

Master's Thesis

---

**Semantic approaches to citation  
recommendation**

---

Tarek Saier

Examiners: Prof. Dr. Georg Lausen  
Prof. Dr. Christian Schindelhauer

Albert-Ludwigs-University Freiburg  
Faculty of Engineering  
Department of Computer Science  
Chair of Databases and Information Systems

April 30<sup>th</sup>, 2019

**Writing Period**

15. 10. 2018 – 30. 04. 2019

**First Examiner**

Prof. Dr. Georg Lausen

**Second Examiner**

Prof. Dr. Christian Schindelhauer

**Supervisor**

Dr. Michael Färber

Master-Thesis

---

**Semantic approaches to citation  
recommendation**

---

Tarek Saier

Gutachter: Prof. Dr. Georg Lausen  
Prof. Dr. Christian Schindelhauer

Albert-Ludwigs-Universität Freiburg

Technische Fakultät

Institut für Informatik

Lehrstuhl für Datenbanken und Informationssysteme

30. April 2019

**Bearbeitungszeit**

15. 10. 2018 – 30. 04. 2019

**Erstgutachter**

Prof. Dr. Georg Lausen

**Zweitgutachter**

Prof. Dr. Christian Schindelhauer

**Betreuer**

Dr. Michael Färber

# Declaration

I hereby declare, that I am the sole author and composer of my thesis and that no other sources or learning aids, other than those listed, have been used. Furthermore, I declare that I have acknowledged the work of others by providing detailed references of said work. I hereby also declare, that my Thesis has not been prepared for another examination or assignment, either wholly or excerpts thereof.

---

Place, Date

---

Signature



# Abstract

foo bar





# **Zusammenfassung**

fu bar



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Problem setting . . . . .	2
1.3	Method . . . . .	3
1.4	Contributions . . . . .	3
1.5	Document structure . . . . .	3
<b>2</b>	<b>Related Work</b>	<b>5</b>
2.1	Semantic approaches to citation recommendation . . . . .	5
2.2	Local citation recommendation . . . . .	7
<b>3</b>	<b>Background</b>	<b>11</b>
<b>4</b>	<b>Data set</b>	<b>13</b>
4.1	Existing data sets . . . . .	13
4.2	Data Set Creation . . . . .	15
4.2.1	Used Data Sets . . . . .	15
4.2.2	Pipeline overview . . . . .	16
4.2.3	L <sup>A</sup> T <sub>E</sub> X Parsing . . . . .	17
4.2.4	Reference resolution . . . . .	19
4.3	Key figures . . . . .	21
4.3.1	Creation process . . . . .	21
4.3.2	Resulting data set . . . . .	23

<b>5</b>	<b>Semantic approaches to citation recommendation</b>	<b>27</b>
5.1	Fields of Study as names entities . . . . .	27
5.2	Claims . . . . .	27
5.2.1	Tools for extracting claims . . . . .	27
5.2.2	A model of aboutness closely tied to claim structure . . . . .	28
<b>6</b>	<b>Evaluation</b>	<b>29</b>
6.1	Special considerations for citation recommendation . . . . .	29
6.2	Offline evaluation . . . . .	29
6.3	Online evaluation . . . . .	30
<b>7</b>	<b>Conclusion</b>	<b>31</b>
<b>8</b>	<b>Future work</b>	<b>33</b>
	<b>Bibliography</b>	<b>33</b>

# List of Figures

1	Four types of numbers. . . . .	12
2	Schematic representation of the data set generation process. . . . .	17
3	Number of citing documents per cited document. . . . .	24
4	Number of citation contexts per reference. . . . .	24
5	Citation flow by field of study for 13.3 million reference items. . . . .	25



# List of Tables

1	Overview of existing data sets. . . . .	14
2	Comparison of tools for parsing $\text{\LaTeX}$ . . . . .	18
3	Confidence intervals for a sample size of 300 with 297 positive results. . . .	22





# List of Algorithms

1	Stochastic Gradient Descent: Neural Network . . . . .	12
---	---	----



# 1 Introduction

## 1.1 Motivation

Citations are a central building block of scholarly discourse. They are the means by which scholars relate their research to existing work—be it in backing up claims, criticising, naming examples or engaging in any other form. Citing in a meaningful way requires an author to be aware of publications relevant to their work. Here, the ever increasing amount of new research publications per year poses a serious challenge. Even with academic search engines like Google Scholar and CiteSeerX at our disposal, identifying publications that are worthwhile to examine and appropriate to reference remains a time consuming task.

It is therefore not surprising that methods to aid researchers in these tasks have been and still are being actively researched. While diverse in nature, the common core of these efforts is the goal to utilize the automated processing of publications. This can be achieved by either extracting information from publications as they are [1, 2], or by introducing explicit semantic representations of their content to facilitate automated processing [3, 4]. Once processed, a typical method is to harvest human made citations, analyze them [5, 6] and use them for example to recommend papers [2] or aid in document exploration [7]. Although systems like this have existed for over 20 years [8, 2], there is not a lot of work looking into the use of explicit semantic representations for the recommendation of papers. This is why this thesis will investigate their application. More specifically, we will concentrate on the task of recommending papers for citation—as opposed to, for example, discovery. What this encompasses will be described in more detail in the following section.

## 1.2 Problem setting

In the broadest sense, recommending papers for citation means given an input text, suggest publications that can be referred to from within that text. In scale this can vary from specific recommendations for a section of a sentence (*local* or *context-based*), to general recommendations for a whole input document (*global*). The task can also include deciding whether or not the input contains parts that would justify inserting a citation in the first place [9]. In this thesis, we will focus on local citation recommendation with the assumption that the input always allows for/requires a citation to be put in.

Another distinction to be made is between personalized and general citation recommendation. Some approaches make use of user specific information such as an author's prior citations. Collaborative filtering approaches by nature include a user model and therefore fall into this category. While personalization can improve recommendation, it limits the approach to users that are willing to share personal information. We therefore limit ourselves to purely content based filtering approaches.

A last clarification has to be made concerning the term *explicit semantic representations*. This is to be understood as a differentiation from the mere use of unstructured text. A most prominent example for explicit semantic representations would be the structure of the Semantic Web [10]. In the context of citation recommendation as briefly outlined above this means representing citations in a semantically meaningful way as opposed to just relying on syntactical information like n-grams or bag-of-words representations.

The problem setting can be summarized as follows. To investigate is, the applicability of and requirements for the use of explicit semantic representations for content based, local citation recommendation. The following section will outline how this investigation is performed.

## 1.3 Method

In order to assess if and how explicit semantic representations can benefit citation recommendation we investigate the use of named entities as well as claim structures. For the evaluation of our models in a realistic setting we generate a large data set that allows for the extraction of precise citation marker positions. To ensure comparability with other approaches we also perform evaluations on existing data sets as far as possible.

Extend to mention offline and online eval

Extend moar

## 1.4 Contributions

The data set

Two models (even though they don't perform that well)

Insights into open problems with building claim models around citations (b/c of non-integral citation styles)

## 1.5 Document structure

foo bar

Copypasta of useful stuff below.

