Improving On-line Scientific Resource Profiling by Exploiting Resource Citation Information in the Literature

Anqing Zheng [a,1], He Zhao [a,1], Zhunchen Luo [b,1], Chong Feng [a,*], Xiaopeng Liu [b], Yuming Ye [b]

[a] *School of Computer Science, Beijing Institute of Technology, Beijing, China*
[b] *Information Research Center of Military Science, PLA Academy of Military Science, Beijing, China*

ARTICLE INFO

ABSTRACT

We study the task of *on-line scientific resource profiling*, which aims at better understanding and summarizing on-line scientific resources to promote resource search and recommendation systems. To this end we propose to exploit the resource citation information in scientific literature by extracting the fine-grained relations between the cited on-line resources and other resource-related scientific terms. In this paper we create a dataset (SciResTR) and develop a framework (SciResTR-IE) which jointly extracts all the related scientific terms and the *resource-term* relations. Extensive experiments demonstrate that our framework outperforms other baselines significantly, by around 5% in scientific information extraction tasks absolutely. We further show that our proposed system can automatically construct several on-line-resource-centered networks from a large corpus of scientific articles, which is a first step towards utilizing resource citation information in the literature to improve on-line scientific resource profiling.

1. Introduction

On-line scientific resources are useful resources, such as *softwares*, *services* or *datasets*, which are available on the Internet and can greatly help researchers on their scientific studies. Recently, some search engines specific for on-line scientific resources, such as the Google Dataset Search,[2] have been developed. However, most resources are roughly indexed by their names and it is still difficult to get comprehensive understandings of more scientific resources. Therefore, more effective approaches to better depicting and summarizing on-line scientific resources are in great demand. As the number of scientific publications increases rapidly, more and more on-line resources are cited in papers. The context of these resource citations directly reflects how the authors conduct their research with the resources, as the examples shown in Fig. 1(a). It makes that the scientific literature becomes a high-quality corpus to study the on-line scientific resource profiling.

In this paper, we are the first to exploit resource citation information in scientific literature to study the task of on-line scientific resource profiling, formulated as extracting the fine-grained relations between the cited on-line resources and other resource-related scientific terms in the sentences that cite the resource. It should be noted that this task is similar to an extraction task, but the difference between them is that the extraction task is a necessary step to implement the task of resource profiling and our ultimate
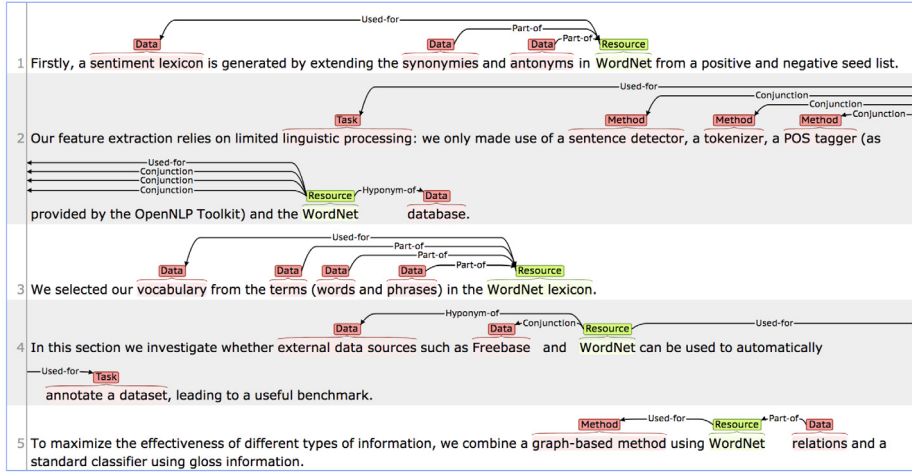
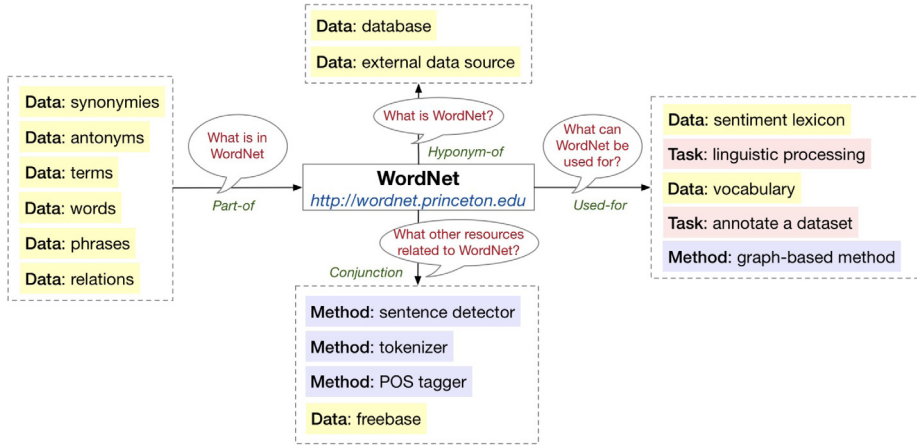(a) Annotation examples for the resource "WordNet".

(b) Resource profiling example for "WordNet".

Fig. 1. An illustrative example of resource profiling for a scientific resource "WordNet". In Fig. 1(a), there are five resource citation sentences from different scientific papers. The resource mentions highlighted in green are the names or descriptions of the resources, which are definitely given. The related scientific terms highlighted in red, e.g. *Sentiment Lexicon (Data)*, and the *resource-term* relations, e.g. *Used-for*, are to be extracted. The extracted knowledge and the resource profiling structure for "WordNet" are shown in Fig. 1(b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

goal is to build a structured description for each on-line resource item to help researchers better understand and summarize on-line scientific resources. Since the unstructured resource citation sentences cannot be used directly for the purpose, we propose to extract more fine-grained information, that is, to discover relationships between the resource and other scientific terms in the sentence where the resource is cited. An illustrative example for a scientific resource "WordNet" is shown in Fig. 1: given the resource citation sentences and the resource citation mentions shown in Fig. 1(a), an automatic framework is supposed to extract other scientific terms in the sentence and the *resource-term* relations between them, where the scientific term is a word or expression indicating a particular scientific entity, and the *resource-term* relation indicates how the resource interacts with a scientific term. Note that, if a term is not related to the target resource, it will not be extracted. Then by categorizing the scientific terms and relations, a profiling structure of "WordNet" can be built for the target resource as is shown in Fig. 1(b). With the help of the resource profiling, it will be easy to answer some questions like *"What is WordNet?"*, *"What is included in WordNet?"*, *"What can WordNet be used for?"*, or *"What other resources are related to WordNet?"*. Therefore, such a structured description for resource profiling is potentially useful to both academia and industry.

