domains, e.g. biomedical [31,60], computer graphics [43], computer science [24], chemistry and computational linguistics [105], or algorithmic metadata [94]. They cover either only abstracts [24,31,60] or full articles [43,69,94,105]. The datasets differentiate between five and twelve concept classes (e.g. Background, Objective, Results). Machine learning approaches for datasets consisting of abstracts achieve an F1 score ranging from 66 to 92% and for datasets with full papers F1 scores ranging from 51 to 78% (see Table 2).

More recent corpora, annotated at *phrasal level*, aim at constructing a fine-grained KG from scholarly abstracts with the tasks of concept extraction [4,15,44,71,89], binary relation extraction [4,45,71], *n*-ary relation extraction [55, 57,59], and coreference resolution [17,25,71]. They cover several domains, e.g. material sciences [44]; computational linguistics [45,89]; computer science, material sciences, and physics [4]; machine learning [71]; biomedicine [25,57,64]; or a set of ten scientific, technical and medical domains [15, 17,37]. The datasets differentiate between four to seven concept classes (like task, method, tool) and between two to seven binary relation types (like used-for, part-of, evaluate-for). The extraction of *n*-ary relations involves extraction of relations among multiple concepts such as drug-gene-mutation interactions in medicine [57], experiments related to solid oxide fuel cells with involved material and measurement conditions in material sciences [44], or task-dataset-metric-score tuples for leaderboard construction for machine learning tasks [59].

Approaches for concept extraction achieve F1 scores ranging from 56.6 to 96.9% (see Table 4), for coreference resolution F1 scores range from 46.0 to 61.4% [17,25,71], and for binary relation extraction from 28.0 to 83.6% (see Table 3). The task of *n*-ary relation extraction with an F1 score from 28.7 to 56.4% [57,59] is especially challenging, since such relationships usually span beyond sentences or even sections and thus, machine learning models require an understanding of the whole document. The inter-coder agreement for the task of concept extraction ranges from 0.6 to 0.96 (Table 4), for relation extraction from 0.6 to 0.9 (see also Table 3), while for coreference resolution the value of 0.68 was reported in two different studies [17,71]. The results suggest that these tasks are not only difficult for machines but also for humans in most cases.

2.4 Quality of knowledge graphs

KGs may contain billions of machine-readable facts about the world or a certain domain. However, do the KGs have also an appropriate quality? Data quality (DQ) is defined as *fitness for use by a data consumer* [110]. Thus, to evaluate data quality, it is important to know the needs of the data consumer since, in the end, the consumer judges whether or not a product is fit for use. Wang et al. [110] propose a data quality evaluation framework for information systems consisting of 15 dimensions grouped into four categories, i.e.:

1. *Intrinsic DQ*: accuracy, objectivity, believability, and reputation.
2. *Contextual DQ*: value-added, relevancy, timeliness, completeness, and an appropriate amount of data.
3. *Representational DQ*: interpretability, ease of understanding, representational consistency, and concise representation.
4. *Accessibility DQ*: accessibility and access security.

Bizer [10] and Zaveri [114] propose further dimensions for the Linked Data context like consistency, verifiability, offensiveness, licensing, and interlinking. Pipino et al. [87] subdivide completeness into *schema completeness*, i.e. the extent to which classes and relations are missing in the ontology to support a certain use, *column completeness* (also known as *Partial Closed World Assumption* [47]), i.e. the extent to which facts are not missing, and *population completeness*, i.e. the extent to which instances for a certain class are missing. Färber et al. [39] comprehensively evaluate and

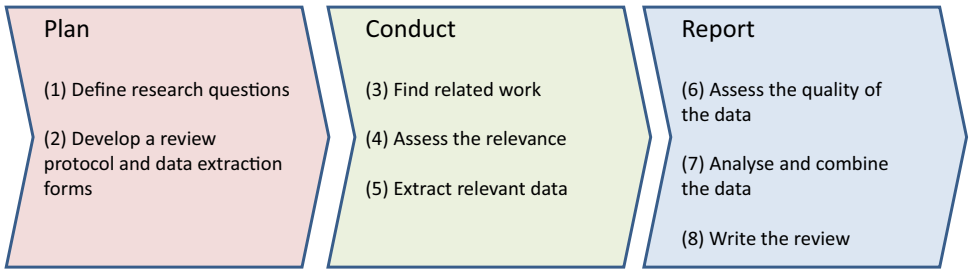

Fig. 1 Activities within a systematic literature review

compare the data quality of popular KGs (e.g. DBpedia, Freebase, WikiData, YAGO) using such dimensions.

To evaluate the correctness of instance data (also known as *precision*), the facts in the KG have to be compared against a ground truth. For that, humans annotate a set of facts as true or false. YAGO found to be 95% correct [103]. The automatically populated AI-KG has a precision of 79% [32] . The KG automatically populated by the Never-Ending Language Learner (NELL) has a precision of 74% [21].

To evaluate the *completeness of instance data* (also known as *coverage and recall*), small collections of ground-truth capturing *all* knowledge for a certain ontology is necessary, that are usually difficult to obtain [111]. However, some studies estimate the completeness of several KGs. Galarrage et al. [46] suggest a rule mining approach to predict missing facts. In Freebase [12] 71% of people have an unknown place of birth, and 75% have an unknown nationality [36]. Suchanek et al. [102] report that 69%-99% of instances in popular KGs (e.g. YAGO, DBPedia) do not have at least one property that other instances of the same class have. The AI-KG has a recall of 81.2% [32].

2.5 Systematic literature reviews

Literature reviews are one of the main tasks of researchers, since a clear identification of a contribution to the present scholarly knowledge is a crucial step in scientific work [51]. This requires a comprehensive elaboration of the present scholarly knowledge for a certain research question. Furthermore, systematic literature reviews help to identify research gaps and to position new research activities [61].

A literature review can be conducted systematically or in a non-systematic, narrative way. Following Fink's [42] definition, a systematic literature review is *"a systematic, explicit, comprehensive, and reproducible method identifying, evaluating, and synthesising the existing body of completed and recorded work"*. Guidelines for systematic literature reviews have been suggested for several scientific disciplines, e.g. for software engineering [61], for information systems [79] and for health sciences [42]. A systematic literature review consists typically of the activities depicted in Fig. 1 subdivided into the phases *plan*, *conduct*, and *report*. The activities may differ in detail for the specific scientific domains [42,61,79]. In particular, a *data extraction form* defines which data has to be extracted from the reviewed papers. Data extraction requirements vary from review to review so that the form is tailored to the specific research questions investigated in the review.

3 Requirements analysis

As the discussion of related work reveals, existing knowledge graphs for research information focus on specific use cases (e.g. improve search engines, help to reproduce research results) and mainly manage metadata and research artefacts about articles. We envision a KG in which research articles are linked through a deep semantic representation of their content to enable further use cases. In the following, we formulate the problem statement and describe our research method. This motivates our use case analysis in Sect. 3.1, from which we derive requirements for an ORKG.

Problem statement: Scholarly knowledge is very heterogeneous and diverse. Therefore, an ontology that conceptualises scholarly knowledge comprehensively does not exist. Besides, due to the complexity of the task, the population of comprehensive ontologies requires domain and ontology experts. Current automatic approaches can only populate rather simple ontologies and achieve moderate accuracy (see Sect. 2.3 and "Appendix 1)" section. *On the one hand, we desire an ontology that can comprehensively capture scholarly knowledge, and instance data with high correctness and completeness. On the other hand, we are faced with a "knowledge acquisition bottleneck".*

Research method: To illuminate the problem statement, we perform a *requirements analysis*. We follow the *design science research (DSR)* methodology [18,53]. The requirements analysis is a central phase in DSR, as it is the basis for design decisions and selection of methods to construct effective solutions systematically [18]. The objective of DSR in general is the innovative, rigorous, and relevant design of information systems for solving important business problems, or the improvement of existing solutions [18,51].

To elicit requirements, we studied guidelines for (a) systematic literature reviews (see Sect. 2.5), (b) data quality

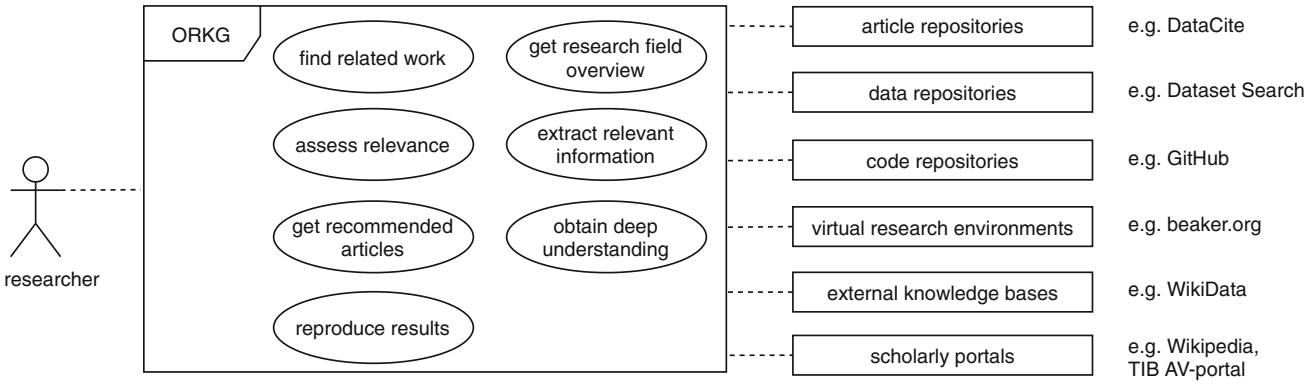

Fig. 2 UML use case diagram for the main use cases between a researcher, an Open Research Knowledge Graph (ORKG), and external systems

requirements for information systems (see Sect. 2.4), and (c) interviewed members of the ORKG and Visual Analytics team at TIB,[2] who are software engineers and researchers in the field of computer science and environmental sciences. Based on the requirements, we elaborate possible approaches to construct an ORKG, which were identified through a literature review (see Sect. 2.3). To verify our assumptions on the presented requirements and approaches, ORKG and Visual Analytics team members reviewed them in an iterative refinement process.