- Put a tilde (nbsp) in front of citations [11].
- **(TODO: Do this!)**
- **(EXTEND: Write more when new results are out!)**
- **(DRAFT: Hacky text!)**

- Chapter 1
- the colors of the Uni
  - UniBlue
  - UniRed
  - UniGrey
- a command for naming matrices **G**, and naming vectors **a**. This overwrites the default behavior of having an arrow over vectors, sticking to the naming conventions normal font for scalars, bold-lowercase for vectors, and bold-uppercase for matrices.
- named equations:

$$d(a, b) = d(b, a) \tag{1}$$

symmetry

- Use “these” for citing, not "these"
- If an equation is at the end of a sentence, add a full stop. If it’s not the end, add a comma:  $a = b + c$  (1),
- <https://en.wikipedia.org>
- Do not overuse footnotes<sup>1</sup> if possible.

---

<sup>1</sup><https://en.wikipedia.org>

## 2 Related Work

To the best of our knowledge there is, so far, almost no work investigating (1) the use of explicit semantic representations for (2) the task of local citation recommendation. We will therefore present related work in two areas. First, semantic approaches to citation/paper recommendation in general (global as well as local). Second, local citation recommendation regardless of the specifics of the approach taken.

Note that SemCir [12] (see below) is the only case of a semantic approach to local citation recommendation we are aware of. The explicit semantic representations are, however, not generated from citation contexts (local) but from papers (global) that are textually (not necessarily semantically) similar to the citation contexts.

### 2.1 Semantic approaches to citation recommendation

At a point in time where publishing research papers online was an emerging trend, Middleton et al. [13] propose a system for paper recommendation making use of a topic ontology. Based on classifying papers into topics and recording which papers a researcher would access on the web, they employ content-based filtering, collaborative filtering and a feedback mechanism to suggest papers from new topics to users. Comparing the topic ontology to a flat list of topics in two user studies, they report 7–15% more user satisfaction for the ontology case.

In a similar vein, Zhang et al. [14] propose a hybrid recommender system for papers based on semantic concept similarity. They derive concepts from CiteULike<sup>1</sup> tags and use these to measure the semantic similarity of papers and users' interest. In their evaluation they compare different settings of the approach but do not compare to other work or alternative techniques.

Jiang et al. [15] use CiteULike tags as academic concepts to build a topic model applied to paper abstracts. In a content-based recommendation setting they let volunteers judge the relevance of recommendations for a test set of 30 papers. The evaluation includes a TFIDF baseline, latent Dirichlet allocation (LDA) and an approach combining LDA with their concept model. The reported MAP@5 and NDCG@5 values are best for the LDA+concept method.

In [16] Zarrinkalam et al. enrich their metadata on research papers using multiple Linked Open Data (LOD) sources to drive a hybrid recommender system. They compare a purely content-based method using only text similarity with a second method additionally utilizing collaborative filtering and a third method furthermore using the LOD enriched data. They report recall, co-cited probability and NDCG values for various cut-off values for which the LOD enriched method consistently achieves the best performance.

With SemCiR [12] Zarrinkalam et al. introduce a content-based, global citation recommendation approach that utilizes a semantic distance measure between papers. They furthermore introduce a method for extending the measure to determine the semantic distance between an input text and a paper, which is achieved by representing the input by textually similar papers. The distance measure suggested builds on six different relational features including shared authors, venue, and overlapping in- and outgoing citations. The approach is evaluated on a 12,500 paper subset of CiteSeerX [17] in a citation re-prediction setting, using as input a paper's title, abstract and contexts in other papers where it was referred to. An evaluation of different scenarios measuring recall, co-cited probability and NDCG leads the authors to conclude that recommendation results can be improved by

---

<sup>1</sup>See <http://citeulike.org/>.



using their semantic distance measure and including citation contexts in the measurement textual similarity.

## 2.2 Local citation recommendation

Probably one of the first investigations into local citation recommendation is the work of He et al. [18]. They propose a two-step system that first identifies recommendation candidates and then re-ranks them by concept similarity. While also discussing global citation recommendation in detail, for the local case they compare recommending for a single context and recommending for all contexts within a document simultaneously. In an evaluation on the CiteSeerX data set measured by recall, co-cited probability and NDCG they find that the single context task is harder, but also, that their approach to the all contexts task achieves results comparable to and even better than some global citation recommendation methods.

In a follow-up work Huang et al. [19] build upon above work by swapping out the computationally complex concept based re-ranking method with a translational model. In this model citation contexts are treated as the source language and cited papers as words in the target language. The resulting system, RefSeer, is evaluated on two smaller data sets (CiteULike and a CiteSeer subset) and one large one (all of CiteSeer). The authors report precision, recall, Bpref and MRR values for the two smaller data sets and conclude that their system can give correct recommendations in a realistic setting—such as when only the top 10 recommendations are shown.

Huang et al. improve RefSeer with a neural probabilistic model that learns distributed representations of words and documents in [20]. They evaluate their model for local citation recommendation on the whole of CiteSeer, splitting between train and test set at the year 2011 (9M contexts train, 1.5M contexts test). Measuring MAP, MRR and NDCG they show that their model outperforms 4 different state-of-the-art approaches. An analysis on the

influence of papers' citation counts on recommendation performance shows that their approach especially exceeds other work in case of lesser cited papers (<100 citations).

In [21] Duma et al. test the effectiveness of using a variety of document internal and external text inputs to a TF-IDF model. Their data set is built from the ACL Anthology and contains 5446 citations. In a re-prediction setting the authors measure how reliably their models rank the correct paper at the top position. They conclude that a mixture of internal and external inputs outperforms either of the aforementioned used on their own.

The work presented in [22] by Duma et al. focusses on the rhetorical function of sentences. The authors classify sentences using the Core Scientific Concepts (CoreSC) scheme and investigate how their distinction can be used to improve recommendation. Evaluating on one million papers from the PubMed Central Open Access Subset they measure NDCG values and find that for several classes of input sentences significant gains in recommendation quality can be made by focussing on certain rethoric passages of candidate documents when ranking text similarity.

The Neural Citation Network (NCN) proposed by Ebesu et al. in [23] is inspired by neural machine translation, learns citation context representations as well as auhtor representations and includes an attention mechanism. For their evaluation the authors use 4.5 million citation contexts from the RefSeer data set and report NDCG, MAP, MRR and recall values. They compare against a BM25 baseline, a citation translation model as well as two variations of their model that do not make use of author representations. While BM25 is outperformed by all of the other approaches to some degree the NCN's results lead the evaluation by a distinct margin.

Kobayashi et al. [24] describe a variation of local citation recommendation they call *context-based co-citation recommendation*. The input here is a citation context *plus* one publication referred to in that contexts. The goal then is to recommend other publications that also can be used as citations in that contexts. By classifying text sections into the discourse facets "objective", "method", and "result" the authors are able to train distributed vector

representations per facet which are then used for the recommendation. They evaluate their approach on contexts from the ACL Anthology containing “enumerated co-citations” (e.g. [27,42]) and report NDCG values at a cut-off of 100. In comparison with two baseline methods their discourse facets are shown to be effective.

