

# **BibTeX Consistency Tool**

## **Web Service for Consistency Checks of BibTeX Files**

BACHELOR'S THESIS

submitted in partial fulfillment of the requirements for the degree of

**Bachelor of Science**

in

**Software & Information Engineering**

by

**Johannes Breitenecker**

Registration Number 01346716

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Ass. Dr.techn. Manuela Waldner, MSc

Vienna, 26<sup>th</sup> March, 2020

---

Johannes Breitenecker

---

Manuela Waldner



# Erklärung zur Verfassung der Arbeit

Johannes Breitenecker

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 26. März 2020

---

Johannes Breitenecker



# Abstract

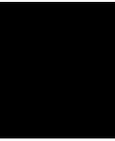
This thesis presents `cleanBibTeX`, an application that can detect and correct inconsistencies in `BIBTeX` files. `BIBTeX` is a popular reference management system for `LATEX`. Unfortunately, when `BIBTeX` files grow with respect to size or come from different sources, they can get inconsistent. For example, titles may have an inconsistent capitalization style, or the bibliography includes abbreviated author names. For a user, it can take a long time to detect such inconsistencies manually. `cleanBibTeX` allows the user to clean up a bibliography with a few mouseclicks.



# Contents

<b>Abstract</b>	<b>v</b>
<b>Contents</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Related Work</b>	<b>3</b>
2.1 Reference Managers . . . . .	3
2.2 Scripts . . . . .	4
2.3 Conclusion . . . . .	5
<b>3 Challenges</b>	<b>7</b>
3.1 Title Capitalization . . . . .	7
3.2 Inconsistent Author Names . . . . .	8
3.3 Mandatory Fields . . . . .	8
3.4 Journal Titles . . . . .	10
<b>4 Methodology</b>	<b>11</b>
4.1 User Interface . . . . .	12
4.2 Capitalization . . . . .	13
4.3 Author Names . . . . .	14
4.4 Mandatory Fields . . . . .	14
<b>5 Implementation</b>	<b>17</b>
5.1 Server . . . . .	17
5.2 Client . . . . .	20
5.3 Server-Client Communication . . . . .	21
5.4 Hosting . . . . .	22
<b>6 Results</b>	<b>23</b>
6.1 Bibliography Tests . . . . .	23
6.2 Feedback . . . . .	25
<b>7 Conclusion and Future Work</b>	<b>27</b>
	vii

7.1 Conclusion . . . . .	27
7.2 Future Work . . . . .	27
<b>List of Figures</b>	<b>29</b>
<b>List of Tables</b>	<b>31</b>
<b>Appendix</b>	<b>33</b>
<b>Bibliography</b>	<b>51</b>



# Introduction

$\LaTeX$  is one of the most popular tools for writing scientific publications at technical universities. The reason for this is that it provides useful formatting features and allows the easy integration of special symbols, for example, mathematics notation.

$\BIBTeX$  is a reference management system mostly used in combination with  $\LaTeX$ . Although, it is possible to just insert citations in  $\LaTeX$  without using  $\BIBTeX$ , using it has many advantages: First, it allows the automation of tedious tasks like sorting and formatting. The references can be sorted either alphabetically or as they appear in the text. The bibliography style defines how the author names should appear in the text [14].

The references in  $\BIBTeX$  will be saved in a separate document. This separation allows to easily switch bibliographies and add or remove references. Another advantage is that in  $\BIBTeX$  only cited references will be shown, this makes it easy to use large bibliographies that print only the references which are needed. Unfortunately,  $\BIBTeX$  files can become inconsistent over time. Especially, when the references come from different sources, for example, from reference managers. In the publication titles, it can occur that the capitalization style was enforced by reference managers and is not consistent across the file. Another challenge is that authors can be written differently but refer to the same person. Incomplete information in references is another inconsistency which can occur.

The objective of this work is to develop a tool that allows the user to solve these inconsistencies. It should give the user the possibility to detect and correct inconsistencies in a  $\BIBTeX$  file before submitting a publication. Unlike reference managers, it should not collect or retrieve references from different sources like PDFs or other data formats. The user should be able to upload a  $\BIBTeX$  file and have a list of inconsistencies, which can be either solved separately or all at once. It is up to the user to make a different suggestion for individual entries.



## Related Work

The goal of this chapter is to examine some tools to help with the organization of `BIBTEX` files. The following will explore the advantages, disadvantages, and reasons why they are unsuitable for this work.

### 2.1 Reference Managers

Mendeley [9] is a free but proprietary reference manager for scientific articles and literature. The original reference manager Mendeley Desktop runs on Microsoft Windows, macOS, and Linux. It allows the organization of many different file formats like PDFs and `BIBTEX`. A web interface exists, where the user can access the library, but many editing features are only available in the desktop version. Additionally, it is also a social media platform and allows the sharing of bibliographies with other users. Mendeley has many features for the organization of a library, which go beyond the purpose of this work. Although the tool can detect duplicates or missing fields, it lacks features to detect and correct inconsistent author names or capitalization. Also, it requires the creation of an online profile and local installation.

Zotero [12] is an open-source reference management software that runs on Windows, macOS, and Linux. It allows the collection and management of citations from many sources. It is possible to save publication data from websites, journal articles and newspapers, and PDFs. The created bibliography can also be saved to an online account. With Zotero, it is not possible to automatically clean up an imported bibliography. Although looking up missing tags online is possible, it only allows the browser to show search results from platforms like CrossRef or Google Scholar. The tags must then be changed manually. Like Mendeley [9], it requires the creation of a user account and installation of the software locally.

JabRef [8] is an open-source bibliography reference manager. It uses  $\text{BIB}\text{T}_\text{E}\text{X}$  as its native file format, but it is possible to import and export publication data in various other formats. The application runs on the Java VM and is, therefore, platform-independent. It provides an interface for editing imported  $\text{BIB}\text{T}_\text{E}\text{X}$  files. If the publication data includes a unique identifier like ISBN or DOI, it is possible to fetch the complete bibliographic information from IEEEExplore or Springer, among others. JabRef as desktop application needs to be installed with Java. A user account is not necessary, and the bibliography is saved locally.

RefWorks [11] is a proprietary cloud-based reference manager. Licenses are usually available for university institutions. It can find, access, and capture research materials from many sources. Publication data can be managed in multiple formats stored in one central workplace. It allows access to relevant citation metadata through Ex Libris' discovery services. As it runs entirely in the cloud, no installation is necessary.

CrossTeX [6] is a bibliography management tool for the command line. It is an alternative to  $\text{BIB}\text{T}_\text{E}\text{X}$  and uses an object-oriented data model that is built to minimize redundant information. CrossTeX, like cleanBibTeX, enforces consistency in titles and automates tedious tasks such as the abbreviation of author names, to prevent human errors. It is backward compatible with existing  $\text{BIB}\text{T}_\text{E}\text{X}$  bibliographies. CrossTeX can not use custom  $\text{BIB}\text{T}_\text{E}\text{X}$  files, which can be problematic when using a template is required. cleanBibTeX, on the other hand, works with standard  $\text{BIB}\text{T}_\text{E}\text{X}$  and allows the use of templates.