Although mostly previous work concentrates on information extraction from scientific literature (Augenstein, Das, Riedel, Vikraman, & McCallum, 2017; Gábor et al., 2018; Luan, Ostendorf, & Hajishirzi, 2017; Rei, 2017), the specific studies for the meta-data of on-line resources are still extremely rare. Zhao, Luo, Feng and Ye (2019) and Zhao, Luo, Feng, Zheng and Liu (2019) introduced the task of modeling roles and functions for on-line resource citations, while they only conducted the sentence-level

classification for citation contexts. More fine-grained information, for instance, other relevant scientific terms which interacts with the resource citation and the *resource-term* relations between them were not taken into full consideration. Hence, for the problem addressed in this paper, we develop a framework to extract both the correlative scientific terms and the *resource-term* relations when given the sentence of resource citation and the mention of target on-line resource. Concretely, we treat both the term extraction and the relation extraction tasks as multiclass classification problems with shared term representations and learn contextualized embeddings based on SciBERT (Beltagy, Lo, & Cohan, 2019), a pre-trained language model optimized for scientific text.

In summary, we make the following contributions. To the best of our knowledge, our work is the first attempt towards the on-line scientific resource profiling by using the resource citation information in scientific text. Based on previous taxonomies, we propose a new annotation scheme and develop a dataset (SciResTR) for fine-grained resource citation analysis. Taking advantage of the SciBERT encoder, we develop a joint framework (SciResTR-IE) for the term extraction and the relation extraction tasks, which outperforms other baselines significantly. Above all, we propose an automated system to construct several resource-centered networks from any given scientific articles, which is potentially useful to both academia and industry by analyzing such a structured description for resource profiling. The annotated dataset and our code will be released to help future research.

2. Related work

The Google Dataset Search (Hussain, 2019), announced in September 2018, is a search engine specific to find research data published on the Internet. While developing the search system, some efforts on the profiling for on-line dataset resources are necessary (Canino, 2019). More recently, Vrandecic (2019) used Wikidata to provide metadata for datasets and aimed at increasing findability and accessibility of certain datasets. However, the explorations are still restricted on the on-line datasets. More kinds of on-line resources, such as tasks, methods and services, have not received sufficient attention.

There has been growing interest in research on data mining, text analysis and information extraction for scientific literature. Past studies have explored the analysis of citations (Abu-Jbara & Radev, 2011; Bu, Lu, Wu, Chen, & Huang, 2021; Gábor, Zargayouna, Buscaldi, Tellier, & Charnois, 2016; Qazvinian & Radev, 2008), for example the citation function (Athar & Teufel, 2012; Athar & Teufel, 2012; Jurgens, Kumar, Hoover, McFarland, & Jurafsky, 2018; Teufel, Siddharthan, & Tidhar, 2006) which is to analyze the sentiment and reason for citing a paper; the citation recommendation (He, Kifer, Pei, Mitra, & Giles, 2011; He, Pei, Kifer, Mitra, & Giles, 2010; Huang et al., 2012; Huang, Wu, Liang, Mitra, & Giles, 2015; Tang & Zhang, 2009) which is to predict the papers that need to be cited when given the context; and the citation network (Do, Chandrasekaran, Cho, & Kan, 2013; Eto, 2019; Jaidka et al., 2014; Prabhakaran, Hamilton, McFarland, & Jurafsky, 2016; Sim, Smith, & Smith, 2012; Vogel & Jurafsky, 2012) which is to find key authors in a field or observe the over-time trends. Besides (Ji, Tao, Fei, & Ren, 2020; Oral, Emekligil, Arslan, & Eryigit, 2020; Safder et al., 2020) analyzed the complex relation extraction problem and retrieval of massive textual information on text-intensive documents. Though there is no doubt that scientific papers can be a valuable corpus for on-line resource analyzing, little prior work studied the on-line resources citations. Some previous work undertook limited attempts to the on-line resource detection (de la Calle, García-Remesal, Chiesa, de la Iglesia, & Maojo, 2009; Duck, Nenadic, Brass, Robertson, & Stevens, 2013, 2014; Duck et al., 2016; Yamamoto & Takagi, 2007) and resource citation context analysis (Zhao, Luo, Feng, Ye, 2019; Zhao, Luo, Feng, Zheng et al., 2019). Zhao, Luo, Feng, Zheng et al. (2019) analyzed the role and function of on-line resource citations by sentence-level classification. However, the phrase-level identification within a scientific sentence is not well considered, such as fine-grained entities and relationships.

Nowadays, increasing work has developed methods for meta-information extraction, for example, extracting scientific entities and relations (Gábor et al., 2016; Gupta & Manning, 2011; Tsai, Kundu, & Roth, 2013). SemEval 2017 (Augenstein et al., 2017) and 2018 (Gábor et al., 2018) formalized two new tasks of scientific information extraction and provided datasets for relevant research. Luan, He, Ostendorf, and Hajishirzi (2018) extended the previous SemEval tasks and proposed a new problem of identifying scientific entities, their relations and coreference clusters. Then Hou, Jochim, Gleize, Bonin, and Ganguly (2019) introduced another practical work of automatically extracting tasks, datasets, evaluation metrics and scores from papers in order to construct the scientific leaderboards. And many neural-based methods have shown superior performance for these Science-IE tasks (Ammar et al., 2018; Ammar, Peters, Bhagavatula, & Power, 2017; Luan et al., 2017; Rei, 2017). Luan et al.'s DyGIE (Luan et al., 2019) uses dynamically constructed span graphs to capture the interaction of spans for entity and relation extraction. Besides Wadden et al.'s DyGIE++ (Wadden, Wennberg, Luan, & Hajishirzi, 2019) has replaced BiLSTM encoder with BERT (Devlin, Chang, Lee, & Toutanova, 2019) or SciBERT, and achieves state-of-the-art results on three Science-IE tasks.

Different from previous work, we are committed to profile another important meta-data, the on-line resource citations in the scientific literature. We are the first to conduct a more fine-grained study by extracting resource mentions, scientific terms and their interactive relations to get more detailed knowledge about the on-line scientific resources. In addition, we also automatically organize the extracted information from a large collection of scientific on-line resource citation sentences into several resource-centered networks, towards the scientific on-line resource profiling.

3. Research objectives and dataset

3.1. Research objectives

As described in the Introduction section, it is difficult to get comprehensive understandings of more scientific resource with what existed before. We find that there are a huge number of on-line resources cited in the literature, which imply rich information about

Case 1: the in-line resources in body texts
To extract the three components of the coseismic offsets, we process the data with GIPSY-OASIS software (***https://gipsy-oasis.jpl.nasa.gov/***) and Jet Propulsion Laboratory (JPL) flinnR orbit and clock products.

Case 2: the out-line resources in footnotes
Using Bikel's randomized parsing evaluation comparator[5], we find that both reranking models outperform the baseline generative model to statistical significance for recall and precision.

[5]***http://www.cis.upenn.edu/~dbikel/software.html***

Fig. 2. Examples of the in-line resources in body texts and the out-line resources in footnotes. Specially, the cited on-line resources are highlighted in red and the corresponding hyperlinks are underlined. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the scientific research activities, but have barely been considered. To better understand and summarize on-line scientific resources, we study the task of on-line scientific resource profiling, formulated as extracting the fine-grained relations between the cited on-line resources and other resource-related scientific terms in the sentences that cite the resource.