In [25] Jeong et al. introduce an approach to local citation recommendation using Graph Convolutional Networks (GCN) and Bidirectional Encoder Representations from Transformers (BERT). The GCN is used to capture information from the citation relationships between papers, while the pre-trained BERT is applied on the citation contexts themselves. The authors evaluate their approach on a subset of 6500 papers from the ACL Anthology and a self-created data set of close to 5000 papers. They report MAP, MRR and recall values at different cut-offs demonstrating that their BERT+CGN approach outperforms several reduced versions of the aforementioned as well as a baseline model.



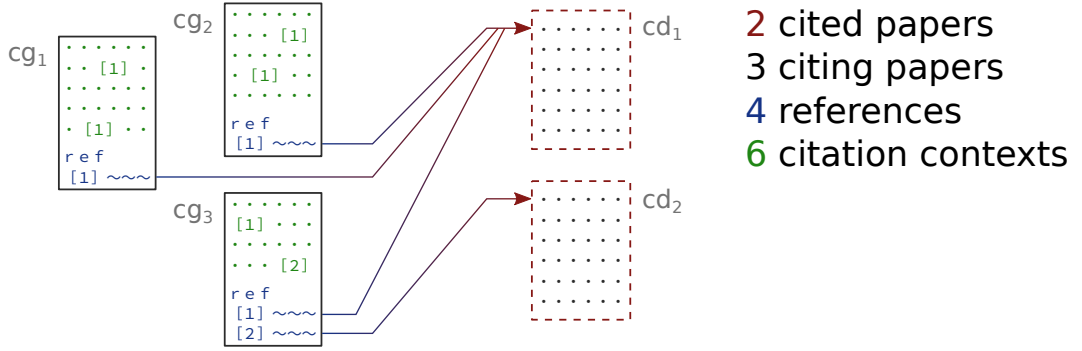
## 3 Background

explain all the things.

global/local citation

four types of numbers (citing/cited documents, reference items citation context)

reference resolution



**Figure 1:** Four types of numbers. A toy example with citation pairs  $cg_1 \rightarrow cd_1$ ,  $cg_2 \rightarrow cd_1$ ,  $cg_3 \rightarrow cd_1$ ,  $cg_3 \rightarrow cd_2$  resulting in 2 cited papers, 3 citing papers, 4 references and 6 citation contexts.

---

**Algorithm 1** Stochastic Gradient Descent: Neural Network

---

Create a mini batch of  $m$  samples  $\mathbf{x}_0 \dots \mathbf{x}_{m-1}$

**foreach** sample  $\mathbf{x}$  **do**

$\mathbf{a}^{x,0} \leftarrow \mathbf{x}$

▷ Set input activation

**foreach** Layer  $l \in \{1 \dots L-1\}$  **do**

▷ Forward pass

$\mathbf{z}^{x,l} \leftarrow \mathbf{W}^l \mathbf{a}^{x,l-1} + \mathbf{b}^l$

$\mathbf{a}^{x,l} \leftarrow \varphi(\mathbf{z}^{x,l})$

**end for**

$\delta^{x,L} \leftarrow \nabla_{\mathbf{a}} C_{\mathbf{x}} \odot \varphi'(\mathbf{z}^{x,L})$

▷ Compute error

**foreach** Layer  $l \in L-1, L-2 \dots 2$  **do**

▷ Backpropagate error

$\delta^{x,l} \leftarrow ((\mathbf{W}^{l+1})^T \delta^{x,l+1}) \odot \varphi'(\mathbf{z}^{x,l})$

**end for**

**end for**

**foreach**  $l \in L, L-1 \dots 2$  **do**

▷ Gradient descent

$\mathbf{W}^l \leftarrow \mathbf{W}^l - \frac{\eta}{m} \sum_{\mathbf{x}} \delta^{x,l} (\mathbf{a}^{x,l-1})^T$

$\mathbf{b}^l \leftarrow \mathbf{b}^l - \frac{\eta}{m} \sum_{\mathbf{x}} \delta^{x,l}$

**end for**

---

## 4 Data set

Recommender systems rely on data for their development, training and evaluation. It is therefore important to properly assess potential data sets in terms of their strengths and shortcomings—especially with regards to the task at hand. In citation recommendation, the goal is to identify papers relevant to a user input. Because of the large amount of available research, this means a recommender has to be able to find relevant publications in a large set of possible candidates in order to be considered fit for the task. As a consequence, evaluation results are likely to be more meaningful when a large data set is used. Apart from the size, the quality of data is also crucial. For local citation recommendation this means that a clean citation context, precise position of citation markers and valid citation information are desirable. With these criteria in mind we assessed existing data sets, leading us to the conclusion that—for the relatively new task of local citation recommendation—it would be worth the effort to create a new data set.

The following sections describe the details of our assessment as well as the creation process and evaluation of our new data set.

### 4.1 Existing data sets

Table 1 gives an overview of relevant existing data sets. While various recommendation domains have established quasi standard data sets, this is not yet the case in citation

---

<sup>1</sup><https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/>

**Table 1:** Overview of existing data sets.

Listed are the number of papers, nature of citation contexts, covered disciplines of citing papers and the type of global reference identifiers.

(*extractable\** is to indicate that extraction might be error-prone due to papers only being available as in PDF format.)

Data set	#Papers	Cit. context	Disciplines	Ref. IDs
CiteSeerX [17] / RefSeer [19]	5M	400 chars	(unrestricted)	internal
PubMed Central OAS <sup>1</sup>	2.3M	extractable	Biomed./Life Sci.	mixed
arXiv CS [26]	90K	1 sentence	CS	DBLP
Scholarly [27]	100K	extractable*	CS	no
ACL-AAN	18K	extractable*	CS/Comp. Ling.	no
ACL-ARC	11K	extractable*	CS/Comp. Ling.	no

recommendation. CiteSeerX is currently the most used in the field [2]. It is comparatively large, but many approaches only use subsets and generate them with varying filtering criteria. It includes pre-extracted citation contexts of 400 characters in length, whereby references are resolved to an internal set of identifiers. Unfortunately there are several quality issues with the data set. The main ones being inaccurate citation information, noisy citation contexts and cut off words at the borders of citation contexts [28].

The PubMed Central Open Access Subset is another large data set that has been used for citation based tasks [22, 29, 30, 31]. Contained publications are already processed and available in XML format. While the data set overall is comparatively clean, heterogeneous annotation of citations within the text and mixed usage of global reference identifiers (PubMed, MEDLINE, DOI, ...) make it difficult to retrieve high quality citation interlinkings of documents from the data set<sup>2</sup> [29].

Consistent global reference identifiers are given in the arXiv CS data set in the form of DBLP IDs. Linking to an existing repository of publication (meta) data has the advantage that

<sup>2</sup>To be more precise, the heterogeneity makes the usage of the data set *as is* problematic. Resolving references retrospectively would be an option but comparatively challenging in the case of PubMed because of the frequent usage of special notation in publication titles; see also: [http://www.sciplore.org/files/citrec/CITREC\\_Parser\\_Documentation.pdf](http://www.sciplore.org/files/citrec/CITREC_Parser_Documentation.pdf)



information about cited papers is readily available. The choice of DBLP restricts resolved references to the field of computer science though. Citations to, for example, publications in maths or statistics can not be resolved to a DBLP ID. A strength of the data set is that it was generated from  $\LaTeX$  source files, which makes it possible to get very clean data.