In general, the goal of reference managers is to organize many references in one place. They offer many features to retrieve references from different formats, not only  $\text{BIB}\text{T}_\text{E}\text{X}$ . CleanBibTeX focuses only on  $\text{BIB}\text{T}_\text{E}\text{X}$  references. It can detect and correct inconsistencies in one file only and does not store any data.

## 2.2 Scripts

Some scripts can perform checks on  $\text{BIB}\text{T}_\text{E}\text{X}$  files. The following will introduce some of them and describe their features. BibTeX Check [4] is a script written in Python which can perform checks in  $\text{BIB}\text{T}_\text{E}\text{X}$  files. It checks if specific required fields are present and set. For example, if every  $\text{BIB}\text{T}_\text{E}\text{X}$  entry from type "publication" is assigned a year. It can also perform some basic checks for names and conference proceedings, but they are very limited. For example, it detects inconsistency only in journals when they got a point in the journal name. The results of a check are exported as an HTML file with links to Google Scholar or other databases of the flawed references.

Biblib [3] is a Python script that provides several features when working with  $\text{BIB}\text{T}_\text{E}\text{X}$ . It can be installed via the Python Package Index (PIP) and can be easily integrated into other Python applications. Features include reading a  $\text{BIB}\text{T}_\text{E}\text{X}$  file, creation, deletion, and update of the entries. It can also perform some checks, for example, checking if all required tags are there and set. Additionally, it checks if the authors in the entry are in

a valid `BIBTEX` format, otherwise, an error gets thrown. It can also merge two `BIBTEX` databases together.

Autobib [2] is a Python script that helps with organizing a `BIBTEX` bibliography. It makes tasks such as retrieving bibliographical information from the web easier. It can automatically create a `BIBTEX` file from a folder that contains the documents, but unlike reference managers, it is not able to read files in multiple formats. For the queries, it uses CrossRef and Google Scholar as backend. The file-name of the documents must include the author and a specifically formatted title. All files must be named according to this schema: `(Author1, Author2) Title of the paper.pdf`.

## 2.3 Conclusion

In summary, all these described alternatives are not fully suitable for the motivation of this work. Mendeley, Zotero, and other reference managers focus more on organizing references and not on performing checks. Their goal is to build up a bibliography of many references from different sources. They include many features beyond finding and correction inconsistencies in `BIBTEX` files. Also, these applications require a local installation, which is not necessary for this work. Due to their many features, they require time to learn how to use the application. Mendeley and Zotero also require creating a user profile, which for this work, is not necessary. RefWorks is cloud-based and does not require installation, but it is not free, and licenses are mostly for university faculties.

They also lack some basic checks. In none of the applications, is it possible to change the case style of the entry title, nor is it possible to find inconsistencies in author names. The described scripts offer some checks but are more suitable to be integrated into other applications than to be used by a user. Because the backend of cleanBibTeX was written in JavaScript and not in Python, it was not possible to integrate them in the application.

cleanBibTeX differs from most reference managers in that as it offers the user a web interface, which is easy to access. The goal is not to build up a bibliography, but rather to perform consistency checks of one self-contained bibliography file and make it much faster for the user to correct these inconsistencies, than having to correct them manually.



# Challenges

The goal of this chapter is to describe some challenges when working with `BIBTEX`. In the following section, I describe four challenges and provide some background information on how `BIBTEX` is handling these.

## 3.1 Title Capitalization

Capitalization defines whether specific words will be written in lower- or uppercase. There are different bibliographical styles that have different rules for capitalization. The American Psychological Association (APA) published a style guide, which defines title case as a capitalization style in which most words are capitalized. In particular, the first word will always be capitalized, then all nouns, verbs, adjectives, adverbs, and pronouns, and at last all words with more than three letters. This leaves just short words like “and” or “but” in lowercase [1].

Sentence case in APA style is defined by the first word being capitalized, then all abbreviations and proper nouns, which are nouns which refer only to a single entity, for example, “Microsoft” or “London”. Everything else should be written in lowercase [1].

Titles in `BIBTEX` are not case-sensitive. They will be shown according to the bibliography style. This is the case even for abbreviations. For example, “CPU” will be shown as “cpu” or as “Cpu” depending on the style. To force `BIBTEX` to use a specific capitalization style, one must wrap the word into curly braces [15].

```
title = {Comparison of {CPU} Schedulers}
```

It is also possible to wrap the whole title in curly braces and preserve the capitalization of every word in the title.

```
title = {{Comparison of CPU Schedulers}}
```

If the `BIBTEX` file contains many entries from different sources, then it may result in these titles having different styles of capitalization. Usually, one must go through all titles and manually curate the titles which have an inconsistent capitalization and put braces around the words which should be capitalized. For a bibliography with many entries, this can be cumbersome.

## 3.2 Inconsistent Author Names

`BIBTEX` expects the author field to have a specific format. First comes the last name, then a comma and then the first name. If there is more than one author, they must be separated by an “and” [15]. For example:

```
author={Tesla, Nikola and Edison, Thomas Alva}
```

Author names in `BIBTEX` can often lead to inconsistencies if a bibliography contains entries with the same author but from different sources. For example, if the author’s first name is in one entry abbreviated and in another one, it is written out fully. Another problem might be that the author’s name contains a non-ASCII character like, for example, in non-English names, and in another, the character is omitted or replaced by ASCII-characters, which look similar. This can lead, for instance, to the following:

```
author={Müller, Maximillian}  
author={Mueller, Maximillian}
```

This can lead to inconsistencies, which can take a long time to fix. For every name, one must search if the same author occurs in another entry differently, which can be time-consuming for a large bibliography.