The closest work to ours is probably SciIE (Luan et al., 2018) and SciResCLF (Zhao, Luo, Feng, Zheng et al., 2019). SciIE studies the problem of identifying scientific entities, their relations and coreference clusters in scientific articles. Compared to SciIE, our task aims to extract the fine-grained relations between the cited on-line resources and scientific terms in the sentences, instead of extracting the relationship between any two probably scientific entities in the sentence. In other words, our obtained is knowledge that resembles parts of a knowledge graph which is relevant to on-line scientific resource. Specially, our task can improve on-line resource profiling and could be utilized in different ways. In addition, SciResCLF is the first to model roles and functions for on-line resource citations, while they only conducted the sentence-level classification in scientific full text. We extend their work to analyze the more fine-grained information by extracting both the correlative scientific terms and the resource-term relations when given the sentence of resource citation and the mention of target on-line

As shown in Fig. 1, our work can build such a structured description for on-line resource profiling, which could be utilized in different ways. You can use it to (1) learn about the interrelations between the resources and the scientific terms more easily and intuitively; (2) construct indexes or knowledge graphs for on-line scientific resources; (3) develop scientific resource search engine and recommendation systems and so on. To use the example of WordNet, we suspect that most people refer to one of the two core literature references (the Miller et al. paper Miller, 1995 and Fellbaum et al. book Fellbaum & Vossen, 2012) rather than the URL. So why we focus on improving on-line scientific resource profiling? First, on-line scientific resources are useful resources, such as software, tools, services and datasets, which are available on the Internet and can greatly help researchers on their scientific studies. We can directly use the URLs to access on-line scientific resources. Second, the less frequently referenced on-line resources benefit more from our approach than the ones, which everybody knows about already and which are more likely to be obtained elsewhere. But limited the size of our dateset and our proposed framework, the extracted results in the top 200 resources as the dataset are nearly 30%–40% errors in the inference phase. We can infer that resources that are rarely used will perform worse than the average result, and it is not consistent with our original intention. Specially, the total effectiveness of our proposed framework still has much room for improvement.

3.2. Dataset

To the best of our knowledge, there is no ready-to-use dataset for our task. So we construct a dataset (called SciResTR) which includes annotations for on-line **res**ource mentions, **sci**entific **t**erms and the *resource-term* **r**elations between them.

3.2.1. Data collection
Our dataset is collected from the ACL Anthology Reference Corpus[3] (ARC), of which the papers are in the domain of Natural Language Processing (NLP) and in the PDF format. As shown in Fig. 2, we divide the resources into two types by their hyperlink locations: the in-line resources in body texts and the out-line resources in footnotes. Above all, we parse the PDFs using pdf2xml[4] to get the sentences containing the citations for both two types of resources and find the corresponding hyperlinks by recognizing the superscripts or using regular expressions in the output XML files. And the script for our parser is already accessible online at https://github.com/aqzheng/extract-link-from-pdf-using-pdf2xml-in-Windows. Then an extracted data sample is composed of a hyperlink and a sentence having at least one on-line resource involved. Thus the same resource can be uniquely identified with its on-line hyperlink. We evaluated our parser on a set of 100 papers randomly sampled from the ARC, with various venues (e.g. *ACL, EMNLP, NAACL-HLT, etc*). 179 of the total 188 on-line resources citation sentences were correctly extracted and the accuracy achieved 95.21%. We collected all the extracted data samples from the ARC. But for annotating, we selected the top 200 most frequent resources by their hyperlinks and random sampled 2000 sentences in the subset.

[3] http://acl-arc.comp.nus.edu.sg/.
[4] https://github.com/kermitt2/pdf2xml.

Table 1
Statistics of term categories and relation categories in the 2000 sentences of the SciResTR dataset.

Term category	Count	Relation category	Count
Task	626	Used-for	1679
Method	935	Feature-of	130
Data	1244	Hyponym-of	492
Service	163	Part-of	382
Metric	50	Compare	25
Generic Term	135	Conjunction	545

3.2.2. Annotation scheme

In the scope of a target sentence where an on-line resource is cited, we first ask the annotators to identify the head position of the on-line resource, whose tail position can be found based on the location of the hyperlinks (e.g. *software or comparator* in Fig. 2). For the given citation sentence and the resource mention, another two types of information are annotated: (1) all the extracted scientific terms which are related to the resource; (2) the relations between the resource and the scientific terms. We extend the previous scheme addressed by Luan et al. (2018) and defined 6 scientific term categories, including *Task*, *Method*, *Material*, *Service*, *Metric* and *Generic Term*. For the *resource-term* relations, we defined 4 asymmetric relation categories (*Used-for*, *Hyponym-of*, *Part-of*, *Feature-of*), together with two symmetric relation categories (*Conjunction*, *Compare*). The categories along with their counts are shown in Table 1. Annotations were performed by a group of 3 PhD students and 3 Master students, who major in NLP. We measured the agreement between annotators and observed that Fleiss Kappa is 0.67 for the scientific terms and 0.85 for the relations. Most disagreements arose from the bounds of the scientific terms and then the annotators reached consensuses after discussions.

3.2.3. Term categories

- **Task**: Applications, problem to solve, systems to construct.
  For instance: Language modeling, relation classification, transductive inference, tree parsing, etc.
- **Method**: Methods, models, toolkits, softwares, libraries, frameworks, systems to use, or components of a system.
  For instance: language model, POS tagger, Apache OpenNLP tools, TreeTagger, CRF+ +, etc.
- **Data**: Data, datasets, corpus, knowledge base.
  For instance: WordNet, Wikipedia, Arabic Wikipedia articles, lexical semantic resource, list of phrases and tokens, part of speech tags, etc.
- **Service**: Website, on-line service, platform.
  For instance: Google Translate, Twitter's Streaming API Service, information retrieval platform, etc.
- **Metric**: Metrics, measures, or terms that can express quality of a method or system.
  For instance: F1, BLEU, recall, precision, ROC curve, mean-squared error, robustness, time complexity, etc.
- **Generic Term**: General terms or pronouns that may refer to a term but do not belong to the above categories, often used as feature expressions.
  For instance: social sciences, humanities, languages, phrase-based, open-source, multi-task, etc.

3.2.4. Relation categories

It is worth notice that B always points to A for asymmetric relations.

- **Used-for**: B is used for A, B models A, A is trained on B, B exploits A, A is based on B.
  For instance: Please Data was collected from Twitter's streaming API service.
- **Feature-of**: B belongs to A, B is a feature of A, B is under A domain.
  For instance: We used the skip-gram model of word2vec with a window size of 5 and negative sampling to train the SKWIKI model.
- **Hyponym-of**: B is a hyponym of A, B is a type of A.
  For instance: The setup allows us to efficiently process large amounts of data points using a Map-Reduce framework via Hadoop.
- **Part-of**: B is a part of A.
  For instance: We adopt Chinese Wikipedia Dump to train our models as well as the original CBOW and SkipGram, implemented in the Word2Vec tool for comparison.
- **Compare**: Opposite of conjunction, compare two models/methods, or listing two opposing entities.
  For instance: We also ran another implementation of LDA, which was 30 times slower than Mallet.
- **Conjunction**: Function as similar role or use/incorporate with.
  For instance: Existing online translation systems such as Google Translate and Bing Translator are thus a great service.