3.1 Overview of the use cases

We define functional requirements with use cases which are a popular technique in software engineering [13]. A use case describes the interaction between a user and the system from the *user's perspective* to achieve a certain goal. Furthermore, a use case introduces a motivating scenario to guide the design of a supporting ontology and the use case analysis helps to figure out which kind of information is necessary [29].

There are many use cases (e.g. literature reviews, plagiarism detection, peer reviewer suggestion) and several stakeholders (e.g. researchers, librarians, peer reviewers, practitioners) that may benefit from an ORKG. Ngyuen et al. [76] discuss some research-related tasks of scientists for information foraging at a broader level. In this study, we focus on use cases that support *researchers* (a) conducting literature reviews (see also Sect. 2.5), (b) obtaining a deep understanding of a research article and (c) reproducing research results. A full discussion of all possible use cases of graph-based knowledge management systems in the research environment is far beyond the scope of this article. With the chosen focus, we hope to cover the most frequent, literature-oriented tasks of scientists.

Figure 2 depicts the main identified use cases, which are described briefly in the following. Please note that we focus on how *semantic content* can improve these use cases and not further metadata.

Get research field overview: Survey articles provide an overview of a particular research field, e.g. a certain research problem or a family of approaches. The results in such surveys are sometimes summarised in structured and comparative tables (an approach usually followed in domains such as computer science, but not as systematically practised in other fields). However, once survey articles are published they are no longer updated. Moreover, they usually represent only the perspective of the authors, i.e. very few researchers of the field. To support researchers to obtain an up-to-date overview of a research field, the system should maintain such surveys in a structured way, and allow for dynamics and evolution. A researcher interested in such an overview should be able to search or to browse the desired research field in a user interface for ORKG access. Then, the system should retrieve related articles and available overviews, e.g. in a table or a leaderboard chart.

While an ORKG user interface should allow for showing tabular leaderboards or other visual representations, the backend should semantically represent information to allow for the exploitation of overlaps in conceptualisations between research problems or fields. Furthermore, faceted drill-down methods based on the properties of semantic descriptions of research approaches could empower researchers to quickly filter and zoom into the most relevant literature.

Find related work: Finding relevant research articles is a daily core activity of researchers. The primary goal of this use case is to find research articles which are relevant to a certain research question. A broad research question is often broken down into smaller, more specific sub-questions which are then converted to search queries [42]. For instance, in this paper, we explored the following sub-questions: (a) *Which ontologies do exist to represent scholarly knowledge?* (b) *Which scientific knowledge graphs do exist and which*

[2] https://projects.tib.eu/orkg/project/team/, https://www.tib.eu/en/research-development/visual-analytics/staff.

information do they contain? (c) *Which datasets do exist for scientific information extraction?* (d) *What are current state-of-the-art methods for scientific information extraction?* (e) *Which approaches do exist to construct a knowledge graph?*

An ORKG should support the answering of queries related to such questions, which can be fine-grained or broad search intents. Preferably, the system should support natural language queries as approached by semantic search and question answering engines [6]. The system has to return a set of relevant articles.

Assess relevance: Given a set of relevant articles the researcher has to assess whether the articles match the criteria of interest. Usually researchers skim through the title and abstract. Often, also the introduction and conclusions have to be considered, which is cumbersome and time-consuming. If only the most important paragraphs in the article are presented to the researcher in a structured way, this process can be boosted. Such information snippets might include, for instance, text passages that describe the problem tackled in the research work, the main contributions, the employed methods or materials, or the yielded results.

Extract relevant information: To tackle a particular research question, the researcher has to extract relevant information from research articles. In a systematic literature review, the information to be extracted can be defined through a *data extraction form* (see Sect. 2.5). Such extracted information is usually compiled in written text or comparison tables in a related work section or survey articles. For instance, for the question *"Which datasets do exist for scientific sentence classification?"* a researcher who focuses on a new annotation study could be interested in (a) domains covered by the dataset and (b) the inter-coder agreement (see Table 2 as an example). Another researcher might follow the same question but focusing on machine learning and thus could be more interested in (c) evaluation results and (d) feature types used.

The system should support the researcher with tailored information extraction from a set of research articles: (1) The researcher defines a data extraction form as proposed in systematic literature reviews (e.g. the fields (a)–(d)), and (2) the system presents the extracted information as suggestions for the corresponding data extraction form and articles in a comparative table. Figure 3 illustrates a data extraction form with corresponding fields in form of questions, and a possible approach to visualise the extracted text passages from the articles for the respective fields in a tabular form.

Get recommended articles: When the researcher focuses on a particular article, further related articles could be recommended by the system utilising an ORKG, for instance, articles that address the same research problem or apply similar methods.

Obtain deep understanding: The system should help the researcher to obtain a deep understanding of a research article (e.g. equations, algorithms, diagrams, datasets). For this purpose, the system should connect the article with artefacts such as conference videos, presentations, source code, datasets, etc., and visualise the artefacts appropriately. Also text passages can be linked, e.g. explanations of methods in Wikipedia, source code snippets of an algorithm implementation, or equations described in the article.

Reproduce results: The system should offer researchers links to all necessary artefacts to help to reproduce research results, e.g. datasets, source code, virtual research environments, materials describing the study, etc. Furthermore, the system should maintain semantic descriptions of domain-specific and standardised evaluation protocols and guidelines such as in machine learning reproducibility checklists [86] and bioassays in the medical domain.

3.2 Knowledge graph requirements

As outlined in Sect. 2.4, data quality requirements should be considered within the context of a particular use case ("fitness for use"). In this section, we first describe dimensions we used to define non-functional requirements for an ORKG. Then, we discuss these requirements within the context of our identified use cases.

3.2.1 Dimensions for KG requirements

In the following, we describe the dimensions that we use to define the requirements for ontology design and instance data. We selected these dimensions since we assume that they are most relevant and also challenging to construct an ORKG with appropriate data to support the various use cases.

For *ontology design*, i.e. how comprehensively should an ontology conceptualise scholarly knowledge to support a certain use case, we use the following dimensions:

(A) *Domain specialisation of the ontology:* How domain-specific should the concepts and relation types be in the ontology? An ontology with *high domain specialisation* targets a specific (sub-)domain and uses domain-specific terms. An ontology with *low domain specialisation* targets a broad range of domains and uses rather domain-independent terms. For instance, various ontologies (e.g. [15,84]) propose domain-independent concepts (e.g. process, method, material). In contrast, Klampanos et al. [62] present a very domain-specific ontology for artificial neural networks.

(B) *Granularity of the ontology:* Which granularity of the ontology is required to conceptualise scholarly knowledge? An ontology with *high granularity* conceptualises scholarly knowledge with a lot of classes that have very detailed and a lot of fine-grained properties and relations. An ontology with a *low granularity* has only a few classes

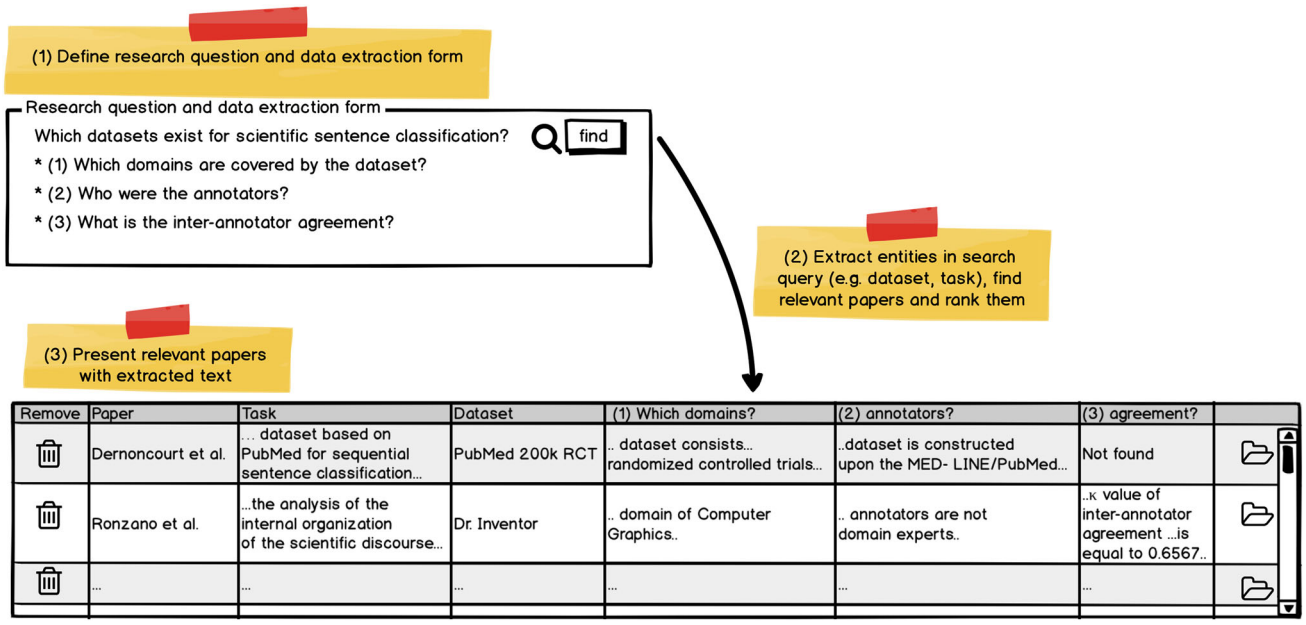

Fig. 3 An example research questions with a corresponding data extraction form, and the extracted text passages from relevant research articles for the respective (data extraction form) fields presented in a tabular form

and relation types. For instance, the annotation schemes for scientific corpora (see Sect. 2.3) have a rather low granularity, as they do not have more than 10 classes and 10 relation types. In contrast, various ontologies (e.g. [50,84]) with more than 20 to 35 classes and over 20 to 70 relations and properties are fine-grained and have a relatively high granularity.