## 3.3 Mandatory Fields

`BIBTEX` differentiates between specific types of references. Every type has different mandatory or optional fields. The following table shows reference types and their mandatory fields:

<b>field:</b>	article	book	booklet	Inbook	Incollection	Inproceedings, conference	manual	masterthesis, phdthesis	misc	proceedings	techreport	unpublished
address	x	O	O	O	O	O	O	O	x	O	O	x
author	M	M1	O	M1	M	M	O	M	x	O	M	M
booktitle	x	x	x	x	M	M	x	x	x	x	x	x
chapter	x	x	x	M2	O	x	x	x	x	x	x	x
edition	x	O	x	O	O	x	O	x	x	x	x	x
editor	x	M1	x	M1	O	O	x	x	x	O	x	x
howpublished	x	x	O	x	x	x	x	x	O	x	x	x
institution	x	x	x	x	x	x	x	x	x	x	M	x
journal	M	x	x	x	x	x	x	x	x	x	x	x
month	O	O	O	O	O	O	O	O	O	O	O	O
note	O	O	O	O	O	O	O	O	O	O	O	M
number	O	O1	x	O1	O1	O1	x	x	x	O1	O	x
organization	x	x	x	x	x	O	O	x	x	O	x	x
pages	O	x	x	M2	O	O	x	x	x	x	x	x
publisher	x	M	x	M	M	O	x	x	x	O	x	x
school	x	x	x	x	x	x	x	M	x	x	x	x
series	x	O	x	O	O	O	x	x	x	O	x	x
title	M	M	M	M	M	M	M	M	O	M	M	M
type	x	x	x	O	O	x	x	O	x	x	O	x
volume	x	O1	x	O1	O1	O1	x	x	x	O1	x	x
year	M	M	O	M	M	M	O	M	O	M	M	O

Table 3.1: Types and fields of references in `BIBTEX`

M → mandatory field, M1 → one of them is mandatory, M2 → either one or both are mandatory, O → optional field, O1 → one of them is optional, x → not required  
 Additional fields can be added to the entry but will not be shown in the bibliography [13].

It may occur that mandatory fields are missing or empty. For the user, it can be difficult to find out which entries have missing fields. The opposite can also happen, so that some entries of a bibliography have way more additional and unnecessary fields than other ones. In this case, the user might want to delete non-required fields so that they are consistent with the other ones.

#### 3.4 Journal Titles

Journal titles can also lead to inconsistencies. If the references come from different sources, one journal-title may be written in abbreviated form and the other one in the full title. For example the journal title “IEEE Transactions on Visualization and Computer Graphics” could also appear in the abbreviated form “IEEE TVCG 2002”. It can be challenging to detect if both refer to the same journal title. `cleanBibTeX` does not include detection for inconsistencies in journal titles because the algorithm for it would be too complex.

# Methodology

The goal of this work is to develop cleanBibTeX, an application that helps the user to solve the challenges which were mentioned in the previous chapter. Figure 4.1 gives an overview of the workflow of cleanBibTeX. It will be developed as a web application for the following reasons:

- **Widely accessible:** The user can have access to the application on every machine with an internet connection and does not require installation or administration rights.
- **Platform-independent:** It works on all operating systems with a web browser installed.
- **Deployment:** Updates are immediately available for the end-user and deployment on each machine is not necessary.

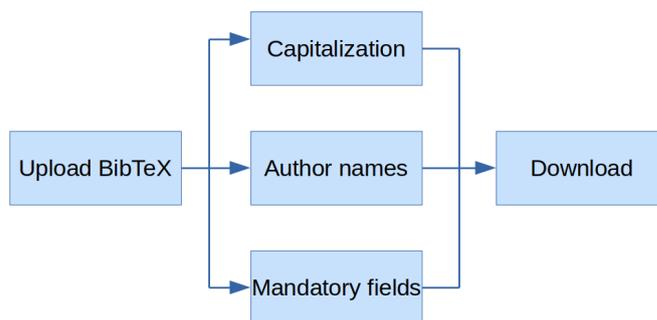


Figure 4.1: Workflow of cleanBibTeX.

## 4.1 User Interface

The UI should be straight-forward to use. At first, the user will be presented with only two options, either upload the file directly or copy the bibliography in the textbox. A status line under the textbox shows information after every user action. After uploading the file, another screen opens, and three tabs are shown: Capitalization, Author name, and Mandatory fields (see Figure 4.2). Every tab shows either a red error sign if any inconsistencies are found or a green success sign when the entries are consistent. On every tab is a summary section, where the total number of entries, and the number of found inconsistencies are listed.

Under the Summary section, there is a table with entries. In general, the consistent entries have a green background color, and the ones which are not consistent are shown in red. The user can select entries with the checkboxes on the left side and fix them with one of the buttons above the table (see Figure 4.2). After cleaning all entries from all different inconsistencies, the user can download the corrected BIB<sub>T</sub>E<sub>X</sub> file with the button at the bottom of the screen.

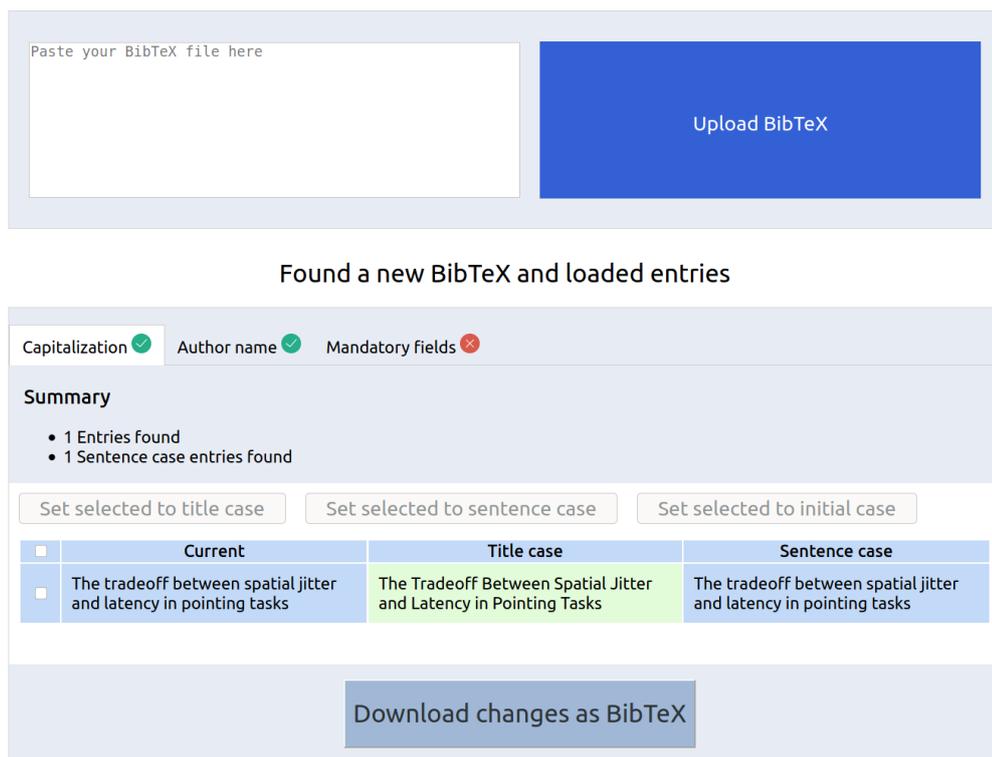


Figure 4.2: cleanBibTeX user interface.