Specially, *Compare* and *Conjunction* are two symmetric relations, which use blue to denote entity. Note that *Feature-of* and *Part-of* are different. B in *Feature-of* is often a component or property of A, and A must realize its function with B. But in *Part-of*, A and B are congeneric entities and have similar functions.
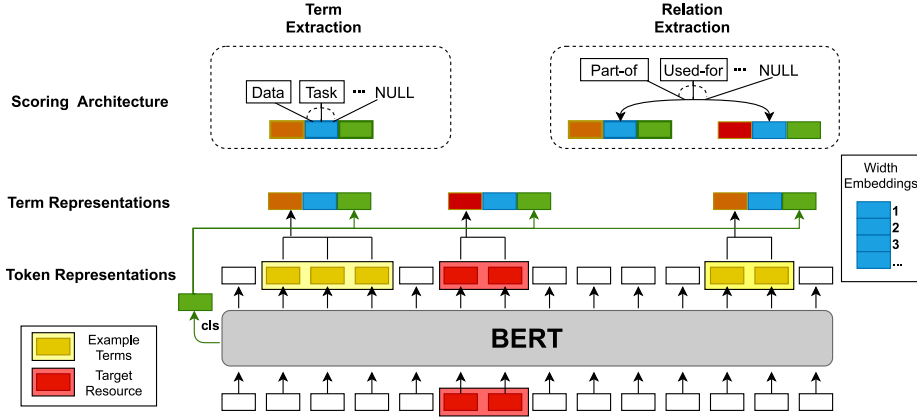
Fig. 3. The overall architecture of the SciResTR-IE framework.

4. SciResTR-IE framework

4.1. Problem definition

Given a resource citation sentence $s = w_1, w_2, ..., w_n$ and a target resource mention $resource = (r_h, r_t)$, where r_h and r_t are the resource indices in the sentence, our goal is to extract a set of resource-related scientific terms $T = \{(i, j, l_t)|1 \leq i \leq j \leq n; j-i+1 \leq L\}$, where i and j are the word indices in the sentence, l_t represents the label of the term, and L is the maximum length of all the scientific terms. Besides, we also predict the *resource-term* relations $R = \{(resource, term, l_r) \cup (term, resource, l_r)|term \in T\}$, where l_r represents the label of the *resource-term* relation. Note that relations are asymmetric in general, we need to classify both $(resource, term)$ and $(term, resource)$.

4.2. Term and relation extraction model

We propose to exploit the resource citation information in scientific literature by automatically extracting the fine-grained relations between the cited on-line resources and other scientific terms. Different from a standard tagging system, our framework (called SciResTR-IE) enumerates all possible terms during decoding and can effectively detect overlapped terms. It avoids cascading errors between tasks by jointly extracting all the related scientific terms and the *resource-term* relations when given a resource citation sentence and the resource citation mention. As shown in Fig. 3, SciResTR-IE framework consists of three critical components, i.e., token representations part, term representations part and scoring architecture part. In the following parts, we will give the detailed introduction of these three components in our framework.

4.2.1. Token representations

Based on previous work (Cohan, Ammar, van Zuylen, & Cady, 2019; Jiang, D'Souza, Auer, & Downie, 2020), our model takes advantage of the pre-trained SciBERT model to learn contextualized representations. Taking a tokenized sentence as input, we obtain a sequence of n byte-pair encoded (BPE) tokens (Sennrich, Haddow, & Birch, 2016). Byte-pair encoding represents infrequent words (such as toolkit) by common subwords (tool and kit) and is utilized in SciBERT to limit the vocabulary size and to map out-of-vocabulary word. Finally, we use the pre-trained SciBERT as a hidden encoder to generate token representations for BPE tokens.

4.2.2. Term representations

We only consider scientific terms with up to L words when we enumerate terms. For each possible related term, we first utilize the *boundary representation* and the *merged representation* to capture the term-level information. The boundary representation is directly obtained by the outputs of the SciBERT final hidden representation layer corresponding to the boundary words, which are respectively the start and end words in the term. The merged representation, which reflects all the information of the tokens included in a term, is also important to the prediction. Following previous work, we use an attention mechanism based on a feed forward neural network to get the merged representation $\hat{x}_{i,j}$. Finally, we concatenate the above term-level embeddings to get the representation of term(i, j) as follows:

$$t(i, j) = [x_i^*, x_j^*, \hat{x}_{i,j}, size(\lambda), \mathbf{c}] \tag{1}$$

where x_i^* and x_j^* are the boundary representations, and $size(\lambda)$ is a width embedding, which encodes the size of term (i, j). The width embedding allows the model to incorporate a prior over the term width (note that scientific terms which are too long are unlikely to represent entities). Finally, we add the hidden output $x_{[CLS]}^*$ as the classifier representation $\mathbf{c}$, which represents the context of the overall sentence. Especially, the representation of the target resource is represented as $t(r_s, r_e)$.
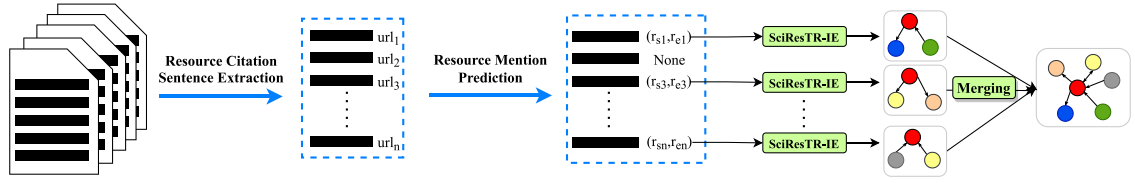
Fig. 4. Resource-centered network construction process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2.3. Scoring architecture

We use feed forward neural networks (FFNNs) over shared term representations $t(i, j)$ to calculate a set of term and pairwise term scores.

Term Extraction In this part, we predict the label for each term. Given these term representations, we calculate a vector of label scores, including the NULL label. For each $term(i, j)$, we get the distribution over the term categories by feeding the term label scores to a softmax function as follows:

$$score^{TE}_{(i,j)} = \text{FFNN}_{\text{TE}}(t(i, j))$$
$$y^{TE}_{(i,j)} = softmax(score^{TE}_{(i,j)}) \tag{2}$$

where $score^{TE}_{(i,j)}$ measures the probability that a $term(i, j)$ has a label e, and scores involving the NULL label are set to 0. Only the terms whose label is not predicted to NULL are selected for relation extraction.

Relation Extraction In this part, we predict the relation label between the target resource and each possible related term. Due to the asymmetric relation types, we put the target resource mention in the logical subject position and logical object position of the relationship in turn. For each possible related term, we calculate two ordered pairwise embeddings:

$$r(i, j)_s = [t(i, j), t(r_s, r_e), t(i, j) \circ t(r_s, r_e)],$$
$$r(i, j)_o = [t(r_s, r_e), t(i, j), t(r_s, r_e) \circ t(i, j)]. \tag{3}$$

We use the ordered pairwise embedding to calculate a vector of relation labels. For each ordered pairwise embedding, the softmax function is applied to generate a distribution over the relation categories using its relation label scores:

$$score^{RE}_{(i,j,x)} = \text{FFNN}_{\text{RE}}(r(i, j)_x)$$
$$y^{RE}_{(i,j,x)} = softmax(score^{RE}_{(i,j,x)}) \tag{4}$$

where $x \in \{s, o\}$ and $score^{RE}_{(i,j,x)}$ measures how likely the target resource and each possible related term are associated in a relation r.