Although there is usually a correlation between domain specialisation and granularity of the ontology (e.g. an ontology with high domain specialisation has also a high granularity), there exist also rather domain-independent ontologies with a high granularity, e.g. scholarly ontology [84]), and ontologies with high domain specialisation and low granularity, e.g. the PICO criterion in Evidence-Based Medicine [60,92]) which stands for population (P), intervention (I), comparison (C), and outcome (O). Thus, we use both dimensions independently. Furthermore, a high domain specialisation requirement for a use case implies that each sub-domain requires a separate ontology for the specific use case. These domain-specific ontologies can be organised in a taxonomy.

For the *instance data*, we use the following dimensions:

(C) *Completeness of the instance data:* Given an ontology, to which extent do *all* possible instances (i.e. instances for classes and facts for relation types) in *all* research articles have to be represented in the KG? *Low completeness:* it is tolerable for the use case when a considerable amount of instance data is missing for the respective ontology. *High completeness:* it is mandatory for the use case that for the respective ontology, a considerable amount of instances are present in the instance data. For instance, given an ontology with a class "Task" and a relation type "subTaskOf" to describe a taxonomy of tasks, the instance data for that ontology would be complete if all tasks mentioned in all research articles are present (population completeness) and "subTaskOf" facts between the tasks are not missing (column completeness).

(D) *Correctness of the instance data:* Given an ontology, which correctness is necessary for the corresponding instances? *Low correctness:* it is tolerable for the use case, that some instances (e.g. 30%) are not correct. *High correctness:* it is mandatory for the use case, that instance data must not be wrong i.e. all present instances in the KG must conform to the ontology and reflect the content of the research articles properly. For instance, an article is correctly assigned to the task addressed in the article, the F1 score in the evaluation results are correctly extracted, etc.

It should be noted that completeness and correctness of instance data can be evaluated only for a given ontology. For instance, let A be an ontology having the class "Deep Learning Model" without properties, and let B be an ontology that also has a class "Deep Learning Model" and additionally further relation types describing the properties of the deep learning model (e.g. drop-out, loss functions, etc.). In this example, the instance data of ontology A would be considered to have high completeness, if it covers most of the important deep learning models. However, for ontology B,

Table 1 Requirements and approaches for the main use cases

		Extract relevant info	Research field overview	Deep understanding	Reproduce results	Find related work	Recommend articles	Assess relevance
Ontology	Domain specialisation	high	high	med	med	low	low	med
	Granularity	high	high	med	med	low	low	low
Instance data	Completeness	low	med	low	med	high	high	med
	Correctness	med	high	high	high	low	low	med
	Maintain terminologies	–	X	–	–	X	X	–
	Define templates	X	X	–	–	–	–	–
	Fill in templates	X	X	X	X	–	–	–
	Maintain overviews	X	X	–	–	–	–	–
Automatic curation	Entity/relation extraction	(x)	(x)	(x)	(x)	X	X	X
	Entity linking	(x)	(x)	(x)	(x)	X	X	X
	Sentence classification	(x)	–	(x)	–	X	–	X
	Template-based extraction	(x)	(x)	(x)	(x)	–	–	–
	Cross-modal linking	–	–	(x)	(x)	–	–	–

The upper part describes the minimum requirements for the ontology (domain specialisation and granularity) and the instance data (completeness and correctness). The bottom part lists possible approaches for manual, automatic, and semi-automatic curation of the KG for the respective use cases. "X" indicates that the approach is suitable for the use case, while "(x)" denotes that the approach is only appropriate with human supervision. The left part (delimited by the vertical triple line) groups use cases suitable for manual, and the right side for automatic approaches. Vertical double lines group use cases with similar requirements

the completeness of the same instance data would be rather low since the properties of the deep learning models are missing. The same holds for correctness: If ontology B has, for instance, a sub-type "Convolutional Neural Network", then the instance data would have a rather low correctness for ontology B if all "Deep Learning Model" instances are typed only with the generic class "Deep Learning Model".

3.2.2 Discussion of the KG requirements

Next, we discuss the seven main use cases with regard to the required level of ontology domain specialisation and granularity, as well as completeness and correctness of instance data. Table 1 summarises the requirements for the use cases along the four dimensions at ordinal scale. The use cases are grouped together, when they have (1) similar justifications for the requirements and (2) a high overlap in ontology concepts and instances.

Extract relevant information & get research field overview: The information to be extracted from relevant research articles for a data extraction form within a literature review is very heterogeneous and depends highly on the intent of the researcher and the research questions. Thus, the ontology has to be domain-specific and fine-grained to offer all possible kinds of desirable information. However, missing information for certain questions in the KG may be tolerable for a researcher. Furthermore, it is tolerable for a researcher if some of the extracted suggestions are wrong since the researcher can correct them.

Research field overviews are usually the result of a literature review. The data in such an overview has also to be very domain-specific and fine-grained. Also, this information must have high correctness, e.g. an F1 score of an evaluation result must not be wrong. Furthermore, an overview of a particular research field should have appropriate completeness and must not miss any relevant research papers. However, it is acceptable when overviews for some research fields are missing.

Obtain deep understanding & reproduce results: The information required for these use cases has to achieve a high level of correctness (e.g. accurate links to dataset, source code, videos, articles, research infrastructures). An ontology for the representation of default artefacts can be rather domain-independent (e.g. Scholix [20]). However, semantic representation of evaluation protocols requires domain-dependent ontologies (e.g. EXPO [99]). Missing information is tolerable for these use cases.

Find related work & get recommended articles: When searching for related work, it is essential not to miss relevant articles. Previous studies revealed that more than half of search queries in academic search engines refer to scientific entities [112]. However, the coverage of scientific entities in general-purpose KGs (e.g. WikiData) is rather low, since the introduction of new concepts in research literature occurs at a faster pace than KG curation [1]. Despite the low completeness, Xiong et al. [112] could improve the ranking of search results in academic search engines by exploiting general-purpose KGs. Hence, the instance data for the "find related work" use case should have high completeness with fine-grained scientific entities. However, semantic search engines leverage latent representations of KGs and text (e.g. graph and word embeddings) [6]. Since a non-perfect ranking of the search results is tolerable for a researcher, lower correctness of the instance data could be acceptable. Furthermore, due to latent feature representations, the ontology can be kept rather simple and domain-independent. For instance, the STM corpus [15] introduces four domain-independent concepts.

Graph- and content-based research paper recommendation systems [8] have similar requirements since they also leverage latent feature representations and require fine-grained scientific entities. Also, non-perfect recommendations are tolerable for a researcher.

Assess relevance: To help the researcher to assess the relevance of an article according to her needs, the system should highlight the most essential zones in the article to get a quick overview. The completeness and correctness of the presented information must not be too low, as otherwise the user acceptance may suffer. However, it can be suboptimal, since it is acceptable for a researcher when some of the highlighted information is not essential or when some important information is missing. The ontology to represent essential information should be rather domain-specific (i.e. using terms that the researchers understand) and quite simple (cf. ontologies for scientific sentence classification in Sect. 2.3.2).

4 ORKG construction strategies

In this section, we discuss the implications for the design and construction of an ORKG and outline possible approaches, which are mapped to the use cases in Table 1. Based on the discussion in the previous section, we can subdivide the use cases into two groups: (1) requiring high correctness and high domain specialisation with rather low requirements on the completeness (left side in Table 1) and (2) requiring high completeness with rather low requirements on the correctness and domain specialisation (right side in Table 1). The first group requires manual approaches, while the second group could be accomplished with fully automatic approaches. To ensure trustworthiness, data records should contain provenance information, i.e. who or what system curated the data.

Manually curated data can also support use cases with automatic approaches, and vice versa. Furthermore, automatic approaches can complement manual approaches by providing suggestions in user interfaces. Such synergy

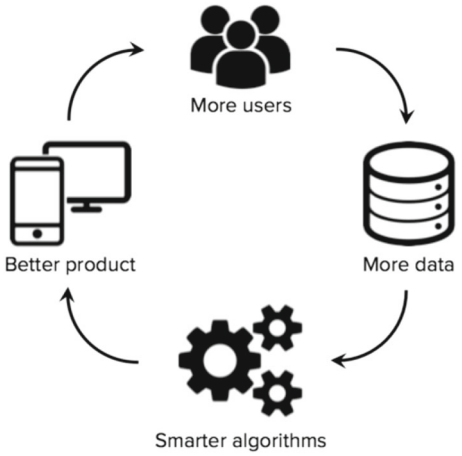

Fig. 4 The virtuous cycle of data network effects by combining manual and automatic data curation approaches [22]

between humans and algorithms may lead to a "data flywheel" (also known as data network effects, see Fig. 4): Users produce data which enable to build a smarter product with better algorithms so that more users use the product and thus produce more data, and so on.