For the remaining data sets—Scholarly, ACL-AA and ACL-ARC—citing papers are only available in PDF format and references are not resolved. The two ACL sets have the additional drawback of being very small.

Above observations lead us to the conclusion that it would be worthwhile to tackle the creation of a data set that is large (in the order of CiteSeerX/RefSeer/PubMed Central OAS), clean (like the PubMed Central OAS and arXiv CS) and also offers consistent global reference IDs that don't restrict the data set to citations within the same discipline. The creation and evaluation of this data set is described in the following sections.

## 4.2 Data Set Creation

Scientific publications are usually distributed in formats targeted at human consumption (e.g. PDF) or, in cases like arXiv.org, also as source files *for* the aforementioned (e.g.  $\LaTeX$  sources for generating PDFs). Citation-based tasks, such as context-dependent citation recommendation, in contrast, require automated processing of the publications' textual contents as well as the documents' interlinking through citations. The creation of a data set for such tasks therefore encompasses two main steps: extraction of plain text and resolution of references. In the following we will describe how we approached these two steps using arXiv publication sources and the Microsoft Academic Graph (MAG) [32].

### 4.2.1 Used Data Sets

The following two resources are the basis of the data set creation process.

**arXiv.org** hosts over 1.4 million submissions from August 1991 onward [33]. They are available not only as PDF, but (in most cases) also as  $\LaTeX$  source files. The discipline most prominently represented is physics, followed by mathematics, with computer science seeing a continued increase in percentage of submissions ranking third (see Fig. 5). The availability of  $\LaTeX$  sources makes arXiv submissions particularly well suited for extracting high quality plain text and accurate citation information. So much so, that it has been used to generate ground truths for the evaluation of PDF to text conversion tools [34]. Approaches to automatically extract citation interlinks from arXiv sources by parsing  $\LaTeX$  files have existed for over 20 years [35]. Nevertheless we are not aware of any data sets for citation based tasks generated from the whole of arXiv.

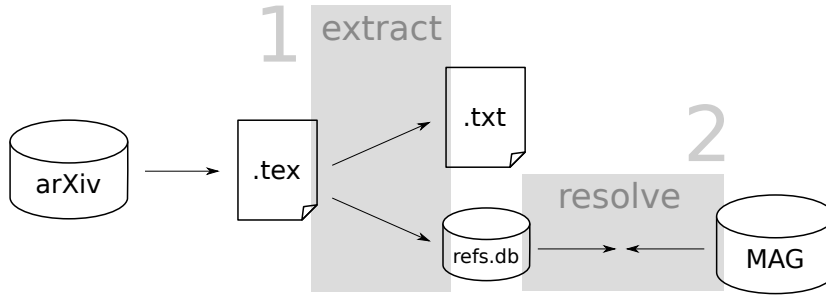
**Microsoft Academic Graph (MAG)** is a very large<sup>3</sup>, automatically generated data set on publications, related entities (authors, venues, etc.) and their interconnections through citation. While citation contexts are available to some degree, full text documents are not. The size of the MAG makes it a good target for matching reference items against it, especially given that arXiv spans several fields of study.

#### 4.2.2 Pipeline overview

As depicted in Figure 2, we start out with arXiv sources to create the data set. From these we generate, per publication, a plain text file with the document’s textual contents and a set of database entries reflecting the document’s reference section. Association between reference items and citations in the text are preserved by placing citation markers in the text. In a second step, we then iterate through all reference items in the database and match them against paper metadata records in the MAG. The result of this process are MAG paper records associated with one or more reference items, that in turn are associated with citation contexts in the plain text files. In other words, we end up with cited documents

---

<sup>3</sup>At the time of writing the MAG contains data on over 200 million publications.



**Figure 2:** Schematic representation of the data set generation process.

described by their MAG metadata and a distributed description of the document, consisting of citation contexts over many citing documents.

#### 4.2.3 $\text{\LaTeX}$ Parsing

In the following we will describe the tools considered for parsing  $\text{\LaTeX}$ , the challenges we faced in general and with regard to arXiv sources in particular, and our resulting approach.

##### Tools

We took several tools for the conversion from  $\text{\LaTeX}$  to plain text or to intermediate formats into consideration and evaluated them. Table 2 gives an overview of our results. Half of the tools failed to produce any output for a large amount of arXiv submissions we used as test input and were therefore deemed not robust enough. *GrabCite* is able to parse 78.5% of arXiv CS submissions but integrates resolving references against DBLP into the parsing process and therefore would require significant modification to fit our system architecture. *LaTeXML* and *Tralics* are both robust and can be used as  $\text{\LaTeX}$  conversion

<sup>4</sup><https://github.com/tiarno/plastex>

<sup>5</sup><https://github.com/alvinwan/texsoup>

<sup>6</sup><https://github.com/pkubowicz/opendetex>

<sup>7</sup><https://www.freebsd.org/cgi/man.cgi?query=detex>

<sup>8</sup><https://github.com/bruce-miller/LaTeXML>

<sup>9</sup><https://www-sop.inria.fr/marelle/tralics/>

**Table 2:** Comparison of tools for parsing  $\LaTeX$ .

Tool	Output	Robust	Usable w/o modification
plastex <sup>4</sup>	DOM	no	yes
TexSoup <sup>5</sup>	document tree	no	yes
opendetex <sup>6</sup> /detex <sup>7</sup>	plain text	no	yes
GrabCite[26]	plain text + resolved ref.	yes	no
LaTeXML <sup>8</sup>	XML	yes	yes
Tralics <sup>9</sup>	XML	yes	yes

tools as is. On subsequent tests we note that *LaTeXML* needs on average 7.7 seconds (3.3 if formula environments are heuristically removed beforehand) to parse an arXiv submission while *Tralics* needs 0.09. Because the quality of their output seemed comparable we chose to use *Tralics*.

## Challenges

Apart from the general difficulty of parsing  $\LaTeX$  due to its feature richness and people’s free-spirited use of it, we especially note difficulty in dealing with extra packages not included in submissions’ sources<sup>10</sup>. While *Tralics* is supposed to for example deal with *natbib* citations,<sup>11</sup> normalization of such citations lead to a decrease of citation markers not being able to be matched to their reference item from 30% to 5% in a sample of 565,613 citations we tested.