4.2.4. Training objective

The goal of training is to optimize all the parameters so as to minimize the cross entropy loss function as much as possible. We use $\hat{y}^{TE}_{(i,j)}$ to represent the correct term label for $term(i, j)$ and $\hat{y}^{RE}_{(i,j,x)}$ to represent the correct relation label between $term(i, j)$ and the target resource. Then the final multi-task learning loss function is

$$loss = - \alpha \sum_{i,j} \hat{y}^{TE}_{(i,j)} \log(y^{TE}_{(i,j)})$$
$$- \beta \sum_{x \in \{s,o\}} \sum_{i,j} \hat{y}^{RE}_{(i,j,x)} \log(y^{RE}_{(i,j,x)}) \tag{5}$$

where α and β are hyper-parameters to control the importance of each task.

4.3. Resource-centered network construction

As shown in Fig. 4, we propose an automated system to construct several resource-centered networks from a large corpus of scientific articles, towards better profiling on-line scientific resource. Give a set of scientific publications which are in the PDF format, we first apply our parser[5] to extract the sentences containing the citations and find the corresponding hyperlinks (urls). Then we train a resource mention prediction model to predict the head position of the on-line resource, and combine the given tail position to form a mention of the resource (r_h, r_t). Notice that, even if we get the sentence containing the citations in the first step, we may not find the corresponding resource mention (e.g. Our framework is accessible online at http://***.). In order to build the resource-centered networks for the whole corpus, we finally apply our SciResTR-IE model over single cited sentence and then extract other related scientific terms and *resource-term* relations in a single sentence. Finally, we integrate each on-line resource

---

[5] https://github.com/aqzheng/extract-link-from-pdf-using-pdf2xml-in-Windows.

Table 2
Comparison with previous models on the SciResTR dataset.

Methods	Term			Relation		
	P	R	F1	P	R	F1
LSTM + CRF	0.662	0.581	0.619	0.470	0.361	0.408
E2E Rel	0.637	0.566	0.600	0.450	0.325	0.377
SciIE	0.684	0.609	0.645	0.520	0.418	0.464
DYGIE	0.686	0.617	0.650	0.526	0.431	0.474
SciResTR-IE	**0.723**	**0.673**	**0.697**	**0.688**	**0.464**	**0.554**

Table 3
Results (F1-score) on each category predicted by the best model for each of the two tasks.

Term	F1	Relation	F1
Task	0.688	Used-for	0.658
Method	0.814	Feature-of	0.558
Data	0.725	Hyponym-of	0.724
Service	0.810	Part-of	0.476
Metric	0.721	Compare	0.398
Generic Term	0.664	Conjunction	0.510

entity and its *resource-term* relations across multiple sentences, towards constructing several resource-centered networks. As shown in Fig. 4, the red node is merged as the same on-line resource entity, according to the same hyperlink or the same resource name. The *resource-term* relationships between the cited on-line resources and other related scientific terms can be organized into different groups according to our resource-centered network construction approach. The target on-line resource can be seen as the center of a series of layers of increasing size. And Fig. 1(b) shows a part of visualized resource-centered network created by our approach.

5. Experiments

5.1. Data and metrics

To test the performance of SciResTR-IE, we do experiments on our SciResTR dataset, which is split into 3 parts: 80% for training, 10% for developing and 10% for testing. We evaluate SciResTR-IE on both term extraction and relation extraction. Moreover, we report results using the *strict* evaluation which is used in most related works. A prediction is considered correct only if its term boundary and the label of the term are both correct. And a relation is correct only if its relation label as well as the two related ordered terms are both correct (in boundary and label). Following previous work, we measure the precision, recall and F1 score. Finally, we manually evaluate the quality of the constructed resource-centered networks.

5.2. Baselines

We compare the SciResTR-IE with the following baselines on the SciResTR dataset: **LSTM + CRF** (Zheng et al., 2017) converts the joint extraction task to a sequence labeling problem based on a novel tagging scheme; **E2E Rel** (Miwa & Bansal, 2016) employs an end-to-end neural model, which is a pipelined system to extract entities and relations; **SciIE** (Luan et al., 2018) utilizes a multi-task learning framework in scientific information extraction over shared span representations; and **DyGIE** (Luan et al., 2019) uses dynamically constructed span graphs to capture the interaction of spans for entity and relation extraction.

5.3. Implementation details

We use the SciBERT-scivocab (uncased) model[6] as a sentence encoder for term extraction and relation extraction. For SciBERT fine-tuning, we use the BertAdam with the learning rate of $2e^{-5}$ and perform linear decay of the learning rate following the warm up period. We prune the scientific terms such that the maximum term width $L = 10$. When we enumerate all the possible terms, the terms that overlap with the target resource mention are not considered. Besides, we use two 100-dimensional hidden feed forward neural networks with the ReLU to compute a set of term and pairwise term scores. We select best hyper-parameters $\alpha = 0.4$ and $\beta = 0.6$ on the dev set.

5.4. Results and discussions

5.4.1. Results on SciResTR

Table 2 compares the results of our model with baselines on the two tasks: term extraction and relation extraction. And the F1-score for each term or relation type is shown in Table 3. What stands out in Table 2 is that our model outperforms all the

---

[6] https://github.com/allenai/scibert, 12-layer, 768-hidden, 12-heads.

Table 4
Comparison results of different pre-training models.

Methods	Term			Relation		
	P	R	F1	P	R	F1
SciBERT	**0.723**	**0.673**	**0.697**	**0.688**	**0.464**	**0.554**
SpanBERT	0.712	0.655	0.682	0.673	0.447	0.537
BERT	0.706	0.631	0.666	0.648	0.423	0.511

Table 5
Ablation experiments.

Methods	Term			Relation		
	P	R	F1	P	R	F1
SciResTR-IE	**0.723**	**0.673**	**0.697**	**0.688**	**0.464**	**0.554**
-Term	0.639	0.654	0.646	–	–	–
-Relation	–	–	–	0.644	0.436	0.519
-Overlap	0.678	0.578	0.624	0.656	0.413	0.506
-BPE token	0.695	0.615	0.652	0.529	0.409	0.461

baselines and achieves state-of-the-art F1 score. Specially, compared to SciIE, our model achieves significant improvements of 8.1% and 19.4% in F1 scores. Since our model is benefiting from the SciBERT encoder, which is pre-trained on a large multi-domain corpus of scientific publications and fine-tuned on our task-specific. It shows an effective solution to the data limitation. For more in-depth performance analysis, we note that the E2E Rel, a pipelined system, has poor performance on the SciResTR dataset, which is attributed to the cascading errors. While others models avoid it by jointly modeling terms extraction and *resource-term* relation extraction. It indicates that multi-task models are better at predicting term boundaries and labels on term and relation extraction.