4.1 Manual approaches

Ontology design: The first group of use cases requires rather domain-specific and fine-grained ontologies. We suggest to develop novel or reuse ontologies that fit the respective use case and the specific domain (e.g. EXPO [99] for experiments). Moreover, appropriate and simple user interfaces are necessary for efficient and easy population.

However, such ontologies can evolve with the help of the community, as demonstrated by WikiData and Wikipedia with "infoboxes" (see Sect. 2.3). Therefore, the system should enable the maintenance of *templates*, which are predefined and very specific forms consisting of fields with certain types (see Fig. 5). For instance, to automatically generate leaderboards for machine learning tasks, a template would have the fields task, model, dataset, and score, which can then be filled in by a curator for articles providing such kind of results in a user interface generated from the template. Such an approach is based on *meta-modelling* [13], as the meta-model for templates enables the definition of concrete templates, which are then instantiated for articles.

Knowledge graph population: Several user interfaces are required to enable manual population: (1) populate semantic content for a research article by (1a) choosing relevant templates or ontologies and (1b) fill in the values; (2) terminology management (e.g. domain-specific research fields); (3) maintain research field overviews by (3a) assigning relevant research articles to the research field, (3b) define corresponding templates, and (3c) fill in the templates for the relevant research articles.

Furthermore, the system should also offer *Application Programming Interfaces (APIs)* to enable population by third-party applications, e.g.:

- Submission portals such as https://www.easychair.org/ during submission of an article.
- Authoring tools such as https://www.overleaf.com/ during writing.
- Virtual research environments [101] to store evaluation results and links to datasets and source code during experimenting and data analysis.

To encourage stakeholders like researchers, librarians, crowd workers to contribute content, we see the following options:

- *Top-down enforcement* via submission portals and publishers.
- *Incentive models*: Researchers want their articles to be cited; semantic content helps other researchers to find, explore and understand an article. This is also related to the concept of *enlightened self-interest*, i.e. act to further interests of others to serve the own self-interest.
- Provide *public acknowledgements* for curators.
- Bring together *experts* (e.g. librarians, researchers from different institutions) who curate and organise content for specific research problems or disciplines.

4.2 (Semi-)automatic approaches

Ontology design: The second group of use cases require a high completeness, while a relatively low correctness and domain specialisation are acceptable. For these use cases, rather simple or domain-independent ontologies should be developed or reused. Although approaches for automatic ontology learning exist (see Sect. 2.3), the quality of their results is not sufficient to generate a meaningful ORKG with complex conceptual models and relations. Therefore, meaningful ontologies should be designed by human experts.

Knowledge graph population: Various approaches can be used to (semi-)automatically populate an ORKG. Methods for *entity and relation extraction* (see Sect. 2.3) can help to populate fine-grained KGs with high completeness and *entity linking* approaches can link mentions in text with entities in KGs. For cross-modal linking, Singh et al. [97] suggest an approach to detect URLs to datasets in research articles automatically, while the Scientific Software Explorer [52] connects text passages in research articles with code fragments. To extract relevant information at sentence level, approaches for *sentence classification* in scientific text can be applied (see Sect. 2.3). To support the curator fill in templates semi-automatically, *template-based extraction* can (1) suggest relevant templates for a research article and (2) pre-fill fields of templates with appropriate values. For pre-filling,

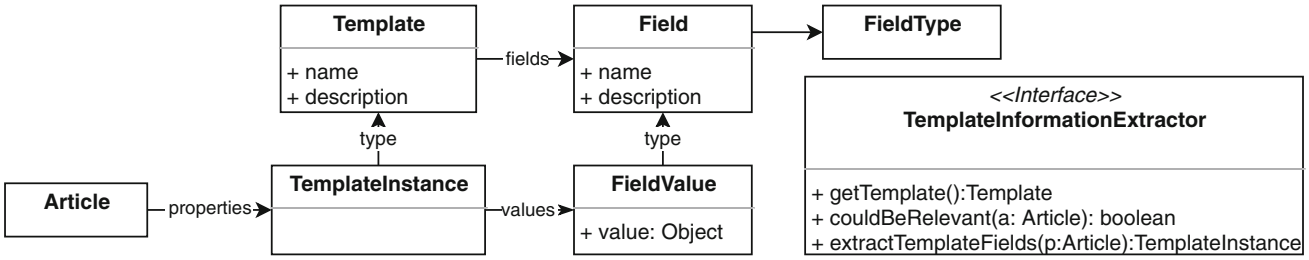

Fig. 5 Conceptual meta-model in UML for templates and interface design for an external template-based information extractor

approaches such as *n*-ary relation extraction [44,54,57,59] or end-to-end question answering [33,91] could be applied.

Furthermore, the system should enable to plugin *external information extractors*, developed for certain scientific domains to extract specific types of information. For instance, as depicted in Fig. 5, an external template information extractor has to implement an interface with three methods. This enables the system (1) to filter relevant template extractors for an article and (2) extract field values from an article.

5 Conclusions and future work

In this paper, we have presented a requirements analysis for an Open Research Knowledge Graph (ORKG). An ORKG should represent the content of research articles in a semantic way to enhance or enable a wide range of use cases. We identified literature-related core tasks of a researcher that can be supported by an ORKG and formulated them as use cases. For each use case, we discussed specificities and requirements for the underlying ontology and the instance data. In particular, we identified two groups of use cases: (1) the first group requires instance data with high correctness and rather fine-grained, domain-specific ontologies, but with moderate completeness; (2) the second group requires a high completeness, but the ontologies can be kept rather simple and domain-independent, and a moderate correctness of the instance data is sufficient. Based on the requirements, we have described possible manual and semi-automatic approaches (necessary for the first group), and automatic approaches (appropriate for the second group) for KG construction. In particular, we propose a framework with lightweight ontologies that can evolve by community curation. Furthermore, we have described the interdependence with external systems, user interfaces, and APIs for third-party applications to populate an ORKG.

The results of our work aim to give a holistic view of the requirements for an ORKG and guide further research. The suggested approaches have to be refined, implemented, and evaluated in an iterative and incremental process (see www.orkg.org for the current progress). Users from different scientific domains should be deeply involved in the development process to build proper solutions. Furthermore, since ontologies and instance data will evolve in the ORKG, solutions are required to adequately support this evolution process (e.g. editing, versioning, support to report inconsistencies, etc.). Finally, our analysis can serve as a foundation for a discussion on ORKG requirements with other researchers and practitioners.

Acknowledgements This work was co-funded by the European Research Council for the Project ScienceGRAPH (Grant agreement ID: 819536).

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Comparative overviews for information extraction datasets from scientific text

Tables 2, 3, and 4 show comparative overviews for some datasets from research papers of various disciplines for the tasks sentence classification, relation extraction, and concept extraction, respectively. These comparisons aim to provide an overview of the current state of the art on information extraction methods in scientific texts to give a sense about which kind of information can be extracted with what accuracy.

Table 2 Characteristics of datasets and performance measures for sentence classification in research papers

Dataset	Domains	# Papers	Coverage	Sentence classes	Inter-coder agreement	Performance
PubMed-20k [31]	Biomedicine	20, 000	Abstracts	Background Objective Methods Results, Conclusion	n/a	92.9% F1 [24]
NICTA-PIBOSO [60]	Biomedicine	1000	Abstracts	Background Intervention Study Population Outcome, Other	62.0% k	84.7% F1 [24]
CSABSTRUCT [24]	Computer Science	2189	Abstracts	Background Objective Method Result, Other	75.0% k	83.1% F1 [24]
CS-Abstracts [48]	Computer Science	654	Abstracts	Background Objective Methods Results, Conclusions	n/a	74.6% F1 [48]
Emerald 100k [100]	Management Information Science Engineering	103, 457	Abstracts	Purpose Design/methodology/approach Findings Originality/value Social implications Practical implications Research limitations/implications	n/a	n/a

Table 2 continued

Dataset	Domains	# Papers	Coverage	Sentence classes	Inter-coder agreement	Performance
MAZEA [28]	Physics Engineering Life and Health Sciences	1335	Abstracts	Background Gap, Purpose Method Result, Conclusion	59,4% k	66.0% accuracy [28]
Safder et al. [94]	Computer Science	92	Full text	Algorithmic efficiency Dataset description Algorithmic time complexity Other	n/a	78.5% accuracy [94]
Dr. Inventor [43]	Computer Graphics	40	Full text	Background Challenge Approach Outcome, Future work	66.7% k	72.5% accuracy [5]
ART/CoreSC [69]	Chemistry Computational Linguistic	225	Full text	Background Motivation, Goal Hypothesis Object Model, Method Experiment, Result Observation, Conclusion	57.0% k	51.6% F1 [68]

Table 3 Characteristics of datasets and performance measures for binary and *n*-ary relation extraction in research papers

Dataset	Domains	# Papers	Coverage	Cardinality	Relation types	Scope	Inter-coder agreement	# Relations	Performance
SemEval17 [4]	Computer Science Material Sciences Physics	500	Abstract	Binary	Synonym-of Hyponym-of	Intra-sentence	60.0% k	672	28.0% F1 [4]
SemEval18 [90]	Comp. Linguistics	500	Abstract	Binary	Usage Result Model Part-whole Topic Comparison	Intra-sentence	90.8% F1	1595	49.3% F1 [90]
ChemProt [64]	Biomedicine	2482	Abstract	Binary	UPREGULATOR ACTIVATOR DOWNREGULATOR INHIBITOR AGONIST ANTAGONIST SUBSTRATE	Intra-sentence	n/a	10, 031	83.64% F1 [9]
SciERC [71]	Art. Intelligence	500	Abstract	Binary	Hyponym-of Compare Part-of Conjunction Evaluate-for Feature-of Used-for	Cross-sentence	67.8% F1	4716	39.3% F1 [71]
PWC [59]	Art. Intelligence	731	Full text	*n*-ary	(Task, Dataset, Metric,	Document-level	n/a	2295	28.7% F1 [59]