## Resulting approach

Our  $\LaTeX$  parsing solution consists of two steps. First, we flatten each arXiv submission’s sources to a single  $\LaTeX$  file using *latexexpand*<sup>12,13</sup> and normalize `\cite` commands to prevent

<sup>10</sup>arXiv.org specifically suggest the omission of such (see [https://arxiv.org/help/submit\\_tex#wegotem](https://arxiv.org/help/submit_tex#wegotem))

<sup>11</sup>See <https://www-sop.inria.fr/marelle/tralics/packages.html#natbib>

<sup>12</sup>See <https://ctan.org/pkg/latexexpand>

<sup>13</sup>We also tested flatex (<https://ctan.org/pkg/flatex>) and flap (<https://github.com/fchauvel/flap>) but got the best results with *latexexpand*.

parsing problems later on. In the second step, we then generate an XML representation of the  $\text{\LaTeX}$  document using *Tralics*, replace formulas, figures, tables and non citation references with replacement tokens and extract the plain text. Furthermore, each reference item is assigned a unique identifier, its text is stored in a database and corresponding citation markers are placed in the plain text.

#### 4.2.4 Reference resolution

A single publication's reference section follows a coherent style, many publications can follow many different bibliography styles though. Furthermore, the amount of information included in a reference item is variable and can range from the inclusion of unique identifiers like a DOI up to the omission of the cited work's title. This makes the automated identification of cited documents based on reference items a challenging and still unsolved task[1].

**Listing 4.1:** Examples of reference items.

- 
- (1) V. N. Senoguz and Q. Shafi, arXiv:hep-ph/0412102
  - (2) V.N. Senoguz and Q. Shafi, Phys. Rev. D 71 (2005) 043514.
  - (3) V. N. Şenoğuz and Q. Shafi, ‘‘Reheat temperature in supersymmetric hybrid inflation models,’’ Phys. Rev. D 71, 043514 (2005) [hep-ph/0412102].
  - (4) V.Sauli, JHEP 02, 001 (2003).
  - (5) Aaij, Roel, et al. "Search for the  $B^0_{(s)} \rightarrow \eta' \pi^0$  decay" Journal of High Energy Physics 2017.5 (2017): 158.
  - (6) According to the numerous discussions with my colleagues <removed> and <removed> an experimental verification of our theoretical predictions is feasible.
- 

Given it is, by itself, the most distinctive part of a publication, we base our reference resolution on the title of the cited work and use other pieces of information (e.g. authors' names) only in secondary steps. In the following we will describe the challenges we faced, matching arXiv.org submissions' reference items against MAG paper records and how we approached the task.

## Challenges

The reference resolution process depends on three pieces. Both data records (arXiv side and MAG side) and the matching procedure. Considering the arXiv side, reference items can be problematic when they do not contain a title or contain formulas in the title. Citing by only authors, journal name, volume and article number was often observed by us in physics papers. Formulas can become problematic because of inconsistent ways of transcribing them as plain text. One also comes across reference items that are mere comments (and would have probably better been included as footnotes) or ones that refer to non publications. These cases are less problematic because they just fail to match. The most significant challenge on the MAG side is noise. This can either prevent matches or lead to false matches.

The matching procedure ideally has a certain tolerance for minor inconsistencies (abbreviations or typos for example) while remaining accurate. Another requirement, because of the amount of data, is again speed. As every citing document can have many reference items, the number of reference items to process can easily be an order of magnitude higher than the number of citing documents. This circumstance required us to implement the matching procedure in a parallelized fashion which brought its own set of technical challenges.

## Detailed description

Our reference resolution approach can be broken down in two steps: title identification and matching. For identification of the title we first look for arXiv IDs and DOIs within the reference item. If such an ID is present we obtain the cited work's title from an arXiv metadata dump or via [crossref.org](https://www.crossref.org/)<sup>14</sup> respectively. Otherwise we use Neural ParsCit[36] to identify the portion of the item that makes up the title. Because there are cases where the identified range is just slightly off, we identify several title candidates, beginning with the one identified by Neural ParsCit and then varying the range by a few words. For

---

<sup>14</sup><https://www.crossref.org/>

the matching we then normalize the title the same way the MAG's normalized titles are generated. This means replacing everything not matching the regex word character class `\w` with a space, normalizing modified alphabet characters like umlauts (`ö→o`) and finally replacing multiple spaces with single ones. The normalized title is matched against all MAG papers. The resulting match candidates are then checked by author names. A candidate is considered good, if at least one of the author's normalized names in the MAG appears in the normalized reference item string. If, after this, still multiple candidates are left, we order them by the citation count given in the MAG and choose the first one. The last step particularly helps to mitigate rouge almost-duplicate entries in the MAG that often have few to no citations.

## 4.3 Key figures

### 4.3.1 Creation process

We used an arXiv.org source dump containing all submissions up until the end of 2017 (1,340,770 documents). 100,240 of these were only available in PDF format, leaving 1,240,530 sources. Our pipeline output 1,151,707 (92.8%) plain text files, 1,018,976 (82.1%) of which contained citation markers (for the missing 10% the parsing of `\cite` and `\bibitem` commands most likely failed). The number of reference items identified is 35,053,329, for which 56,077,906 citation markers were placed within the plain text files. This first part of the process took 59 hours to run, unparallelized on a 8 core Intel Core i7-7700 3.60GHz machine with 60 GB of memory.

Of the 35,053,329 reference items, we were able to match 14,046,239 (40.07%). For 33.14% of the reference items we could neither find an arXiv ID or DOI, nor was Neural ParsCit able to identify a title. For the remaining 26.79% a title was identified but could not be matched with the MAG. Of the matched 14 million items' titles, 50.67% were identified via Neural ParsCit. 29.67% by DOI and 19.66% by arXiv ID. Of the identified DOIs 26.8% were found as

**Table 3:** Confidence intervals for a sample size of 300 with 297 positive results as given by Wilson score interval and Jeffreys interval [37].

Confidence level	Method	Lower limit	Upper limit
0.99	Wilson	0.9613	0.9975
	Jeffreys	0.9666	0.9983
0.95	Wilson	0.9710	0.9966
	Jeffreys	0.9736	0.9972

is while 73.2% were heuristically determined<sup>15</sup>. The matching process took 103 hours, run in 10 parallel processes on a 64 core Intel Xeon Gold 6130 @ 2.10GHz machine with 500 GB of memory.

### Quality assessment of matches

To test the quality of our matches we take a random sample of 300 matched reference items and manually check if the correct record in the MAG was identified. For 300 items we note 3 errors, giving us an accuracy estimate of 96% at the worst, as shown in table 3.

The three incorrectly identified references were as follows (MAG IDs in square brackets):

1. "Eddy, J.A.: 1983, *The maunder minimum - a reappraisal*. *Solar Phys.* 89, 195. ADS."
  - matched: [2024069573]
 

"*The Maunder Minimum*" (John A. Eddy; 1976)
  - correct: [2080336740]
 

"*The Maunder Minimum: A reappraisal*" (John A. Eddy; 1983)
2. "J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani. *1-norm support vector machines*. In *Advances in Neural Information Processing Systems (NIPS)*, volume 16, pages 49–56, 2004."

<sup>15</sup>This was possible because the DOIs of articles in journals of the American Physical Society follow predictable patterns.