5.4.2. Importance of pre-training selection

In the token representation component, we replace SciBERT with SpanBERT (Joshi et al., 2020) or BERT (Devlin et al., 2019) to evaluate the importance of pre-training model selection. Here SpanBERT is designed to better represent and predict spans of text, which is consistent with the purpose of our mission. Especially, we use SpanBERT-Base-Cased model[7] and BERT-Base-Uncased model,[8] whose the number of layers, the hidden size and the number of self-attention heads are the same as the SciBERT model we chose. In addition, there is only a cased version of SpanBERT, so we have to choose this. As shown in Table 4, comparing with SpanBERT and BERT, SciBERT significantly boosts performance for term extraction and relation extraction on our SciResTR dataset. This demonstrates that using unlabeled text of similar domains for pre-training can significantly improve performance. Moreover, SpanBERT performs better than BERT, where most of this gain stems from single-sentence training through the contribution of span masking and the span boundary objective.

5.4.3. Ablation studies

To investigate the effect of each component in our model, we conducted ablation experiments as shown in Table 5. Firstly, we set $\alpha = 0$ or $\beta = 0$ to evaluate the effect of multi-task learning. We observed that the performance of joint tasks was better than individual tasks, which demonstrated the effectiveness of the multi-task setup. Different from previous entity extraction tasks, we did not enumerate all the possible scientific terms, and ignored the terms that overlap with the target resource mention. The comparison results indicated that our enumeration method could improve the quality of our model effectively. Finally, to better understand how the byte-pair encoding (BPE) influences our model performance, we presented each token by matching directly in the dictionary. The effect of BPE was somewhat surprising, where removal of BPE dramatically declined the performance by 8.6% in relation extraction.

5.4.4. Resource-centered networks analysis and evaluation

Fig. 5 shows the top 10 on-line resources that other related scientific terms are directly connected to in our constructed resource-centered networks from a large corpus of scientific articles. We observe that most researchers frequently use *word2vec*, *Stanford Parser*, *Stanford Core NLP*, *WordNet* and *OpenNLP* to solve their problems, which are currently available tools. Besides, researchers are also keen to studying tasks on *Wikipedia* and *Twitter*. To the best of our knowledge, there is no approach can automatically extract fine-grained information and construct structured knowledge for on-line resources from articles in PDF format. Limited the size of selected corpus of scientific articles and no ready-to-use model, we just evaluate the quality of automatically generated resource-centered networks by asking 5 domain experts to annotate, instead of using some indicators to evaluate the knowledge graph or comparing other state-of-the-art models. More specifically, we ask five domain experts to annotate each of these 523 extracted *resource-term* relations to define ground truth labels, which corresponding to the ten on-line resources described above.
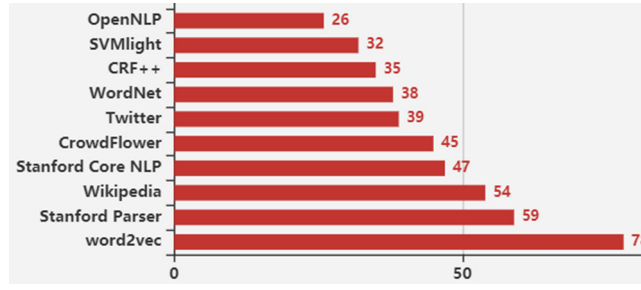
---

[7] https://github.com/facebookresearch/SpanBERT.
[8] https://github.com/google-research/bert

Fig. 5. The top 10 on-line resources that other related scientific terms are directly connected to in our constructed resource-centered networks.
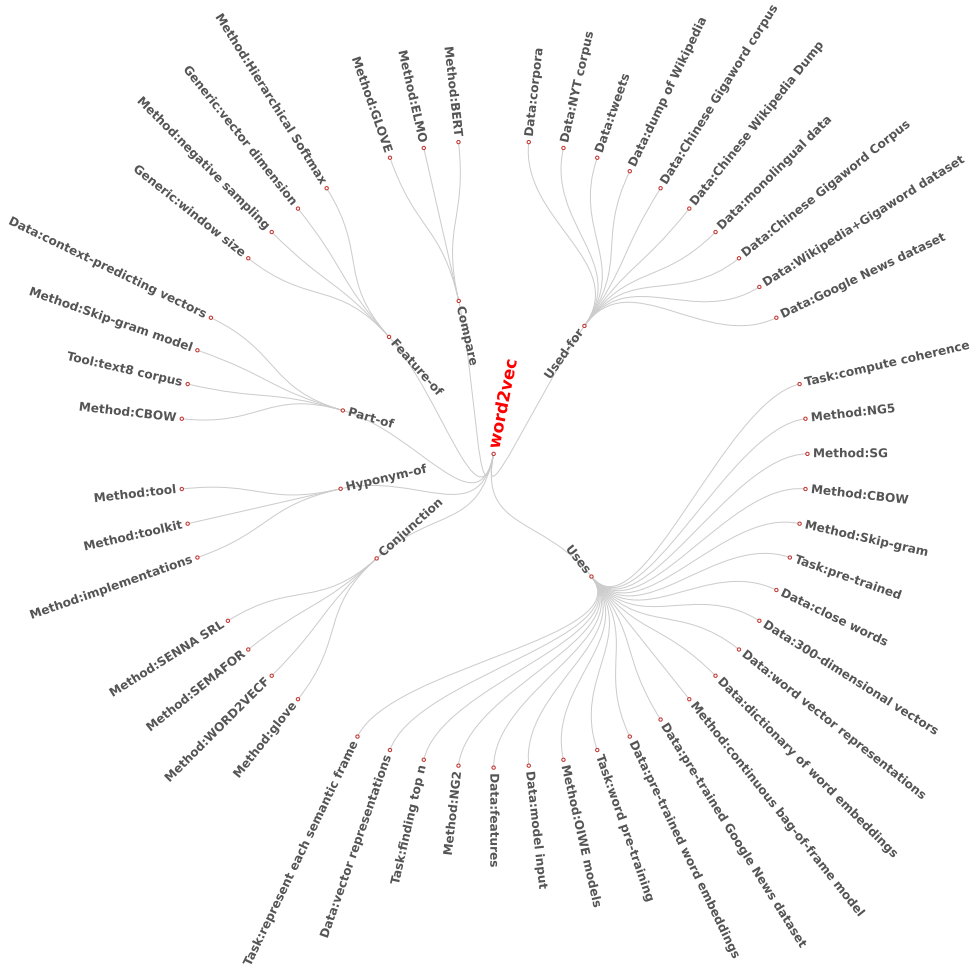

Fig. 6. The constructed resource-centered network with the most frequent neighbors of the cited on-line resource *word2vec*. For simplicity we denote Used-for (Reverse) as Uses.

4   The results that 82% of the relations are correct shows that our construction method can effectively improve the on-line scientific
5   resource profiling. Fig. 6 is an example of our automatically generated networks centered on word2vec. With the help of such
6   networks, it will be easy to answer the questions described in the Introduction section. Specially, we observe that *word2vec* can
7   be trained on different datasets, and then get the word vector representations. In addition, we can use GLOVE, ELMO or BERT to
8   replace word2vec. Our study can help researchers know more about the referenced resources they are not familiar with.

6. Conclusion

In this paper, we study the task of on-line resource profiling by using the resource citation information in scientific text. For this task we first create an annotation scheme and construct a manually labeled dataset. Then we develop a framework to jointly extract all the related scientific terms and the *resource-term* relations when given the sentence of resource citation and the mention of target on-line resource. By incorporating the associated information, our SciResTR-IE framework effectively improves the performance across all tasks. Using our proposed system, we are able to automatically organize the extracted information from a large collection of scientific literature into several resource-centered networks, towards the scientific on-line resource profiling. For future work, we will explore using our frameworks to help more tasks such as the evaluation, prediction and recommendation for scientific on-line resources.