Table 3 continued

Dataset	Domains	# Papers	Coverage	Cardinality	Relation types	Scope	Inter-coder agreement	# Relations	Performance
					Score)				
CKB [57]	Biomedicine	343	Full text	*n*-ary	(Drug, Gene, Mutation)	Document-level	n/a	2025	52.8% F1 [57]
SOFC-Exp [44]	Material Sciences	45	Full text	*n*-ary	(AnodeMaterial, CathodeMaterial, Device, ElectrolyteMaterial, FuelUsed, InterlayerMaterial, OpenCircuitVoltage, PowerDensity, Resistance, WorkingTemperature)	Document-level	n/a	*n/a*	56.4% F1* [44]

*For SOFC-Exp corpus, performance values were obtained with ground truth concept mentions

Table 4 Characteristics of datasets and performance measures for scientific concept extraction in research papers

Dataset	Domains	# Papers	# Concepts	Coverage	Concept types	Inter-coder agreement	Performance
SemEval17 [4]	Computer Science Material Sciences Physics	500	9946	Abstract	Process Task Material	60.0% κ	56.9% F1 [81]
STM [15]	Agriculture Astronomy Biology Chemistry Computer Science Earth Science Engineering Materials Science Mathematics Medicine	110	6127	Abstract	Process Method Material Data	76.0% κ	65.5% F1 [15]
SciERC [71]	Art. Intelligence	500	8089	Abstract	Task Method Metric Material Other Generic	76.9% κ	75.2% F1 [81]
ACL2 [90]	Comp. Linguistics	300	6818	Abstract	Method Tool Language Resource (LR) LR product Model Measures/Measurements Other	63.0% F1	69.9% F1 [81]
B5CDR [67]	Biomedicine	1500	28, 785	Abstract	Chemical Disease	91.8% F1	88.9% F1 [9]
NCBI-disease [35]	Biomedicine	793	6892	Abstract	Disease	88.0% F1	96.9% F1 [9]
SOFC-Exp [44]	Material Sciences	45	4004	Full text	Material Device Value	95.8% F1	81.5% F1* [44]

* For SOFC-Exp corpus, performance values were obtained with ground truth sentences describing experiments

References

1. Ammar, W., Groeneveld, D., Bhagavatula, C., Beltagy, I., Crawford, M., Downey, D., Dunkelberger, J., Elgohary, A., Feldman, S., Ha, V., Kinney, R., Kohlmeier, S., Lo, K., Murray, T., Ooi, H., Peters, M.E., Power, J., Skjonsberg, S., Wang, L.L., Wilhelm, C., Yuan, Z., van Zuylen, M., Etzioni, O.: Construction of the literature graph in semantic scholar. In: Bangalore, S., Chu-Carroll, J., Li, Y. (eds.) Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1–6, 2018, vol. 3 (Industry Papers), pp. 84–91. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/n18-3011
2. Aryani, A., Wang, J.: Research graph: Building a distributed graph of scholarly works using research data switchboard. Open Repos. Conf. (2017). https://doi.org/10.4225/03/58c696655af8a
3. Auer, S., Mann, S.: Towards an open research knowledge graph. Ser. Libr. **76**(1–4), 35–41 (2019). https://doi.org/10.1080/0361526X.2019.1540272
4. Augenstein, I., Das, M., Riedel, S., Vikraman, L., McCallum, A.: Semeval 2017 task 10: Scienceie—xtracting keyphrases and relations from scientific publications. In: Bethard, S., Carpuat, M., Apidianaki, M., Mohammad, S.M., Cer, D.M., Jurgens, D. (eds.) Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval@ACL 2017, Vancouver, Canada, 2017, pp. 546–555. Association for Computational Linguistics (2017). https://doi.org/10.18653/v1/S17-2091
5. Badie, K., Asadi, N., Mahmoudi, M.T.: Zone identification based on features with high semantic richness and combining results of separate classifiers. J. Inf. Telecommun. **2**(4), 411–427 (2018). https://doi.org/10.1080/24751839.2018.1460083
6. Balog, K.: Entity-Oriented Search. Springer, Berlin (2018). https://doi.org/10.1007/978-3-319-93935-3
7. Bechhofer, S., Buchan, I.E., Roure, D.D., Missier, P., Ainsworth, J.D., Bhagat, J., Couch, P.A., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Michaelides, D.T., Owen, S., Newman, D.R., Sufi, S., Goble, C.A.: Why linked data is not enough for scientists. Future Gener. Comput. Syst. **29**(2), 599–611 (2013). https://doi.org/10.1016/j.future.2011.08.004
8. Beel, J., Gipp, B., Langer, S., Breitinger, C.: Research-paper recommender systems: a literature survey. Int. J. Digit. Libr. **17**(4), 305–338 (2016). https://doi.org/10.1007/s00799-015-0156-0
9. Beltagy, I., Lo, K., Cohan, A.: SciBERT: a pretrained language model for scientific text. In: Inui, K., Jiang, J., Ng, V., Wan, X. (eds.) Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, 2019, pp. 3613–3618. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/D19-1371
10. Bizer, C.: Quality-Driven Information Filtering—In the Context of Web-Based Information Systems. VDM Verlag, Saarbrücken (2007)
11. Bodenreider, O.: The unified medical language system (UMLS): integrating biomedical terminology. Nucl. Acids Res. **32**, 267–270 (2004). https://doi.org/10.1093/nar/gkh061
12. Bollacker, K.D., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: Wang, J.T. (ed.) Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, 2008, pp. 1247–1250. ACM (2008). https://doi.org/10.1145/1376616.1376746
13. Booch, G., Rumbaugh, J., Jacobson, I.: Unified Modeling Language User Guide, The (2nd Edition) (Addison-Wesley Object Technology Series). Addison-Wesley Professional, Boston (2005)
14. Bornmann, L., Mutz, R.: Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. J. Assoc. Inf. Sci. Technol. **66**(11), 2215–2222 (2015). https://doi.org/10.1002/asi.23329
15. Brack, A., D'Souza, J., Hoppe, A., Auer, S., Ewerth, R.: Domain-independent extraction of scientific concepts from research articles. In: Jose, J.M., Yilmaz, E., Magalhães, J., Castells, P., Ferro, N., Silva, M.J., Martins, F. (eds.) Advances in Information Retrieval—42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, 2020, Proceedings, Part I, Lecture Notes in Computer Science, vol. 12035, pp. 251–266. Springer (2020). https://doi.org/10.1007/978-3-030-45439-5_17
16. Brack, A., Hoppe, A., Stocker, M., Auer, S., Ewerth, R.: Requirements analysis for an open research knowledge graph. In: Hall, M.M., Mercun, T., Risse, T., Duchateau, F. (eds.) Digital Libraries for Open Knowledge—24th International Conference on Theory and Practice of Digital Libraries, TPDL 2020, Lyon, France, 2020, Proceedings, Lecture Notes in Computer Science, vol. 12246, pp. 3–18. Springer (2020). https://doi.org/10.1007/978-3-030-54956-5_1
17. Brack, A., Müller, D.U., Hoppe, A., Ewerth, R.: Coreference resolution in research papers from multiple domains. In: Hiemstra, D., Moens, M., Mothe, J., Perego, R., Potthast, M., Sebastiani, F. (eds.) Advances in Information Retrieval—43rd European Conference on IR Research, ECIR 2021, Virtual Event, 2021, Proceedings, Part I, Lecture Notes in Computer Science, vol. 12656, pp. 79–97. Springer (2021). https://doi.org/10.1007/978-3-030-72113-8_6
18. Braun, R., Benedict, M., Wendler, H., Esswein, W.: Proposal for requirements driven design science research. In: Donnellan, B., Helfert, M., Kenneally, J., VanderMeer, D.E., Rothenberger, M.A., Winter, R. (eds.) New Horizons in Design Science: Broadening the Research Agenda—10th International Conference, DESRIST 2015, Dublin, Ireland, 2015, Proceedings, Lecture Notes in Computer Science, vol. 9073, pp. 135–151. Springer (2015). https://doi.org/10.1007/978-3-319-18714-3_9
19. Brodaric, B., Reitsma, F., Qiang, Y.: Skiing with DOLCE: toward an e-science knowledge infrastructure. In: Eschenbach, C., Grüninger, M. (eds.) Formal Ontology in Information Systems, Proceedings of the Fifth International Conference, FOIS 2008, Saarbrücken, Germany, 2008, Frontiers in Artificial Intelligence and Applications, vol. 183, pp. 208–219. IOS Press (2008). https://doi.org/10.3233/978-1-58603-923-3-208
20. Burton, A., Aryani, A., Koers, H., Manghi, P., Bruzzo, S.L., Stocker, M., Diepenbroek, M., Schindler, U., Fenner, M.: The scholix framework for interoperability in data-literature information exchange. D-Lib Mag. **23**(1/2), 1–20 (2017). https://doi.org/10.1045/january2017-burton
21. Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Jr., E.R.H., Mitchell, T.M.: Toward an architecture for never-ending language learning. In: Fox, M., Poole, D. (eds.) Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, 2010. AAAI Press (2010). http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1879
22. CB Insights: The data flywheel: how enlightened self-interest drives data network effects. https://www.cbinsights.com/research/team-blog/data-network-effects/ (2020)