## 4.2 Capitalization

As described in Chapter 3, title capitalization can be challenging when working with  $\text{BIB}\text{T}_{\text{E}}\text{X}$ . Often, entries of the  $\text{BIB}\text{T}_{\text{E}}\text{X}$  file come from reference managers, which force specific capitalization on the title of entries. In the capitalization tab, the goal is to get rid of all the different capitalization styles and get a consistent style for every entry. The first step is to determine the current capitalization. The title gets split up into an array of words. Then every word will be checked to see if it conforms to the rules of title or sentence case. If a word consists only of uppercase letters, it will assume that it is an acronym and will be ignored. The result will be saved in a property inside of the reference.

In the next step, the title gets converted into title and sentence case. For every word in the split title, it changes the first letter according to title and sentence case, respectively. The specific rules are described in 5.1.1. The corrected title will also be saved as property in the reference. When the user downloads the changed  $\text{BIB}\text{T}_{\text{E}}\text{X}$  file, there will be double curly braces around the title, so that the capitalization style will be shown exactly as it appears in the file. After the checks, a table with every title will be shown. The table columns from left to right are the current title form, the title in sentence case, and the title in title case. The background colors have the following meaning: green for title case, blue for sentence case and red if it is neither one of those (see Figure 4.3).

Capitalization ✖ Author name ✔ Mandatory fields ✖

**Summary**

- 3 Entries found
- 1 Title case entries found
- 2 without known case found

Set selected to title case    Set selected to sentence case    Set selected to initial case

<input type="checkbox"/>	Current	Title case	Sentence case
<input type="checkbox"/>	Handbook of mathematical Functions With Formulas, Graphs, and Mathematical Tables	Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables	Handbook of mathematical functions with formulas, graphs, and mathematical tables
<input type="checkbox"/>	Legal Knowledge And Information Systems: JURIX 2016: the Twenty-Ninth Annual Conference	Legal Knowledge and Information Systems: JURIX 2016: the Twenty-Ninth Annual Conference	Legal knowledge and information systems: JURIX 2016: the twenty ninth annual conference
<input type="checkbox"/>	Automatic Assignment of Section Structure to Texts of Dutch Court Judgments	Automatic Assignment of Section Structure to Texts of Dutch Court Judgments	Automatic assignment of section structure to texts of dutch court judgments

Download changes as BibTeX

Figure 4.3: Correction table for capitalization.

The user can select the entries which should be changed and then set it to title case, sentence case, or to the initial case by pressing one of the buttons. When setting the title to the initial case, it changes title to the form it had when it was uploaded. With this action, the title property of the reference will be set to the corrected title property the user selected. Now the user can see that the current title in the table changed to the selected one. If all titles are set to either sentence or title case, then the error sign at the tab changes to a green success sign.

### 4.3 Author Names

To detect inconsistencies in the author's name, cleanBibTeX must do some checks when parsing the uploaded bibliography. First, for every entry, the author field gets split up into the individual names with the help of regular expressions. At the same time, the tool checks if the first name of the author is abbreviated. Middle names that are abbreviated will be ignored. For every abbreviated name check, the tool searches for a name with the same last name and the same first letter for the first name. Only if there is a matching name found, which is not abbreviated, it will be saved as a suggestion.

Finally, the tool checks if there is a misspelled author name. It goes through all author names and checks if there is another author's name with an edit distance of smaller than two but not the same. The edit distance is selected by the difference of characters when substitution of a special character like the German Umlaut. For example, the names: Müller, Maximillian and Mueller, Maximillian have an edit distance of exactly two. On the frontend, there will be a table with all the author's names, which are either abbreviated or are potentially misspelled. If there is no suggestion found, using the steps above, the user can try to find the author's name online, via the "Search suggestion online" button. The CrossRef API will be used for searching because it can find publications from many different publishers. Now the user can change the author's name to the suggestion. Figure 4.4 shows the table with abbreviated author names.

### 4.4 Mandatory Fields

To detect mandatory fields, cleanBibTeX checks the entry type of each entry and looks up if all of the required fields are available and set (see Figure 4.5). If required fields are missing, they will be added to an array, which will be saved as a property inside of the entry object. On the frontend, all entries will be shown in green if they have all the required fields. Otherwise, they are shown in red with a list of missing mandatory fields. If there are missing fields, but there exists a suggestion, they are shown in blue (see Figure 4.5). The user can now select entries, which have missing mandatory fields and search for them online via the CrossRef API. If a suggestion is found, the missing fields can be added to the entry. In the case that there are entries with many non-mandatory fields, the user can select the entries and remove them from the reference.

Capitalization  Author name  Mandatory fields

**Summary**

- 3 entries found
- 3 entries with inconsistent author names found

Search suggestion online    Set author to file suggestion    Set author to online suggestion

<input type="checkbox"/>	Current Author Name	File suggestion	Online suggestion	Entry Title
<input checked="" type="checkbox"/>	Klein, K.	Klein, Karsten	no suggestion found	Immersive Collaborative Analysis of Network Connectivity: Cave-Style or Head-Mounted Display?
<input type="checkbox"/>	Stuerzlinger, Wolfgang	Stürzlinger, Wolfgang	no suggestion found	Immersive Analytics
<input type="checkbox"/>	Stürzlinger, Wolfgang	Stuerzlinger, Wolfgang	no suggestion found	The tradeoff between spatial jitter and latency in pointing tasks

Figure 4.4: Correction table for author name.

Capitalization  Author name  Mandatory fields

**Summary**

- 3 entries found
- 1 entries with missing required fields found

Search missing fields online    Remove not mandatory fields    Add field suggestion to entry

<input type="checkbox"/>	Current entry	Suggestion for missing fields
<input checked="" type="checkbox"/>	TITLE: The Tradeoff Between Spatial Jitter and Latency in Pointing Tasks AUTHOR: Pavlovych, Andriy and Stuerzlinger, Wolfgang PAGES: 187–196 ORGANIZATION: ACM Missing required fields: BOOKTITLE, YEAR,	BOOKTITLE: Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems - EICS '09 YEAR: 2009
<input type="checkbox"/>	TITLE: Immersive Analytics AUTHOR: Marriott, Kim and Schreiber, Falk and Dwyer, Tim and Klein, Karsten and Riche, Nathalie Henry and Itoh, Takayuki and Stuerzlinger, Wolfgang and Thomas, Bruce H YEAR: 2018 PUBLISHER: Springer	-

Figure 4.5: Correction table for mandatory fields.



# Implementation

This chapter describes how cleanBibTeX was implemented and which technological choices were made. In the first section, responsibilities and important code segments of the server will be described. The second section is about the technology and searching capability of the client. The third section deals with the communication of the server and the client and how authentication is realized. The last section describes how the application is made available on the web. As a web application, cleanBibTeX is implementing the client-server architecture. This architecture separates the client from the server and allows to add other client implementations that can communicate with the server.