CRediT authorship contribution statement

Anqing Zheng: Methodology, Software, Writing - original draft. **He Zhao:** Conceptualization, Methodology, Investigation, Writing - original draft. **Zhunchen Luo:** Conceptualization, Supervision, Writing - review & editing. **Chong Feng:** Writing - review & editing, Funding acquisition. **Xiaopeng Liu:** Visualization, Data curation. **Yuming Ye:** Validation, Data curation.

Acknowledgments

References

Abu-Jbara, A., & Radev, D. R. (2011). Coherent citation-based summarization of scientific papers. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA* (pp. 500–509). URL https://www.aclweb.org/anthology/P11-1051/.

Ammar, W., Groeneveld, D., Bhagavatula, C., Beltagy, I., Crawford, M., Downey, D., et al. (2018). Construction of the literature graph in semantic scholar. In *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, Volume 3 (Industry Papers)* (pp. 84–91). http://dx.doi.org/10.18653/v1/n18-3011.

Ammar, W., Peters, M., Bhagavatula, C., & Power, R. (2017). The AI2 system at semeval-2017 task 10 (scienceIE): semi-supervised end-to-end entity and relation extraction. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (pp. 592–596). Vancouver, Canada: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/S17-2097, URL https://www.aclweb.org/anthology/S17-2097.

Athar, A., & Teufel, S. (2012). Detection of Implicit Citations for Sentiment Detection. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse* (pp. 18–26).

Athar, A., & Teufel, S. (2012). Context-enhanced citation sentiment detection. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 3-8, 2012, MontrÉAl, Canada* (pp. 597–601). URL https://www.aclweb.org/anthology/N12-1073/.

Augenstein, I., Das, M., Riedel, S., Vikraman, L., & McCallum, A. (2017). Semeval 2017 task 10: Scienceie - extracting keyphrases and relations from scientific publications. In *Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval@ACL 2017, Vancouver, Canada, August 3-4, 2017* (pp. 546–555). http://dx.doi.org/10.18653/v1/S17-2091.

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. In *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, November 3-7, 2019* (pp. 3613–3618). http://dx.doi.org/10.18653/v1/D19-1371.

Bu, Y., Lu, W., Wu, Y., Chen, H., & Huang, Y. (2021). How wide is the citation impact of scientific publications? A cross-discipline and large-scale analysis. *Information Processing & Management*, *58*(1), Article 102429. http://dx.doi.org/10.1016/j.ipm.2020.102429, URL http://www.sciencedirect.com/science/article/pii/S0306457320309225.

de la Calle, G., García-Remesal, M., Chiesa, S., de la Iglesia, D., & Maojo, V. (2009). BIRI: a new approach for automatically discovering and indexing available public bioinformatics resources from the literature. *BMC Bioinformatics*, *10*, 320. http://dx.doi.org/10.1186/1471-2105-10-320.

Canino, A. (2019). Deconstructing google dataset search. *Public Services Quarterly*, *15*(3), 248–255.

Cohan, A., Ammar, W., van Zuylen, M., & Cady, F. (2019). Structural scaffolds for citation intent classification in scientific publications. In *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, June 2-7, 2019, Volume 1 (Long and Short Papers)* (pp. 3586–3596). http://dx.doi.org/10.18653/v1/n19-1361.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. In *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, June 2-7, 2019, Volume 1 (Long and Short Papers)* (pp. 4171–4186). http://dx.doi.org/10.18653/v1/n19-1423.

Do, H. H. N., Chandrasekaran, M. K., Cho, P. S., & Kan, M. (2013). Extracting and matching authors and affiliations in scholarly documents. In *13th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '13, Indianapolis, in, USA, July 22 - 26, 2013* (pp. 219–228). http://dx.doi.org/10.1145/2467696.2467703.

Duck, G., Nenadic, G., Brass, A., Robertson, D. L., & Stevens, R. (2013). Bionerds: exploring bioinformatics' database and software use through literature mining. *BMC Bioinformatics*, *14*, 194. http://dx.doi.org/10.1186/1471-2105-14-194.

Duck, G., Nenadic, G., Brass, A., Robertson, D. L., & Stevens, R. (2014). Extracting patterns of database and software usage from the bioinformatics literature. *Bioinformatics*, *30*(17), 601–608. http://dx.doi.org/10.1093/bioinformatics/btu471.

Duck, G., Nenadic, G., Filannino, M., Brass, A., Robertson, D. L., & Stevens, R. (2016). A survey of bioinformatics database and software usage through mining the literature. *PLOS ONE*, *11*(6).

Eto, M. (2019). Extended co-citation search: Graph-based document retrieval on a co-citation network containing citation context information. *Information Processing & Management*, *56*(6), http://dx.doi.org/10.1016/j.ipm.2019.05.007.

Fellbaum, C., & Vossen, P. (2012). Challenges for a multilingual wordnet. *Language Resources and Evaluation*, *46*(2), 313–326. http://dx.doi.org/10.1007/s10579-012-9186-z.

Gábor, K., Buscaldi, D., Schumann, A., QasemiZadeh, B., Zargayouna, H., & Charnois, T. (2018). Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers. In *Proceedings of the 12th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2018, New Orleans, Louisiana, USA, June 5-6, 2018* (pp. 679–688). http://dx.doi.org/10.18653/v1/s18-1111.

Gábor, K., Zargayouna, H., Buscaldi, D., Tellier, I., & Charnois, T. (2016). Semantic annotation of the ACL anthology corpus for the automatic analysis of scientific literature. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016* (pp. 3694–3701). URL http://www.lrec-conf.org/proceedings/lrec2016/summaries/870.html.

Gupta, S., & Manning, C. D. (2011). Analyzing the dynamics of research by extracting key aspects of scientific papers. In *Fifth International Joint Conference on Natural Language Processing, IJCNLP 2011, Chiang Mai, Thailand, November 8-13, 2011* (pp. 1–9). URL https://www.aclweb.org/anthology/I11-1001/.

He, Q., Kifer, D., Pei, J., Mitra, P., & Giles, C. L. (2011). Citation recommendation without author supervision. In *Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011* (pp. 755–764). http://dx.doi.org/10.1145/1935826.1935926.

He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, C. L. (2010). Context-aware citation recommendation. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010* (pp. 421–430). http://dx.doi.org/10.1145/1772690.1772734.

Hou, Y., Jochim, C., Gleize, M., Bonin, F., & Ganguly, D. (2019). Identification of tasks, datasets, evaluation metrics, and numeric scores for scientific leaderboards construction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers* (pp. 5203–5213). http://dx.doi.org/10.18653/v1/p19-1513.

Huang, W., Kataria, S., Caragea, C., Mitra, P., Giles, C. L., & Rokach, L. (2012). Recommending citations: translating papers into references. In *21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012* (pp. 1910–1914). http://dx.doi.org/10.1145/2396761.2398542.

Huang, W., Wu, Z., Liang, C., Mitra, P., & Giles, C. L. (2015). A neural probabilistic model for context based citation recommendation. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA* (pp. 2404–2410). URL http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9737.