23. Cohan, A., Ammar, W., van Zuylen, M., Cady, F.: Structural scaffolds for citation intent classification in scientific publications. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, 2019, vol. 1 (Long and Short Papers), pp. 3586–3596. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/n19-1361
24. Cohan, A., Beltagy, I., King, D., Dalvi, B., Weld, D.S.: Pretrained language models for sequential sentence classification. In: Inui, K., Jiang, J., Ng, V., Wan, X. (eds.) Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, 2019, pp. 3691–3697. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/D19-1383
25. Cohen, K.B., Lanfranchi, A., Choi, M.J., Baumgartner, W.A., Panteleyeva, N., Verspoor, K., Palmer, M., Hunter, L.E.: Coreference annotation and resolution in the Colorado richly annotated full text (CRAFT) corpus of biomedical journal articles. BMC Bioinform. **18**(1), 1–14 (2017). https://doi.org/10.1186/s12859-017-1775-9
26. Consortium, T.G.O., Consortium: The gene ontology resource: 20 years and still going strong. Nucl. Acids Res. **47**, D330–D338 (2019). https://doi.org/10.1093/nar/gky1055
27. Constantin, A., Peroni, S., Pettifer, S., Shotton, D.M., Vitali, F.: The document components ontology (DoCo). Semant. Web **7**(2), 167–181 (2016). https://doi.org/10.3233/SW-150177
28. Dayrell, C., Jr., A.C., Lima, G., Jr., D.M., Copestake, A.A., Feltrim, V.D., Tagnin, S.E.O., Aluísio, S.M.: Rhetorical move detection in english abstracts: multi-label sentence classifiers and their annotated corpora. In: Calzolari, N., Choukri, K., Declerck, T., Dogan, M.U., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S. (eds.) Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, 2012, pp. 1604–1609. European Language Resources Association (ELRA) (2012). http://www.lrec-conf.org/proceedings/lrec2012/summaries/734.html
29. Degbelo, A.: A snapshot of ontology evaluation criteria and strategies. In: Hoekstra, R., Faron-Zucker, C., Pellegrini, T., de Boer, V. (eds.) Proceedings of the 13th International Conference on Semantic Systems, SEMANTICS 2017, Amsterdam, The Netherlands, 2017, pp. 1–8. ACM (2017). https://doi.org/10.1145/3132218.3132219
30. Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M., Ashburner, M.: Chebi: a database and ontology for chemical entities of biological interest. Nucl. Acids Res. **36**, 344–350 (2008). https://doi.org/10.1093/nar/gkm791
31. Dernoncourt, F., Lee, J.Y.: Pubmed 200k RCT: a dataset for sequential sentence classification in medical abstracts. In: Kondrak, G., Watanabe, T. (eds.) Proceedings of the Eighth International Joint Conference on Natural Language Processing, IJCNLP 2017, Taipei, Taiwan, 2017, Volume 2: Short Papers, pp. 308–313. Asian Federation of Natural Language Processing (2017). https://www.aclweb.org/anthology/I17-2052/
32. Dessì, D., Osborne, F., Recupero, D.R., Buscaldi, D., Motta, E., Sack, H.: AI-KG: an automatically generated knowledge graph of artificial intelligence. In: Pan, J.Z., Tamma, V.A.M., d'Amato, C., Janowicz, K., Fu, B., Polleres, A., Seneviratne, O., Kagal, L. (eds.) The Semantic Web—ISWC 2020—19th International Semantic Web Conference, Athens, Greece, 2020, Proceedings, Part II, Lecture Notes in Computer Science, vol. 12507, pp. 127–143. Springer (2020). https://doi.org/10.1007/978-3-030-62466-8_9
33. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, 2019, vol. 1 (Long and Short Papers), pp. 4171–4186. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/n19-1423
34. Doerr, M., Kritsotaki, A., Rousakis, Y., Hiebel, G., Theodoridou, M.: Definition of the CRMsci: an extension of CIDOC-CRM to support scientific observation. Tech. rep., FORTH, Version 1.2.8. http://www.cidoc-crm.org/crmsci/ModelVersion/version-1.2.8 (2020)
35. Dogan, R.I., Leaman, R., Lu, Z.: NCBI disease corpus: a resource for disease name recognition and concept normalization. J. Biomed. Inform. **47**, 1–10 (2014). https://doi.org/10.1016/j.jbi.2013.12.006
36. Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., Zhang, W.: Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: Macskassy, S.A., Perlich, C., Leskovec, J., Wang, W., Ghani, R. (eds.) The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA-2014, pp. 601–610. ACM (2014). https://doi.org/10.1145/2623330.2623623
37. D'Souza, J., Hoppe, A., Brack, A., Jaradeh, M.Y., Auer, S., Ewerth, R.: The STEM-ECR dataset: grounding scientific entity references in STEM scholarly content to authoritative encyclopedic and lexicographic sources. In: Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, 2020, pp. 2192–2203. European Language Resources Association (2020). https://www.aclweb.org/anthology/2020.lrec-1.268/
38. Färber, M.: The microsoft academic knowledge graph: A linked data source with 8 billion triples of scholarly data. In: Ghidini, C., Hartig, O., Maleshkova, M., Svátek, V., Cruz, I.F., Hogan, A., Song, J., Lefrançois, M., Gandon, F. (eds.) The Semantic Web—ISWC 2019—18th International Semantic Web Conference, Auckland, New Zealand,, 2019, Proceedings, Part II, Lecture Notes in Computer Science, vol. 11779, pp. 113–129. Springer (2019). https://doi.org/10.1007/978-3-030-30796-7_8
39. Färber, M., Bartscherer, F., Menne, C., Rettinger, A.: Linked data quality of DBpedia, Freebase, Opencyc, Wikidata, and YAGO. Semant. Web **9**(1), 77–129 (2018). https://doi.org/10.3233/SW-170275
40. Fathalla, S., Vahdati, S., Auer, S., Lange, C.: Towards a knowledge graph representing research findings by semantifying survey articles. In: Kamps, J., Tsakonas, G., Manolopoulos, Y., Iliadis, L.S., Karydis, I. (eds.) Research and Advanced Technology for Digital Libraries—21st International Conference on Theory and Practice of Digital Libraries, TPDL 2017, Thessaloniki, Greece, 2017, Proceedings, Lecture Notes in Computer Science, vol. 10450, pp. 315–327. Springer (2017). https://doi.org/10.1007/978-3-319-67008-9_25
41. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. Language, Speech, and Communication. MIT Press, Cambridge (1998)
42. Fink, A.: Conducting Research Literature Reviews: From the Internet to Paper. SAGE Publications, Thousand Oaks (2014)
43. Fisas, B., Saggion, H., Ronzano, F.: On the discoursive structure of computer graphics research papers. In: Meyers, A., Rehbein, I., Zinsmeister, H. (eds.) Proceedings of The 9th Linguistic Annotation Workshop, LAW@NAACL-HLT 2015, 2015, Denver, Colorado, USA, pp. 42–51. The Association for Computer Linguistics (2015). https://doi.org/10.3115/v1/w15-1605