## 5.1 Server

The primary responsibilities of the server are parsing the BIB<sub>T</sub>E<sub>X</sub> file, performing checks, and storing the changed entries for the current session. It is built with the framework Express. Express is a minimalist web framework for Node.js. It provides tooling for an HTTP server and is especially suited for single-page applications. Express was chosen because it allows to quickly set up a web application server and does not include components, which are not needed for this project, like an object-relational mapper (ORM) for database access [7]. For parsing the uploaded BIB<sub>T</sub>E<sub>X</sub> file, the server uses the third-party script “bibtex-parser”, which converts it to JavaScript objects [5]. Unfortunately, the parser is not able to read non-letter characters like “?” or “+” in entry keys. Therefore all of these special characters must be removed manually before uploading the BIB<sub>T</sub>E<sub>X</sub> file.

### 5.1.1 Capitalization Correction

For the detection and correction of the capitalization style, the server removes all previously added curly braces in the title with the following regular expression:

`TITLE.replace(/[{}]+/g, "")`. This is necessary because many reference managers add them to enforce a specific capitalization style, but for this project, the user should specify if the title should be in title or sentence case. The server split the title in an array of words with the regular expressions: `cleanedTitle.split(/[\\s]+)/`. The splitting makes it possible to loop through every word and detect and correct the capitalization. Afterward, the title gets corrected to title case, with the following method:

```
const correctWordTitleCase = word => {
  if (word === word.toUpperCase()) return word;
  if (word.includes("-")) {
    for (let i = 0; i < word.length - 1; i++) {
      const currentLetter = word[i];
      const previousLetter = word[i - 1];
      if (i > 0 && previousLetter === "-" &&
          currentLetter !== currentLetter.toUpperCase()) {
        word = word.substr(0, i) +
          currentLetter.toUpperCase() + word.substr(i + 1);
      }
    }
  }
  let correctedWord = word;
  if (word[0] === word[0].toLowerCase() && word.length >= 4) {
    correctedWord = word.charAt(0).toUpperCase() + word.slice(1);
  }
  if (word[0] === word[0].toUpperCase() && word.length < 4) {
    correctedWord = word.charAt(0).toLowerCase() + word.slice(1);
  }
  return correctedWord;
};
```

---

Listing 5.1: JavaScript code showing how a specific word will be corrected to title case. If a word which consists only of uppercase letters will be immediately returned because they are most likely abbreviations. In the case that the word includes a hyphen, it will be corrected so that after the hyphen follows a capital letter. If the word is longer than three characters it will be capitalized and otherwise written in lowercase.

### 5.1.2 Detection of Inconsistent Author Names

For the detection of abbreviated or misspelled author names, the author field will be split into the individual names with the expression: `author.split(" and ")`. The first name will be extracted and checked if it is abbreviated with the regular expressions: `/, (.+)/.exec(name)` and `/^[A-Za-z]\\./.test(firstName)`.

To search for misspelled author names, every name will be compared to every other name and checked if the edit distance is greater than zero and smaller than three. The following method shows the algorithm for the edit distance:

```
const getEditDistance = (word1, word2) => {
  const len1 = word1.length;
  const len2 = word2.length;
  const dp = new Array(len1 + 1);
```

---

```
for (let i = 0; i < dp.length; i++) {
  dp[i] = new Array(len2 + 1);
}
for (let i = 0; i <= len1; i++) {
  dp[i][0] = i;
}
for (let j = 0; j <= len2; j++) {
  dp[0][j] = j;
}
// iterate though, and check last char
for (let i = 0; i < len1; i++) {
  const c1 = word1.charAt(i);
  for (let j = 0; j < len2; j++) {
    const c2 = word2.charAt(j);

    // if last two chars equal
    if (c1 === c2) {
      // update dp value for +1 length
      dp[i + 1][j + 1] = dp[i][j];
    } else {
      const replaceVal = dp[i][j] + 1;
      const insertVal = dp[i][j + 1] + 1;
      const deleteVal = dp[i + 1][j] + 1;

      let min = replaceVal > insertVal ? insertVal : replaceVal;
      min = deleteVal > min ? min : deleteVal;
      dp[i + 1][j + 1] = min;
    }
  }
}
return dp[len1][len2];
};
```

---

Listing 5.2: JavaScript code showing the computation of the edit distance for two words.

## 5.2 Client

The client uses the React JavaScript library. React is a lightweight library for building interactive user interfaces, particularly single-page applications. React is component-based, which makes it possible to build components that manage their state and compose them to make complex user interfaces [10].

### 5.2.1 Online Search

The client's main responsibility is creating a user interface. Additionally, it also performs online searches for the author or mandatory field suggestions. Initially, it was planned to use Google Scholar to look up the names and fields, but unfortunately, no API exists for it. There are some scripts available which scrape the HTML content of Google Scholar, but it took them on average 30 seconds to get a result, which would be unacceptable from a usability perspective. For that reason, CrossRef is used instead of Google Scholar. CrossRef provides a REST-API to query a large database and is also free of charge. The following method is responsible for searching an author name:

```
searchAuthorSuggestion = (title, author) => {
  return BibtexAPI.searchAuthor(
    title.replace(/\s+/g, "+"),
    author.replace(/\s+/g, "+")
  ).then(result => {
    if (
      result != null &&
      result.message.items.length > 0 &&
      result.message.items[0].author != null
    ) {
      const foundAuthor = result.message.items[0].author
        .find(itemAuthor =>
          author.startsWith(itemAuthor.family.substr(0, 2))
        );
      if (foundAuthor != null) {
        return `${foundAuthor.family}, ${foundAuthor.given}`;
      } else return null;
    } else return null;
  });
};
const crossrefApi = 'https://api.crossref.org';
export const searchAuthor = (title, author) =>
  fetch(`${crossrefApi}/works?query.bibliographic=
    ${title}&query.author=${author}&rows=1`)
    .then(res => res.json());
```

---

Listing 5.3: JavaScript code shows the method to search online for an author name. First, the query for CrossRef will be created with the author and title of the entry. From the result, it takes the first entry and checks if there is an author name whose last name starts with the first two letters like the one in the query. If that is the case, it returns the full name of the author.

## 5.3 Server-Client Communication

The communication between the client and the server is implementing representative state transfer (REST) and the messages are encoded in JavaScript Object Notation (JSON). Following listing shows an example entry:

```
{ "entryType": "ARTICLE",
  "TITLE": "Graph clustering",
  "AUTHOR": [ { "id": 0, "name": "Schaeffer, Satu Elisa",
               "abbreviated": false, "misspelled": false ... } ],
  "JOURNAL": "Computer science review",
  "YEAR": "2007",
  "id": 2,
  "ref": "schaeffer2007graph",
  "correctionTitleCase": "Graph Clustering",
  "correctionSentenceCase": "Graph clustering",
  "capitalization": "sentenceCase",
  "initialCapitalization": "sentenceCase",
  "correctionInitialCase": "Graph clustering",
  "missingRequiredFields": [],
  ... }
```

Listing 5.4: JSON encoded message, showing a reference entry. The fields “correctionTitleCase” and “correctionSentenceCase” show the title with the changed capitalization styles and “capitalization” indicates the current capitalization style of the title. “correctionInitialCase” and “initialCapitalization” save the initial title and its capitalization style to change it back when the user has already changed it.