Hussain, S. A. (2019). Google dataset search. *American Journalism, 36*(3), 416–417.

Jaidka, K., Chandrasekaran, M. K., Elizalde, B. F., Jha, R., Jones, C., Kan, M.-Y., et al. (2014). The computational linguistics summarization pilot task. In *Text Analysis Conference*.

Ji, D., Tao, P., Fei, H., & Ren, Y. (2020). An end-to-end joint model for evidence information extraction from court record document. *Information Processing & Management, 57*(6), Article 102305. http://dx.doi.org/10.1016/j.ipm.2020.102305.

Jiang, M., D'Souza, J., Auer, S., & Downie, J. S. (2020). Improving scholarly knowledge representation: Evaluating BERT-based models for scientific relation classification. In *Digital Libraries At Times of Massive Societal Transition - 22nd International Conference on Asia-Pacific Digital Libraries, ICADL 2020, Kyoto, Japan, November 30 - December 1, 2020, Proceedings* (pp. 3–19). http://dx.doi.org/10.1007/978-3-030-64452-9_1.

Joshi, M., Chen, D., Liu, Y., Weld, D. S., Zettlemoyer, L., & Levy, O. (2020). Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics, 8*, 64–77, URL https://transacl.org/ojs/index.php/tacl/article/view/1853.

Jurgens, D., Kumar, S., Hoover, R., McFarland, D. A., & Jurafsky, D. (2018). Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics, 6*, 391–406, URL https://transacl.org/ojs/index.php/tacl/article/view/1266.

Luan, Y., He, L., Ostendorf, M., & Hajishirzi, H. (2018). Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018* (pp. 3219–3232). http://dx.doi.org/10.18653/v1/d18-1360.

Luan, Y., Ostendorf, M., & Hajishirzi, H. (2017). Scientific information extraction with semi-supervised neural tagging. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017* (pp. 2641–2651). http://dx.doi.org/10.18653/v1/d17-1279.

Luan, Y., Wadden, D., He, L., Shah, A., Ostendorf, M., & Hajishirzi, H. (2019). A general framework for information extraction using dynamic span graphs. In *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, June 2-7, 2019, Volume 1 (Long and Short Papers)* (pp. 3036–3046). http://dx.doi.org/10.18653/v1/n19-1308.

Miller, G. A. (1995). Wordnet: A lexical database for english. *Communications of the ACM, 38*(11), 39–41. http://dx.doi.org/10.1145/219717.219748, URL http://doi.acm.org/10.1145/219717.219748.

Miwa, M., & Bansal, M. (2016). End-to-end relation extraction using LSTMs on sequences and tree structures. *1*, In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers* (pp. 1105–1116). http://dx.doi.org/10.18653/v1/p16-1105.

Oral, B., Emekligil, E., Arslan, S., & Eryigit, G. (2020). Information extraction from text intensive and visually rich banking documents. *Information Processing & Management, 57*(6), Article 102361. http://dx.doi.org/10.1016/j.ipm.2020.102361.

Prabhakaran, V., Hamilton, W. L., McFarland, D. A., & Jurafsky, D. (2016). Predicting the rise and fall of scientific topics from trends in their rhetorical framing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers* (pp. 1170–1180). http://dx.doi.org/10.18653/v1/p16-1111.

Qazvinian, V., & Radev, D. R. (2008). Scientific paper summarization using citation summary networks. In *COLING 2008, 22nd International Conference on Computational Linguistics, Proceedings of the Conference, 18-22 August 2008, Manchester, UK* (pp. 689–696). URL https://www.aclweb.org/anthology/C08-1087/.

Rei, M. (2017). Semi-supervised multitask learning for sequence labeling. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers* (pp. 2121–2130). http://dx.doi.org/10.18653/v1/P17-1194.

Safder, I., Hassan, S., Visvizi, A., Noraset, T., Nawaz, R., & Tuarob, S. (2020). Deep learning-based extraction of algorithmic metadata in full-text scholarly documents. *Information Processing & Management, 57*(6), Article 102269. http://dx.doi.org/10.1016/j.ipm.2020.102269.

Sennrich, R., Haddow, B., & Birch, A. (2016). Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1715–1725). Berlin, Germany: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/P16-1162, URL https://www.aclweb.org/anthology/P16-1162.

Sim, Y., Smith, N. A., & Smith, D. A. (2012). Discovering factions in the computational linguistics community. In *Proceedings of the Special Workshop on Rediscovering 50 Years of Discoveries@ACL 2012, Jeju Island, Korea, July 10, 2012* (pp. 22–32). URL https://www.aclweb.org/anthology/W12-3203/.

Tang, J., & Zhang, J. (2009). A discriminative approach to topic-based citation recommendation. In *Advances in Knowledge Discovery and Data Mining, 13th Pacific-Asia Conference, PAKDD 2009, Bangkok, Thailand, April 27-30, 2009, Proceedings* (pp. 572–579). http://dx.doi.org/10.1007/978-3-642-01307-2_55.

Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation function. In *EMNLP 2006, Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, 22-23 July 2006, Sydney, Australia* (pp. 103–110). URL https://www.aclweb.org/anthology/W06-1613/.

Tsai, C., Kundu, G., & Roth, D. (2013). Concept-based analysis of scientific literature. In *22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013* (pp. 1733–1738). http://dx.doi.org/10.1145/2505515.2505613.

Vogel, A., & Jurafsky, D. (2012). He said, she said: Gender in the ACL anthology. In *Proceedings of the Special Workshop on Rediscovering 50 Years of Discoveries@ACL 2012, Jeju Island, Korea, July 10, 2012* (pp. 33–41). URL https://www.aclweb.org/anthology/W12-3204/.

Vrandecic, D. (2019). Describing datasets in wikidata. In *15th International Conference on EScience, EScience 2019, San Diego, CA, USA, September 24-27, 2019* (pp. 528–529). http://dx.doi.org/10.1109/eScience.2019.00070.

Wadden, D., Wennberg, U., Luan, Y., & Hajishirzi, H. (2019). Entity, relation, and event extraction with contextualized span representations. In *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, November 3-7, 2019* (pp. 5783–5788). http://dx.doi.org/10.18653/v1/D19-1585.

Yamamoto, Y., & Takagi, T. (2007). Orefil: an online resource finder for life sciences. *BMC Bioinformatics*, *8*, http://dx.doi.org/10.1186/1471-2105-8-287.

Zhao, H., Luo, Z., Feng, C., & Ye, Y. (2019). A context-based framework for resource citation classification in scientific literatures. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019* (pp. 1041–1044). http://dx.doi.org/10.1145/3331184.3331348.

Zhao, H., Luo, Z., Feng, C., Zheng, A., & Liu, X. (2019). A context-based framework for modeling the role and function of on-line resource citations in scientific literature. In *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, November 3-7, 2019* (pp. 5205–5214). http://dx.doi.org/10.18653/v1/D19-1524.

Zheng, S., Wang, F., Bao, H., Hao, Y., Zhou, P., & Xu, B. (2017). Joint extraction of entities and relations based on a novel tagging scheme. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers* (pp. 1227–1236). http://dx.doi.org/10.18653/v1/P17-1113.