44. Friedrich, A., Adel, H., Tomazic, F., Hingerl, J., Benteau, R., Marusczyk, A., Lange, L.: The sofc-exp corpus and neural approaches to information extraction in the materials science domain. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, 2020, pp. 1255–1268. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.116
45. Gábor, K., Buscaldi, D., Schumann, A., Qasemi Zadeh, B., Zargayouna, H., Charnois, T.: Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers. In: Apidianaki, M., Mohammad, S.M., May, J., Shutova, E., Bethard, S., Carpuat M. (eds.) Proceedings of The 12th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2018, New Orleans, Louisiana, USA, 2018, pp. 679–688. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/s18-1111
46. Galárraga, L., Razniewski, S., Amarilli, A., Suchanek, F.M.: Predicting completeness in knowledge bases. In: de Rijke, M., Shokouhi, M., Tomkins, A., Zhang, M. (eds.) Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, United Kingdom, 2017, pp. 375–383. ACM (2017). https://doi.org/10.1145/3018661.3018739
47. Galárraga, L.A., Teflioudi, C., Hose, K., Suchanek, F.M.: AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In: Schwabe, D., Almeida, V.A.F., Glaser, H., Baeza-Yates, R., Moon, S.B. (eds.) 22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, 2013, pp. 413–422. International World Wide Web Conferences Steering Committee. ACM (2013). https://doi.org/10.1145/2488388.2488425
48. Gonçalves, S., Cortez, P., Moro, S.: A deep learning classifier for sentence classification in biomedical and computer science abstracts. Neural Comput. Appl. **32**(11), 6793–6807 (2020). https://doi.org/10.1007/s00521-019-04334-2
49. Groza, T., Handschuh, S., Möller, K., Decker, S.: SALT—semantically annotated latex for scientific publications. In: Franconi, E., Kifer, M., May, W. (eds.) The Semantic Web: Research and Applications, 4th European Semantic Web Conference, ESWC 2007, Innsbruck, Austria, 2007, Proceedings, Lecture Notes in Computer Science, vol. 4519, pp. 518–532. Springer (2007). https://doi.org/10.1007/978-3-540-72667-8_37
50. Hars, A.: Structure of Scientific Knowledge, pp. 83–185. Springer, Berlin (2003). https://doi.org/10.1007/978-3-540-24737-1_3
51. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
52. Hoppe, A., Hagen, J., Holzmann, H., Kniesel, G., Ewerth, R.: An analytics tool for exploring scientific software and related publications. In: Méndez, E., Crestani, F., Ribeiro, C., David, G., Lopes, J.C. (eds.) Digital Libraries for Open Knowledge, 22nd International Conference on Theory and Practice of Digital Libraries, TPDL 2018, Porto, Portugal, 2018, Proceedings, Lecture Notes in Computer Science, vol. 11057, pp. 299–303. Springer (2018). https://doi.org/10.1007/978-3-030-00066-0_27
53. Horvath, I.: Comparison of three methodological approaches of design research. In: S.N. (ed.) Proceedings of the 16th International Conference on Engineering Design, ICED'07, pp. 1–11. Ecole Central Paris (2007). Null; Conference date: 28-08-2007 through 30-08-2007
54. Hou, Y., Jochim, C., Gleize, M., Bonin, F., Ganguly, D.: Identification of tasks, datasets, evaluation metrics, and numeric scores for scientific leaderboards construction. In: Korhonen, A., Traum, D.R., Màrquez, L. (eds.) Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, 2019, vol. 1: Long Papers, pp. 5203–5213. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/p19-1513
55. Jain, S., van Zuylen, M., Hajishirzi, H., Beltagy, I.: Scirex: A challenge dataset for document-level information extraction. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, 2020, pp. 7506–7516. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.670
56. Jaradeh, M.Y., Oelen, A., Prinz, M., Stocker, M., Auer, S.: Open research knowledge graph: a system walkthrough. In: Doucet, A., Isaac, A., Golub, K., Aalberg, T., Jatowt, A. (eds.) Digital Libraries for Open Knowledge—23rd International Conference on Theory and Practice of Digital Libraries, TPDL 2019, Oslo, Norway, 2019, Proceedings, Lecture Notes in Computer Science, vol. 11799, pp. 348–351. Springer (2019). https://doi.org/10.1007/978-3-030-30760-8_31
57. Jia, R., Wong, C., Poon, H.: Document-level n-ary relation extraction with multiscale representation learning. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, 2019, vol. 1 (Long and Short Papers), pp. 3693–3704. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/n19-1370
58. Kannan, A.V., Fradkin, D., Akrotirianakis, I., Kulahcioglu, T., Canedo, A., Roy, A., Yu, S., Malawade, A.V., Faruque, M.A.A.: Multimodal knowledge graph for deep learning papers and code. In: d'Aquin, M., Dietze, S., Hauff, C., Curry, E., Cudré-Mauroux, P. (eds.) CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, 2020, pp. 3417–3420. ACM (2020). https://doi.org/10.1145/3340531.3417439
59. Kardas, M., Czapla, P., Stenetorp, P., Ruder, S., Riedel, S., Taylor, R., Stojnic, R.: Axcell: Automatic extraction of results from machine learning papers. In: Webber, B., Cohn, T., He, Y., Liu, Y. (eds.) Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, 2020, pp. 8580–8594. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.emnlp-main.692
60. Kim, S., Martínez, D., Cavedon, L., Yencken, L.: Automatic classification of sentences to support evidence based medicine. BMC Bioinform. **12**(2), S5 (2011). https://doi.org/10.1186/1471-2105-12-S2-S5
61. Kitchenham, B.A., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. Tech. Rep. EBSE 2007-001, Keele University and Durham University Joint Report. https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf (2007)
62. Klampanos, I.A., Davvetas, A., Koukourikos, A., Karkaletsis, V.: ANNETT-O: an ontology for describing artificial neural network evaluation, topology and training. Int. J. Metadata Semant. Ontol. **13**(3), 179–190 (2019). https://doi.org/10.1504/IJMSO.2019.099833
63. Kolitsas, N., Ganea, O., Hofmann, T.: End-to-end neural entity linking. In: Korhonen, A., Titov, I. (eds.) Proceedings of the 22nd Conference on Computational Natural Language Learning, CoNLL 2018, Brussels, Belgium, 2018, pp. 519–529. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/k18-1050
64. Kringelum, J., Kjærulff, S.K., Brunak, S., Lund, O., Oprea, T.I., Taboureau, O.: Chemprot-3.0: a global chemical biology diseases mapping. Database J. Biol. Databases Curation (2016). https://doi.org/10.1093/database/bav123

65. Lange, C.: Ontologies and languages for representing mathematical knowledge on the semantic web. Semant. Web **4**(2), 119–158 (2013). https://doi.org/10.3233/SW-2012-0059
66. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: Dbpedia—a large-scale, multilingual knowledge base extracted from Wikipedia. Semant. Web **6**(2), 167–195 (2015). https://doi.org/10.3233/SW-140134
67. Li, J., Sun, Y., Johnson, R.J., Sciaky, D., Wei, C., Leaman, R., Davis, A.P., Mattingly, C.J., Wiegers, T.C., Lu, Z.: Biocreative V CDR task corpus: a resource for chemical disease relation extraction. Database J. Biol. Databases Curation **2016**, (2016). https://doi.org/10.1093/database/baw068
68. Liakata, M., Saha, S., Dobnik, S., Batchelor, C.R., Rebholz-Schuhmann, D.: Automatic recognition of conceptualization zones in scientific articles and two life science applications. Bioinformatics **28**(7), 991–1000 (2012). https://doi.org/10.1093/bioinformatics/bts071
69. Liakata, M., Teufel, S., Siddharthan, A., Batchelor, C.R.: Corpora for the conceptualisation and zoning of scientific papers. In: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D. (eds.) Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 2010, Valletta, Malta. European Language Resources Association (2010). http://www.lrec-conf.org/proceedings/lrec2010/summaries/644.html
70. Lo, K., Wang, L.L., Neumann, M., Kinney, R., Weld, D.S.: S2ORC: the semantic scholar open research corpus. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, 2020, pp. 4969–4983. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.447
71. Luan, Y., He, L., Ostendorf, M., Hajishirzi, H.: Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In: Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (eds.) Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018, pp. 3219–3232. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/d18-1360
72. Lubani, M., Noah, S.A.M., Mahmud, R.: Ontology population: approaches and design aspects. J. Inf. Sci. (2019). https://doi.org/10.1177/0165551518801819
73. Manghi, P., Bardi, A., Atzori, C., Baglioni, M., Manola, N., Schirrwagen, J., Principe, P.: The OpenAIRE research graph data model. Zenodo (2019). https://doi.org/10.5281/zenodo.2643199
74. Mesbah, S., Fragkeskos, K., Lofi, C., Bozzon, A., Houben, G.: Semantic annotation of data processing pipelines in scientific publications. In: Blomqvist, E., Maynard, D., Gangemi, A., Hoekstra, R., Hitzler, P., Hartig, O. (eds.) The Semantic Web—14th International Conference, ESWC 2017, Portorož, Slovenia, 2017, Proceedings, Part I, Lecture Notes in Computer Science, vol. 10249, pp. 321–336 (2017). https://doi.org/10.1007/978-3-319-58068-5_20
75. Nasar, Z., Jaffry, S.W., Malik, M.K.: Information extraction from scientific articles: a survey. Scientometrics **117**(3), 1931–1990 (2018). https://doi.org/10.1007/s11192-018-2921-5
76. Nguyen, V.B., Svátek, V., Rabby, G., Corcho, Ó.: Ontologies supporting research-related information foraging using knowledge graphs: literature survey and holistic model mapping. In: Keet, C.M., Dumontier, M. (eds.) Knowledge Engineering and Knowledge Management—22nd International Conference, EKAW 2020, Bolzano, Italy, 2020, Proceedings, Lecture Notes in Computer Science, vol. 12387, pp. 88–103. Springer (2020). https://doi.org/10.1007/978-3-030-61244-3_6
77. Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E.: A review of relational machine learning for knowledge graphs. Proc. IEEE **104**(1), 11–33 (2016). https://doi.org/10.1109/JPROC.2015.2483592
78. Oelen, A., Jaradeh, M.Y., Stocker, M., Auer, S.: Generate FAIR literature surveys with scholarly knowledge graphs. In: Huang, R., Wu, D., Marchionini, G., He, D., Cunningham, S.J., Hansen, P. (eds.) JCDL '20: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, Virtual Event, China, 2020, pp. 97–106. ACM (2020). https://doi.org/10.1145/3383583.3398520
79. Okoli, C.: A guide to conducting a standalone systematic literature review. Commun. Assoc. Inf. Syst. **37**, 43 (2015)
80. Papers with code. https://paperswithcode.com/. Accessed 04 Oct 2021
81. Park, S., Caragea, C.: Scientific keyphrase identification and classification by pre-trained language models intermediate task transfer learning. In: Scott, D., Bel, N., Zong, C. (eds.) Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), 2020, pp. 5409–5419. International Committee on Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.coling-main.472
82. Peng, Y., Yan, S., Lu, Z.: Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets. In: Demner-Fushman, D., Cohen, K.B., Ananiadou, S., Tsujii, J. (eds.) Proceedings of the 18th BioNLP Workshop and Shared Task, BioNLP@ACL 2019, Florence, Italy, 2019, pp. 58–65. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/w19-5006
83. Peroni, S., Shotton, D.M.: Fabio and cito: ontologies for describing bibliographic resources and citations. J. Web Semant. **17**, 33–43 (2012). https://doi.org/10.1016/j.websem.2012.08.001
84. Pertsas, V., Constantopoulos, P.: Scholarly ontology: modelling scholarly practices. Int. J. Digit. Libr. **18**(3), 173–190 (2017). https://doi.org/10.1007/s00799-016-0169-3
85. Petasis, G., Karkaletsis, V., Paliouras, G., Krithara, A., Zavitsanos, E.: Ontology population and enrichment: state of the art. In: Paliouras, G., Spyropoulos, C.D., Tsatsaronis, G. (eds.) Knowledge-Driven Multimedia Information Extraction and Ontology Evolution—Bridging the Semantic Gap, Lecture Notes in Computer Science, vol. 6050, pp. 134–166. Springer (2011). https://doi.org/10.1007/978-3-642-20795-2_6
86. Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché-Buc, F., Fox, E.B., Larochelle, H.: Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). CoRR **abs/2003.12206** (2020). arXiv:2003.12206
87. Pipino, L.L., Lee, Y.W., Wang, R.Y.: Data quality assessment. Commun. ACM **45**(4), 211–218 (2002). https://doi.org/10.1145/505248.506010
88. Pujara, J., Singh, S.: Mining knowledge graphs from text. In: Chang, Y., Zhai, C., Liu, Y., Maarek, Y. (eds.) Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, 2018, pp. 789–790. ACM (2018). https://doi.org/10.1145/3159652.3162011
89. Qasemi Zadeh, B., Handschuh, B.S.: The ACL RD-TEC: a dataset for benchmarking terminology extraction and classification in computational linguistics. In: Proceedings of the 4th International Workshop on Computational Terminology (Computerm), pp. 52–63. Association for Computational Linguistics and Dublin City University, Dublin, Ireland (2014). 10.3115/v1/W14-4807. https://www.aclweb.org/anthology/W14-4807
90. Qasemi Zadeh, B., Schumann, A.: The ACL RD-TEC 2.0: a language resource for evaluating term extraction and entity recognition methods. In: Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of