The server must respond to many clients at the same time and always respond to requests with the correct data from the client. For that reason, the client must authenticate itself to the server. Below is the code from the client which shows the creation of an authentication token:

```
let token = localStorage.token;

if (!token) {
  token = localStorage.token = Math.random().toString(36).substr(-8)
}
const headers = {
  Accept: "application/json",
  Authorization: token
};
```

Listing 5.5: JavaScript code showing the creation of a token for authentication. With the “localStorage” property, it is possible to store data within the user’s browser. The saved data does not expire and still exists after the browser has been restarted. If the token key in the “localStorage” does not exist, a new token will be generated, with the last eight characters of a randomly generated string. The client sends the token inside of the header in every request he makes to the server.

The server reads the token from the request and tries to access the data. The following source code gets executed when the client requests data:

```
const db = {};  
const defaultData = {  
  entries: []  
};  
const get = token => {  
  let data = db[token];  
  if (data == null) {  
    data = db[token] = clone(defaultData);  
  }  
  return data;  
};
```

---

Listing 5.6: JavaScript Code for returning the right data for the specific client. The “db” object holds the data with the token as object key. The server tries to access the data from the “db” object with the token key. If no data is found, then it creates a new object in “db” with the cloned defaultData. Otherwise, it returns the data.

### 5.4 Hosting

To make cleanBibTeX available on the web, it must be executed from a web server. cleanBibTeX is available on the following link:  
<https://clean-bibtex.herokuapp.com>

# Results

This chapter describes how effective cleanBibTeX is in detecting and correcting inconsistencies. The first section deals with an example bibliography to show where the limitations of cleanBibTeX are. The second section focuses on feedback from users who tested the tool with their own BIBTEX files.

## 6.1 Bibliography Tests

cleanBibTeX was tested with different bibliographies during the development of the application. The following listing shows a small bibliography. The bibliography consists of entries from Google Scholar, the publisher Springer, the reference managers Mendeley, Zotero, and JabRef.

```
% 1. entry from Google Scholar
@inproceedings{du2017isphere,
  title={isphere: Focus+ context sphere visualization for interactive large
  graph exploration},
  author={Du, Fan and Cao, Nan and Lin, Yu-Ru and Xu, Panpan and Tong,
  Hanghang},
  booktitle={Proceedings of the 2017 CHI Conference on Human Factors in
  Computing Systems},
  pages={2916--2927},
  year={2017}
}
% 2. entry exported from Springer
@InProceedings{10.1007/978-3-642-14527-8_1,
author="Eckersley, Peter",
editor="Atallah, Mikhail J.
and Hopper, Nicholas J.",
title="How Unique Is Your Web Browser?",
booktitle="Privacy Enhancing Technologies",
year="2010",
publisher="Springer Berlin Heidelberg",
```

## 6. RESULTS

---

```
address="Berlin, Heidelberg",
pages="1--18",
abstract="We investigate the degree..."
}
% 3. entry exported from Zotero
@inproceedings{de_groef_flowfox_nodate,
  address = {New York, NY, USA},
  series = {{CCS} '12},
  title = {{FlowFox}: {A} {Web} {Browser} with {Flexible} and {Precise}
    {Information} {Flow} {Control}},
  isbn = {978-1-4503-1651-4},
  url = {https://doi.org/10.1145/2382196.2382275},
  doi = {10.1145/2382196.2382275},
  publisher = {Association for Computing Machinery},
  author = {De Groef, Willem and Devriese, Dominique and Nikiforakis,
    Nick and Piessens, Frank},
  note = {event-place: Raleigh, North Carolina, USA},
  keywords = {information flow, web browser architecture, web security
    },
  pages = {748--759}
}
% 4. entry exported Mendeley
@inproceedings{6547132,
  author = {Nikiforakis, N. and Kapravelos, A. and Joosen, W. and Kruegel, C.
    and Piessens, F. and Vigna, G.},
  booktitle = {2013 IEEE Symposium on Security and Privacy},
  doi = {10.1109/SP.2013.43},
  issn = {1081-6011},
  keywords = {data privacy;Internet...},
  pages = {541--555},
  title = {{Cookieless Monster: Exploring the Ecosystem of Web-Based Device
    Fingerprinting}},
  year = {2013}
}
% 5. entry exported from JabRef
@inproceedings{herrera2013ping,
  title={Ping-pong: Using smartphones to measure distances and relative
    positions},
  author={Herrera, Jorge and Kim, Hyung Suk},
  volume={20},
  number={1},
  pages={055003},
  organization={Acoustical Society of America}
}
```

---

Listing 6.1: BibTeX file example.

First, we want to have a look at the capitalization. The title “isphere: Focus+ context sphere visualization for interactive large graph exploration”, was not correctly detected as sentence case. The reason for that is that it starts with a lowercase letter, which in this case, is correct because it is a proper name. It refers to a specific product that begins with a lowercase letter. It is complicated to detect all proper names, because of this cleanBibTeX ignores them and always expects a capital letter at the beginning of a

title. The second and third entry is detected as not title or sentence case, although they are both in title case. The algorithm for the capitalization does not detect nouns and corrects every word with three or fewer characters as lower case. The fourth and fifth entries are detected correctly.

For the author’s name, the authors of the fourth entry are all abbreviated. For the authors Nikiforakis N. and Piessens F. exists a suggestion from the file, the other ones have to be looked up online. Unfortunately, it can happen that the online suggestion also consists of abbreviated names. For the mandatory fields, the third and fifth entries have missing required fields. In both titles, the fields “booktitle” and “year” are missing. An online search did find all fields for the third entry, but the “booktitle” for the fifth entry could not be found.

The bibliography from Listing 7.2 consisted of edge cases, which were picked to show the limitations of cleanBibTeX. In the following, we will look at two more realistic bibliographies “bibliography1.bib” and “bibliography2.bib”, which are printed in the Appendix. The following table shows how many capitalization styles were wrongly detected, not found author names, and mandatory fields always with the total number of entries. The table shows that most of the entries could be detected, and corrected correctly.

Checks	<b>bibliography1.bib</b>	<b>bibliography2.bib</b>
Capitalization style errors	3/61	1/30
Author name errors	1/61	2/30
Mandatory field errors	0/61	2/30

Table 6.1: Errors and not found data in cleanBibTeX.