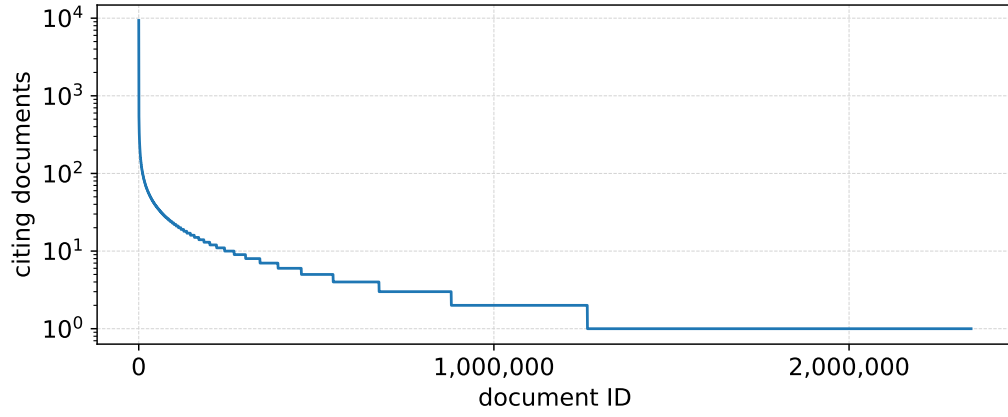


- matched: [2249237221]  
*"Support Vector Machines"* (Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani; 2013)
  - correct: [2130698119]  
*"1-norm Support Vector Machines"* (Ji Zhu, Saharon Rosset, Robert Tibshirani, Trevor J. Hastie; 2003)
3. "D. T. Limmer and D. Chandler. *The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water. The Journal of Chemical Physics*, 135(13):134503, 2011."
- matched: [2599889364]  
*"The Putative Liquid-Liquid Transition is a Liquid-Solid Transition in Atomistic Models of Water"* (David Chandler, David Limmer; 2013)
  - correct: [1977410206]  
*"The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water. II"* (David T. Limmer, David Chandler; 2011)

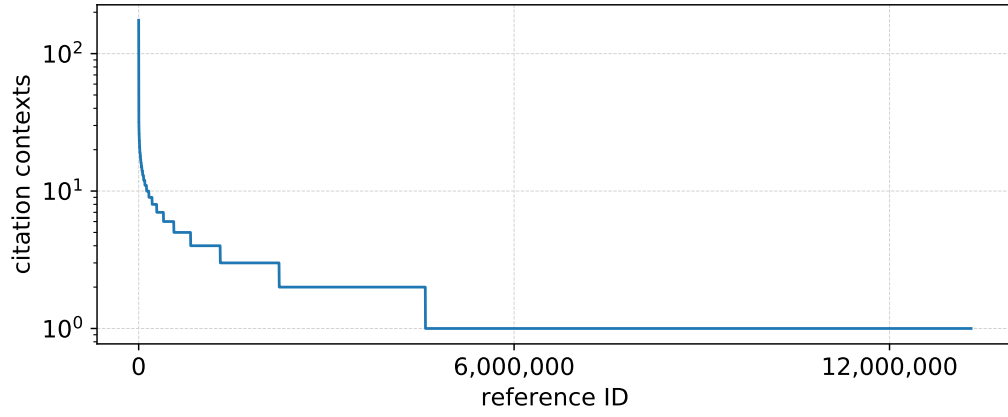
### 4.3.2 Resulting data set

Figure 3 shows the number of citing documents for all cited documents. There is one document with close to 10,000 citations, another 3 with more than 5,000 and another 10 with more than 3,000. 1,262,861 (53.89%) of the documents have at least 2 citations, 547,036 (23.34%) have at least 5. The mean number of citations is 5.68 (SD 26.82). Figure 4 shows the number of citation context per reference. 8,722,795 (65.57%) references have only one citation context, the maximum is 278, the mean 1.85 (SD 2.02). This means a cited document is described by on average  $1.85 \times 5.68 \approx 10.5$  citation contexts.

Figure 5 shows the flow of citations by field of study for all 13.3 million matched reference items. Fields of study with very small numbers of references are combined to *other* for legibility reasons. For the citing document's side, these are economics, electrical engineering and systems science, quantitative biology, quantitative finance and statistics. Combined on



**Figure 3:** Number of citing documents per cited document.

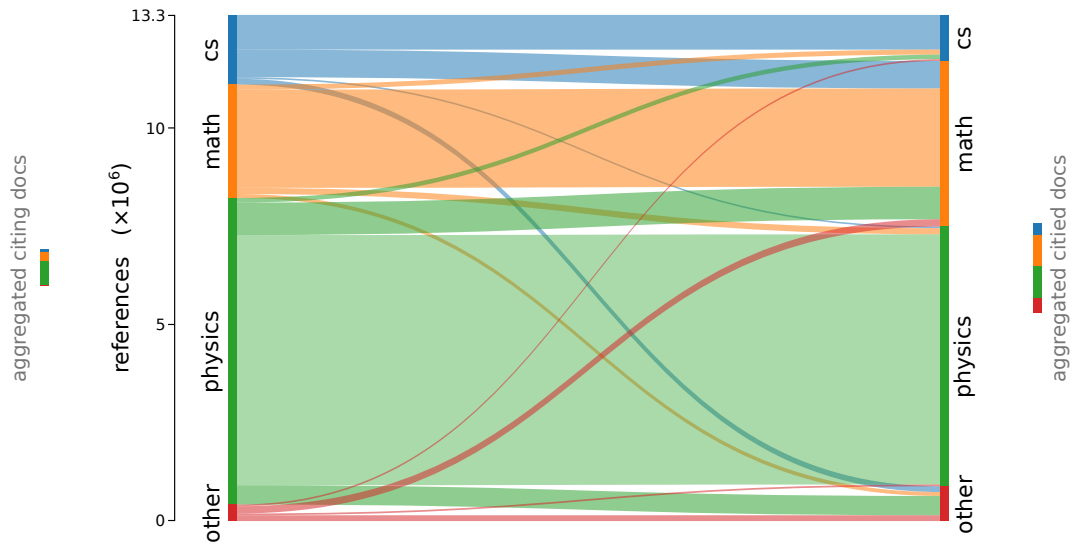


**Figure 4:** Number of citation contexts per reference.

the cited document's side are chemistry, biology, engineering, materials science, economics, geology, psychology, medicine, business, geography, sociology, political science, philosophy, environmental science and art. In rare cases<sup>16</sup> papers in the MAG (cited document's side) can have multiple main fields of study assigned. In such an event we assigned the first one we retrieved from the MAG.

To no surprise, publications in each field are cited the most from within the field itself.

<sup>16</sup>Exact numbers for the whole MAG are as follows: physics 8,682,931; math 6,701,038; cs 14,225,297; cs+math 2,254; math+phys 1,737; phys+cs 287.



**Figure 5:** Citation flow by field of study for 13.3 million reference items. For reference, the number of citing and cited documents per field of study are plotted on the sides.

Notable is, however, that the incoming citations in mathematics are the most varied (physics and computer science combined make up 38% of the citations).