the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, 2016. European Language Resources Association (ELRA) (2016). http://www.lrec-conf.org/proceedings/lrec2016/summaries/681.html
91. Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P.: Squad: 100, 000+ questions for machine comprehension of text. In: Su, J., Carreras, X., Duh, K. (eds.) Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, 2016, pp. 2383–2392. The Association for Computational Linguistics (2016). https://doi.org/10.18653/v1/d16-1264
92. Richardson, S., Wilson, M., Nishikawa, J., Hayward, R.: The well-built clinical question: a key to evidence-based decisions. ACP J. Club **123**(3), A12–13 (1995)
93. Ruiz-Iniesta, A., Corcho, Ó.: A review of ontologies for describing scholarly and scientific documents. In: Castro, A.G., Lange, C., Lord, P.W., Stevens, R. (eds.) Proceedings of the 4th Workshop on Semantic Publishing Co-located with the 11th Extended Semantic Web Conference (ESWC 2014), Anissaras, Greece, 2014, CEUR Workshop Proceedings, vol. 1155. CEUR-WS.org (2014). http://ceur-ws.org/Vol-1155/paper-07.pdf
94. Safder, I., Hassan, S., Visvizi, A., Noraset, T., Nawaz, R., Tuarob, S.: Deep learning-based extraction of algorithmic metadata in full-text scholarly documents. Inf. Process. Manag. **57**(6), 102269 (2020). https://doi.org/10.1016/j.ipm.2020.102269
95. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Birukou, A., Osborne, F., Motta, E.: The computer science ontology: a comprehensive automatically-generated taxonomy of research areas. Data Intell. **2**(3), 379–416 (2020). https://doi.org/10.1162/dint_a_00055
96. Say, A., Fathalla, S., Vahdati, S., Lehmann, J., Auer, S.: Semantic representation of physics research data. In: Aveiro, D., Dietz, J.L.G., Filipe, J. (eds.) Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2020, vol. 2: KEOD, Budapest, Hungary, 2020, pp. 64–75. SCITEPRESS (2020). https://doi.org/10.5220/0010111000640075
97. Singh, M., Barua, B., Palod, P., Garg, M., Satapathy, S., Bushi, S., Ayush, K., Rohith, K.S., Gamidi, T., Goyal, P., Mukherjee, A.: OCR++: a robust framework for information extraction from scholarly articles. In: Calzolari, N., Matsumoto, Y., Prasad, R. (eds.) COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 2016, Osaka, Japan, pp. 3390–3400. ACL (2016). https://www.aclweb.org/anthology/C16-1320/
98. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S., Consortium, T.O.: The obo foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotechnol. **25**(11), 1251–1255 (2007). https://doi.org/10.1038/nbt1346
99. Soldatova, L.N., King, R.D.: An ontology of scientific experiments. J. R. Soc. Interface **3**(11), 795–803 (2006). https://doi.org/10.1098/rsif.2006.0134
100. Stead, C., Smith, S., Busch, P.A., Vatanasakdakul, S.: Emerald 110k: a multidisciplinary dataset for abstract sentence classification. In: Mistica, M., Piccardi, M., MacKinlay, A. (eds.) Proceedings of the The 17th Annual Workshop of the Australasian Language Technology Association, ALTA 2019, Sydney, Australia, 2019, pp. 120–125. Australasian Language Technology Association (2019). https://aclweb.org/anthology/papers/U/U19/U19-1016/
101. Stocker, M., Prinz, M., Rostami, F., Kempf, T.: Towards research infrastructures that curate scientific information: a use case in life sciences. In: Auer, S., Vidal, M. (eds.) Data Integration in the Life Sciences—13th International Conference, DILS 2018, Hannover, Germany, 2018, Proceedings, Lecture Notes in Computer Science, vol. 11371, pp. 61–74. Springer (2018). https://doi.org/10.1007/978-3-030-06016-9_6
102. Suchanek, F.M., Gross-Amblard, D., Abiteboul, S.: Watermarking for ontologies. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N.F., Blomqvist, E. (eds.) The Semantic Web—ISWC 2011—10th International Semantic Web Conference, Bonn, Germany, 2011, Proceedings, Part I, Lecture Notes in Computer Science, vol. 7031, pp. 697–713. Springer (2011). https://doi.org/10.1007/978-3-642-25073-6_44
103. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: Williamson, C.L., Zurko, M.E., Patel-Schneider, P.F., Shenoy, P.J. (eds.) Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, 2007, pp. 697–706. ACM (2007). https://doi.org/10.1145/1242572.1242667
104. Talburt, J.R.: 2—principles of information quality. In: Talburt, J.R. (ed.) Entity Resolution and Information Quality, pp. 39–62. Morgan Kaufmann, Boston (2011). https://doi.org/10.1016/B978-0-12-381972-7.00002-6. http://www.sciencedirect.com/science/article/pii/B9780123819727000026
105. Teufel, S., Siddharthan, A., Batchelor, C.R.: Towards domain-independent argumentative zoning: Evidence from chemistry and computational linguistics. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, Singapore, A Meeting of SIGDAT, a Special Interest Group of the ACL, pp. 1493–1502. ACL (2009). https://www.aclweb.org/anthology/D09-1155/
106. Vahdati, S., Fathalla, S., Auer, S., Lange, C., Vidal, M.: Semantic representation of scientific publications. In: Doucet, A., Isaac, A., Golub, K., Aalberg, T., Jatowt, A. (eds.) Digital Libraries for Open Knowledge—23rd International Conference on Theory and Practice of Digital Libraries, TPDL 2019, Oslo, Norway, 2019, Proceedings, Lecture Notes in Computer Science, vol. 11799, pp. 375–379. Springer (2019). https://doi.org/10.1007/978-3-030-30760-8_37
107. Vandenbussche, P., Atemezing, G., Poveda-Villalón, M., Vatant, B.: Linked open vocabularies (LOV): a gateway to reusable semantic vocabularies on the web. Semant. Web **8**(3), 437–452 (2017). https://doi.org/10.3233/SW-160213
108. Vrandecic, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Commun. ACM **57**(10), 78–85 (2014). https://doi.org/10.1145/2629489
109. Waard, A., Tel, G.: The ABCDE format enabling semantic conference proceedings. In: Völkel, M., Schaffert, S. (eds.) SemWiki2006, First Workshop on Semantic Wikis—From Wiki to Semantics, Proceedings, Co-located with the ESWC2006, Budva, Montenegro, 2006, CEUR Workshop Proceedings, vol. 206. CEUR-WS.org (2006). http://ceur-ws.org/Vol-206/paper8.pdf
110. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. **12**(4), 5–33 (1996)
111. Weikum, G., Dong, L., Razniewski, S., Suchanek, F.M.: Machine knowledge: creation and curation of comprehensive knowledge bases. CoRR **abs/2009.11564** (2020). arXiv:2009.11564
112. Xiong, C., Power, R., Callan, J.: Explicit semantic ranking for academic search via knowledge graph embedding. In: Barrett, R., Cummings, R., Agichtein, E., Gabrilovich, E. (eds.) Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, 2017, pp. 1271–1279. ACM (2017). https://doi.org/10.1145/3038912.3052558
113. Yaman, B., Pasin, M., Freudenberg, M.: Interlinking scigraph and dbpedia datasets using link discovery and named entity recognition techniques. In: Eskevich, M., de Melo, G., Fäth, C., McCrae, J.P., Buitelaar, P., Chiarcos, C., Klimek, B., Dojchinovski, M.

(eds.) 2nd Conference on Language, Data and Knowledge, LDK 2019, Leipzig, Germany, *OASICS*, vol. 70, pp. 15:1–15:8. Schloss Dagstuhl–Leibniz–Zentrum für Informatik (2019). https://doi.org/10.4230/OASIcs.LDK.2019.15

114. Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S.: Quality assessment for linked data: a survey. Semant. Web **7**(1), 63–93 (2016). https://doi.org/10.3233/SW-150175
115. Zhang, Y., Wang, M., Saberi, M., Chang, E.: From big scholarly data to solution-oriented knowledge repository. Front. Big Data **2**, 38 (2019). https://doi.org/10.3389/fdata.2019.00038