## 6.2 Feedback

To get some input from outside, we asked others who are currently working on a paper or thesis to test their bibliography with cleanBibTeX. We got three users responding and sharing their experience with cleanBibTeX. In one response, it was noted that the meaning of the different colors in the table was not clear, and it would be helpful to provide a legend. This is understandable because the meaning of colors is different in the capitalization tab. In the capitalization tab, entries with a green background color refer to title case entries, blue ones to sentence case and red ones if neither case can be found. For the tabs, author names, and mandatory fields, the meaning of the colors are green for consistent entries, blue for inconsistent entries with suggestions, and red for inconsistent entries without suggestion. For the user, the different uses of one color can be confusing. One way to at least minimize this confusion would be to use different colors in the capitalization tab. As an alternative, the suggestion from the responder to provide a legend to explain the meaning of the colors would also be helpful.

There was a question if cleanBibTeX would change the capitalization based on the entry type. In general, capitalization is defined in the bibliography style, but cleanBibTeX allows the user to define the capitalization themselves. The corrected titles from cleanBibTeX will be shown as they are because they force the format which the user selected. The reason for this is to remove all previously defined capitalization and give the user full control of how the titles will appear in the publication.

The tester in one answer could not load his BIBTEX file in cleanBibTeX and got an error message that the file was not valid. The reason was an entry with the key: bederson1994pad++. This key caused the BibTeX parser to crash because it included the special character “+”. After removing the last two characters, the BIBTEX could be loaded as usual. Because the parser uses the third-party library “bibtex-parser” [5], it was not easily fixable but should be corrected in the future by either reimplementing the functionality or using another library.

Another responder noted that cleanBibTeX does not capitalize words like the eye, use, or age. The reason for that is that the algorithm behind capitalization does not detect if the specific word is a noun. The decision if a word is capitalized depends only on how many letters the word have. To solve this problem in the future, part of speech tagging must be used, but this could lead to performance problems when large BIBTEX files are used.

It was also noted that there are many options in the author’s names tab and that it would be helpful if the names could be selected all at once and be merged. Currently, every entry with an abbreviated name will be shown in the correction table, and if the same author appears in many entries, it can lead to many redundant entries. The solution would, be as the tester noted, to show only abbreviated author name once if they appear in many entries and change it for all at once.

Another tester suggested providing a similar detection and correction of inconsistencies for journal titles as that for author names. As noted in the next Chapter 7, the algorithm to detect abbreviated would be difficult to implement, and this feature was therefore let for future work.

# Conclusion and Future Work

This last chapter summarizes the contribution of this project and describes possible future improvements and additions.

## 7.1 Conclusion

cleanBibTeX is a tool that helps BIBTEX users to keep their bibliography consistent before submitting their publication. As described in Chapter 3, there are some challenges when working with BIBTEX. In Chapter 2, it is mentioned that there are some existing tools which can correct some inconsistencies in BIBTEX, but they either are not easy to use, or they have a completely different goal. The results in Chapter 6 showed that the application could effectively minimize inconsistencies but also does have its limitations. With cleanBibTeX, we wanted to build a tool that is easy to access and does only concentrate on cleaning up inconsistencies. The feedback from the tester revealed that such a tool could be useful to many people who write scientific publications regularly.

## 7.2 Future Work

In the following, I want to describe some possible improvements and suggestions which were noted in the project presentation. One suggestion was to parse PDFs because then the exact author name could be extracted directly from the document. The reason for not including this feature is that cleanBibTeX focuses only on detecting and correcting inconsistencies for one BIBTEXfile before submitting the publication, it is not intended to collect references from many sources.

Another noted suggestion was to use ORCID for searching author names. This would make the search process much more reliable. The BIBTEXfiles, which were used to test cleanBibTeX, did not include an ORCID-identifier, which is the reason why it was not

## 7. CONCLUSION AND FUTURE WORK

---

included. Nevertheless, it would be a promising feature for future releases.

The detection and correction of inconsistent journal titles was another suggested improvement. This feature would be very useful because journal titles from many different sources can often lead to inconsistencies. Because the algorithm to detect inconsistent journal titles would be quite complex, it was decided not to implement it and focus more on the other consistency checks. However, as the tool was very well received, it is definitely worth to extend its functionality in the future.

# List of Figures

4.1	Workflow of cleanBibTeX. . . . .	11
4.2	cleanBibTeX user interface. . . . .	12
4.3	Correction table for capitalization. . . . .	13
4.4	Correction table for author name. . . . .	15
4.5	Correction table for mandatory fields. . . . .	15



# List of Tables

3.1	Types and fields of references in BIBTEX . . . . .	9
6.1	Errors and not found data in cleanBibTeX. . . . .	25



# Appendix

## **bibliography1.bib:**

```
@inproceedings{bastian2009gephi,
  title={Gephi: an open source software for exploring and manipulating
  networks},
  author={Bastian, Mathieu and Heymann, Sebastien and Jacomy, Mathieu},
  booktitle={Third international AAAI conference on weblogs and social media
  },
  year={2009}
}

@inproceedings{pavlovych2009tradeoff,
  title={The tradeoff between spatial jitter and latency in pointing tasks},
  author={Pavlovych, Andriy and Stuerzlinger, Wolfgang},
  booktitle={Proceedings of the 1st ACM SIGCHI symposium on Engineering
  interactive computing systems},
  pages={187--196},
  year={2009},
  organization={ACM}
}

@article{schaefffer2007graph,
  title={Graph clustering},
  author={Schaeffer, Satu Elisa},
  journal={Computer science review},
  volume={1},
  number={1},
  pages={27--64},
  year={2007},
  publisher={Elsevier}
}

@inproceedings{butcher2017building,
  title={Building immersive data visualizations for the web},
  author={Butcher, Peter WS and Ritsos, Panagiotis D},
  booktitle={2017 international conference on cyberworlds (CW)},
  pages={142--145},
  year={2017},
  organization={IEEE}
}

@inproceedings{debattista2018frame,
```

```

    title={Frame rate vs resolution: A subjective evaluation of spatiotemporal
        perceived quality under varying computational budgets},
    author={Debattista, Kurt and Bugeja, Keith and Spina, Sandro and Bashford-
        Rogers, Thomas and Hulusic, Vedad},
    booktitle={Computer Graphics Forum},
    volume={37},
    number={1},
    pages={363--374},
    year={2018},
    organization={Wiley Online Library}
}

@inproceedings{davis2014systematic,
    title={A systematic review of cybersickness},
    author={Davis, Simon and Nesbitt, Keith and Nalivaiko, Eugene},
    booktitle={Proceedings of the 2014 Conference on Interactive Entertainment
        },
    pages={1--9},
    year={2014},
    organization={ACM}
}

@book{marriott2018immersive,
    title={Immersive Analytics},
    author={Marriott, Kim and Schreiber, Falk and Dwyer, Tim and Klein, Karsten
        and Riche, Nathalie Henry and Itoh, Takayuki and Stuerzlinger,
        Wolfgang and Thomas, Bruce H},
    volume={11190},
    year={2018},
    publisher={Springer}
}

@incollection{marriott2018immersive_chapter,
    title={Immersive Analytics: Time to Reconsider the Value of 3D for
        Information Visualisation},
    author={Marriott, Kim and Chen, Jian and Hlawatsch, Marcel and Itoh,
        Takayuki and Nacenta, Miguel A and Reina, Guido and Stuerzlinger,
        Wolfgang},
    booktitle={Immersive Analytics},
    pages={25--55},
    year={2018},
    publisher={Springer}
}

@inproceedings{sutherland1968head,
    title={A head-mounted three dimensional display},
    author={Sutherland, Ivan E},
    booktitle={Proceedings of the December 9-11, 1968, fall joint computer
        conference, part I},
    pages={757--764},
    year={1968},
    organization={ACM}
}