## **5 Semantic approaches to citation recommendation**

types of citations (naming an entity, backing up a claim, etc.)

how citations are embedded in sentences (integral/non-integral[38, 39, 40, 41, 42])

### **5.1 Fields of Study as names entities**

name name

### **5.2 Claims**

#### **5.2.1 Tools for extracting claims**

tools tools

also: Survey on open information extraction[43]

context specific claim detection[44]

if only papers where semantically annotated as proposed in [3]

### 5.2.2 A model of aboutness closely tied to claim structure

predpatt[45, 46]

unfeasibility of use of PredPatt output as is

loosened predicate:parameter model

predicates could be grouped/clustered to represent functions as in [47]

alternative view: model gives a selective citation context derived from claim structure (cf.

concept of reference scope as sub part of citation context sentence[48, 49]

## 6 Evaluation

evaluate evaluate

implemetation pain and bad evaluation scores[50]

### 6.1 Special considerations for citation recommendation

train/test splitting (per cited doc, temporal, ...), re-recommendation, number of contexts describing a recommendation item, ...

a cited doc's role (how it is cited) can develop over time[51, 52]

relevance of time[53]

candidates are only citations within current paper[21]

### 6.2 Offline evaluation

pre-filtering experiments (knn[31], lsi, lda, fos, ...)

different evaluation settings (all, COnly, comparison to MAG, ACL (data from [54])...)

FoS alone, restrictively combined w/ BOW, only directly preceeding, ...

PP model alone, combined, ...

-> not *generally* applicable/beneficial but for certain citation types ...

also mention [24] here b/c they specifically target cases where more than one citation is applicable (could be interpreted as either *multiple (simultaneously)* for one context or *several options that are all valid by themselves but in the end a single one is to be chosen* for one contexts)

## 6.3 Online evaluation

online online



## 7 Conclusion

conclude conclude.



## 8 Future work

As a first step identify types of citations more systematically.

For different types, different models.

Proper claim model. (that could also include assessing credibility[55])

Argumentative structures. (Argumentation mining[56, 57, 58])



# Bibliography

- [1] Z. Nasar, S. W. Jaffry, and M. K. Malik, "Information extraction from scientific articles: a survey," *Scientometrics*, vol. 117, pp. 1931–1990, Dec 2018.
- [2] J. Beel, B. Gipp, S. Langer, and C. Breitingner, "Research-paper recommender systems: a literature survey," *International Journal on Digital Libraries*, vol. 17, pp. 305–338, Nov 2016.
- [3] S. Buckingham Shum, E. Motta, and J. Domingue, "Scholonto: an ontology-based digital library server for research documents and discourse," *International Journal on Digital Libraries*, vol. 3, pp. 237–248, Oct 2000. r (ch 1-3).
- [4] J. Schneider, T. Groza, and A. Passant, "A review of argumentation for the social semantic web," *Semant. web*, vol. 4, pp. 159–218, Apr. 2013.
- [5] A. Abu-Jbara, J. Ezra, and D. Radev, "Purpose and polarity of citation: Towards nlp-based bibliometrics," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 596–606, Association for Computational Linguistics, 2013.
- [6] S. Teufel, A. Siddharthan, and D. Tidhar, "Automatic classification of citation function," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, (Stroudsburg, PA, USA), pp. 103–110, Association for Computational Linguistics, 2006.

- [7] M. Berger, K. McDonough, and L. M. Seversky, "Cite2vec: Citation-driven document exploration via word embeddings," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 1–1, 01 2016.
- [8] K. D. Bollacker, S. Lawrence, and C. L. Giles, "Citeseer: An autonomous web agent for automatic retrieval and identification of interesting publications," in *Proceedings of the Second International Conference on Autonomous Agents*, AGENTS '98, (New York, NY, USA), pp. 116–123, ACM, 1998.
- [9] Q. He, D. Kifer, J. Pei, P. Mitra, and C. L. Giles, "Citation recommendation without author supervision," in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, (New York, NY, USA), pp. 755–764, ACM, 2011.
- [10] T. Berners-Lee, J. Hendler, O. Lassila, *et al.*, "The semantic web," *Scientific american*, vol. 284, no. 5, pp. 28–37, 2001.
- [11] M. J. Moravcsik and P. Murugesan, "Some results on the function and quality of citations," *Social Studies of Science*, vol. 5, no. 1, pp. 86–92, 1975.
- [12] F. Zarrinkalam and M. Kahani, "Semcir: A citation recommendation system based on a novel semantic distance measure," *Program: Electronic Library and Information Systems*, vol. 47, pp. 92–112, 2013.
- [13] S. E. Middleton, D. D. Roure, and N. Shadbolt, "Capturing knowledge of user preferences: ontologies in recommender systems," in *K-CAP*, 2001.
- [14] M. Zhang, W. Wang, and X. Li, "A paper recommender for scientific literatures based on semantic concept similarity," in *Digital Libraries: Universal and Ubiquitous Access to Information* (G. Buchanan, M. Masoodian, and S. J. Cunningham, eds.), (Berlin, Heidelberg), pp. 359–362, Springer Berlin Heidelberg, 2008.

- [15] Y. Jiang, A. Jia, Y. Feng, and D. Zhao, "Recommending academic papers via users' reading purposes," *RecSys'12 - Proceedings of the 6th ACM Conference on Recommender Systems*, 09 2012.
- [16] F. Zarrinkalam and M. Kahani, "A multi-criteria hybrid citation recommendation system based on linked data," in *2012 2nd International eConference on Computer and Knowledge Engineering (ICCCKE)*, pp. 283–288, IEEE, 2012.
- [17] C. Caragea, J. Wu, A. Ciobanu, K. Williams, J. Fernández-Ramírez, H.-H. Chen, Z. Wu, and L. Giles, "Citeseerx: A scholarly big dataset," in *Advances in Information Retrieval* (M. de Rijke, T. Kenter, A. P. de Vries, C. Zhai, F. de Jong, K. Radinsky, and K. Hofmann, eds.), (Cham), pp. 311–322, Springer International Publishing, 2014.
- [18] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, "Context-aware citation recommendation," in *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, (New York, NY, USA), pp. 421–430, ACM, 2010.
- [19] W. Huang, , P. Mitra, and C. L. Giles, "Refseer: A citation recommendation system," in *IEEE/ACM Joint Conference on Digital Libraries*, pp. 371–374, Sep. 2014.
- [20] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. L. Giles, "A neural probabilistic model for context based citation recommendation," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15*, pp. 2404–2410, AAAI Press, 2015.
- [21] D. Duma and E. Klein, "Citation resolution: A method for evaluating context-based citation recommendation systems," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, pp. 358–363, 2014.
- [22] D. Duma, E. Klein, M. Liakata, J. Ravenscroft, and A. Clare, "Rhetorical classification of anchor text for citation recommendation," *D-Lib Magazine*, vol. 22, 2016.

- [23] T. Ebesu and Y. Fang, “Neural citation network for context-aware citation recommendation,” in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’17, (New York, NY, USA), pp. 1093–1096, ACM, 2017.
- [24] Y. Kobayashi, M. Shimbo, and Y. Matsumoto, “Citation recommendation using distributed representation of discourse facets in scientific articles,” in *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, JCDL ’18, (New York, NY, USA), pp. 243–251, ACM, 2018. r (ch 1-2).
- [25] C. Jeong, J. Sion, H. Shin, E. Park, and S. Choi, “A context-aware citation recommendation model with bert and graph convolutional networks,” 02 2019.
- [26] M. Färber, A. Thiemann, and A. Jatowt, “A High-Quality Gold Standard for Citation-based Tasks,” in *Proceedings of the 11th International Conference on Language Resources and Evaluation*, LREC 2018, 2018. r.
- [27] K. Sugiyama and M.-Y. Kan, “Exploiting potential citation papers in scholarly paper recommendation,” in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL ’13, (New York, NY, USA), pp. 153–162, ACM, 2013.
- [28] D. Roy, K. Ray, and M. Mitra, “From a Scholarly Big Dataset to a Test Collection for Bibliographic Citation Recommendation,” SBD’16, 2016.
- [29] B. Gipp, N. Meuschke, and M. Lipinski, “Citrec : An evaluation framework for citation-based similarity measures based on trec genomics and pubmed central,” in *iConference 2015 Proceedings*, iSchools, 2015.
- [30] L. Galke, F. Mai, I. Vagliano, and A. Scherp, “Multi-modal adversarial autoencoders for recommendations of citations and subject labels,” in *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, UMAP ’18, (New York, NY, USA), pp. 197–205, ACM, 2018.



- [31] C. Bhagavatula, S. Feldman, R. Power, and W. Ammar, “Content-based citation recommendation,” in *NAACL-HLT*, 2018.
- [32] A. Sinha, Z. Shen, Y. Song, H. Ma, D. Eide, B.-J. P. Hsu, and K. Wang, “An overview of microsoft academic service (mas) and applications,” in *Proceedings of the 24th International Conference on World Wide Web, WWW ’15 Companion*, (New York, NY, USA), pp. 243–246, ACM, 2015. r.
- [33] P. Ginsparg, “First steps towards electronic research communication,” *Computers in Physics*, vol. 8, pp. 390–396, July 1994.
- [34] H. Bast and C. Korzen, “A benchmark and evaluation for text extraction from pdf,” in *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pp. 1–10, June 2017.
- [35] H. Nanba, “Towards multi-paper summarization using reference information,” Master’s thesis, Japan Advanced Institute of Science and Technology, 2 1998. (in Japanese).
- [36] A. Prasad, M. Kaur, and M.-Y. Kan, “Neural parsцит: A deep learning based reference string parser,” *International Journal on Digital Libraries*, vol. 19, pp. 323–337, 2018.
- [37] L. D. Brown, T. T. Cai, and A. DasGupta, “Interval estimation for a binomial proportion,” *Statistical Science*, vol. 16, no. 2, pp. 101–133, 2001.
- [38] J. Swales, *Genre analysis: English in academic and research settings*. Cambridge University Press, 1990.
- [39] K. Hyland, “Academic attribution: citation and the construction of disciplinary knowledge,” *Applied Linguistics*, vol. 20, no. 3, pp. 341–367, 1999.
- [40] P. Thompson, *A pedagogically-motivated corpus-based examination of PhD theses: Macrostructure, citation practices and uses of modal verbs*. PhD thesis, University of Reading, 2001.

- [41] A. Okamura, "Citation forms in scientific texts: Similarities and differences in l1 and l2 professional writing," *Nordic Journal of English Studies*, vol. 7, no. 3, pp. 61–81, 2008.
- [42] W. Lamers, N. J. v. Eck, L. Waltman, and H. Hoos, "Patterns in citation context: the case of the field of scientometrics," in *STI 2018 Conference proceedings*, pp. 1114–1122, Centre for Science and Technology Studies (CWTS), 2018.
- [43] C. Niklaus, M. Cetto, A. Freitas, and S. Handschuh, "A survey on open information extraction," in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3866–3878, Association for Computational Linguistics, 2018.
- [44] R. Levy, Y. Bilu, D. Hershcovich, E. Aharoni, and N. Slonim, "Context dependent claim detection," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, (Dublin, Ireland), pp. 1489–1500, Dublin City University and Association for Computational Linguistics, August 2014. r.
- [45] A. S. White, D. Reisinger, K. Sakaguchi, T. Vieira, S. Zhang, R. Rudinger, K. Rawlins, and B. Van Durme, "Universal decompositional semantics on universal dependencies," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1713–1723, Association for Computational Linguistics, 2016.
- [46] S. Zhang, R. Rudinger, and B. V. Durme, "An evaluation of predpatt and open ie via stage 1 semantic role labeling," in *IWCS 2017 — 12th International Conference on Computational Semantics — Short papers*, 2017.
- [47] K. Gábor, D. Buscaldi, A.-K. Schumann, B. QasemiZadeh, H. Zargayouna, and T. Charnois, "Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, pp. 679–688, Association for Computational Linguistics, 2018.
- [48] A. Abu-Jbara and D. Radev, "Reference scope identification in citing sentences," in *Proceedings of the 2012 Conference of the North American Chapter of the Association for*

*Computational Linguistics: Human Language Technologies*, NAACL HLT '12, (Stroudsburg, PA, USA), pp. 80–90, Association for Computational Linguistics, 2012.

- [49] R. Jha, A.-A. Jbara, V. Qazvinian, and D. R. Radev, “Nlp-driven citation analysis for scientometrics,” *Natural Language Engineering*, vol. 23, no. 1, p. 93–130, 2017.
- [50] J. Beel and S. Dinesh, “Real-world recommender systems for academia: The pain and gain in building, operating, and researching them [long version],” *CoRR*, vol. abs/1704.00156, 2017.
- [51] J. Swales, “Citation analysis and discourse analysis,” *Applied Linguistics*, vol. 7, no. 1, pp. 39–56, 1986.
- [52] J. He and C. Chen, “Temporal representations of citations for understanding the changing roles of scientific publications,” in *Front. Res. Metr. Anal.*, 2018.
- [53] J. Beel, “It’s time to consider “time” when evaluating recommender-system algorithms [proposal],” *CoRR*, vol. abs/1708.08447, 2017.
- [54] M. Färber, A. Thiemann, and A. Jatowt, “To Cite, or Not to Cite? Detecting Citation Contexts in Text,” in *Proceedings of the 40th European Conference on Information Retrieval*, ECIR 2018, 2018.
- [55] K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum, “Credibility assessment of textual claims on the web,” in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, (New York, NY, USA), pp. 2173–2178, ACM, 2016.
- [56] C. Stab and I. Gurevych, “Parsing argumentation structures in persuasive essays,” *CoRR*, vol. abs/1604.07370, 2016.
- [57] M. Lippi and P. Torroni, “Argumentation mining: State of the art and emerging trends,” *ACM Trans. Internet Technol.*, vol. 16, pp. 10:1–10:25, Mar. 2016.

- [58] I. Habernal and I. Gurevych, “Argumentation mining in user-generated web discourse,” *Comput. Linguist.*, vol. 43, pp. 125–179, Apr. 2017.