```

```

@article{ware2008visualizing,
  title={Visualizing graphs in three dimensions},
  author={Ware, Colin and Mitchell, Peter},
  journal={ACM Transactions on Applied Perception (TAP)},
  volume={5},
  number={1},
  pages={2},
  year={2008},
  publisher={ACM}
}

@inproceedings{febretti2013cave2,
  title={CAVE2: a hybrid reality environment for immersive simulation and
  information analysis},
  author={Febretti, Alessandro and Nishimoto, Arthur and Thigpen, Terrance
  and Talandis, Jonas and Long, Lance and Pirtle, JD and Peterka, Tom and
  Verlo, Alan and Brown, Maxine and Plepys, Dana and others},
  booktitle={The Engineering Reality of Virtual Reality 2013},
  volume={8649},
  pages={864903},
  year={2013},
  organization={International Society for Optics and Photonics}
}

@inproceedings{kwon2015spherical,
  title={Spherical layout and rendering methods for immersive graph
  visualization},
  author={Kwon, Oh-Hyun and Muelder, Chris and Lee, Kyungwon and Ma, Kwan-Liu
  },
  booktitle={2015 IEEE Pacific Visualization Symposium (PacificVis)},
  pages={63--67},
  year={2015},
  organization={IEEE}
}

@article{kwon2016study,
  title={A study of layout, rendering, and interaction methods for immersive
  graph visualization},
  author={Kwon, Oh-Hyun and Muelder, Chris and Lee, Kyungwon and Ma, Kwan-Liu
  },
  journal={IEEE transactions on visualization and computer graphics},
  volume={22},
  number={7},
  pages={1802--1815},
  year={2016},
  publisher={IEEE}
}

@article{miksich2014matter,
  title={A matter of time: Applying a data--users--tasks design triangle to
  visual analytics of time-oriented data},
  author={Miksich, Silvia and Aigner, Wolfgang},

```

```

    journal={Computers \& Graphics},
    volume={38},
    pages={286--290},
    year={2014},
    publisher={Elsevier}
}

@article{erd6sl959random,
  title={On random graphs},
  author={Erd6s, P and Renyi, A},
  journal={Publ. Math. Debrecen},
  volume={6},
  pages={290--297},
  year={1959}
}

@article{d3library,
  author = {Bostock, Michael and Ogievetsky, Vadim and Heer, Jeffrey},
  title = {D3 Data-Driven Documents},
  journal = {IEEE Transactions on Visualization and Computer Graphics},
  issue_date = {December 2011},
  volume = {17},
  number = {12},
  month = dec,
  year = {2011},
  issn = {1077-2626},
  pages = {2301--2309},
  numpages = {9},
  doi = {10.1109/TVCG.2011.185},
  publisher = {IEEE Educational Activities Department},
  keywords = {Information visualization, user interfaces, toolkits, 2D
    graphics.},
}

@inproceedings{garcia2016perspectives,
  title={Perspectives for using virtual reality to extend visual data mining
    in information visualization},
  author={Garc{\i}a-Hern{a}ndez, Rub{e}n Jes{u}s and Anthes,
    Christoph and Wiedemann, Markus and Kranzlm{u}ller, Dieter},
  booktitle={2016 IEEE Aerospace Conference},
  pages={1--11},
  year={2016},
  organization={IEEE}
}

@inproceedings{lee2006task,
  title={Task taxonomy for graph visualization},
  author={Lee, Bongshin and Plaisant, Catherine and Parr, Cynthia Sims and
    Fekete, Jean-Daniel and Henry, Nathalie},
  booktitle={Proceedings of the 2006 AVI workshop on BEyond time and errors:
    novel evaluation methods for information visualization},
  pages={1--5},
  year={2006},
}

```

```

    organization={ACM}
}

@inproceedings{wagner2017immersive,
  title={Immersive analytics of dimensionally-reduced data scatterplots},
  author={Wagner Filho, Jorge A and Rey, Marina F and Freitas, CM and Nedel,
    Luciana},
  booktitle={2nd Workshop on Immersive Analytics. IEEE},
  year={2017}
}

@inproceedings{wagner2018immersive,
  title={Immersive visualization of abstract information: An evaluation on
    dimensionally-reduced data scatterplots},
  author={Wagner Filho, Jorge A and Rey, Marina F and Freitas, Carla MDS and
    Nedel, Luciana},
  booktitle={2018 IEEE Conference on Virtual Reality and 3D User Interfaces (
    VR)},
  pages={483--490},
  year={2018},
  organization={IEEE}
}

@article{demiralp2006cave,
  title={Cave and fishtank virtual-reality displays: A qualitative and
    quantitative comparison},
  author={Demiralp, Cagatay and Jackson, Cullen D and Karelitz, David B and
    Zhang, Song and Laidlaw, David H},
  journal={IEEE Transactions on Visualization & Computer Graphics},
  number={3},
  pages={323--330},
  year={2006},
  publisher={IEEE}
}

@article{cordeil2016immersive,
  title={Immersive collaborative analysis of network connectivity: Cave-style
    or head-mounted display?},
  author={Cordeil, Maxime and Dwyer, Tim and Klein, Karsten and Laha,
    Bireswar and Marriott, Kim and Thomas, Bruce H},
  journal={IEEE transactions on visualization and computer graphics},
  volume={23},
  number={1},
  pages={441--450},
  year={2016},
  publisher={IEEE}
}

@inproceedings{raja2004exploring,
  title={Exploring the benefits of immersion in abstract information
    visualization},
  author={Raja, Dheva and Bowman, Doug and Lucas, John and North, Chris},
  booktitle={Proc. Immersive Projection Technology Workshop},

```

```

    pages={61--69},
    year={2004}
}

@article{ware1996evaluating,
  title={Evaluating stereo and motion cues for visualizing information nets
in three dimensions},
  author={Ware, Colin and Franck, Glenn},
  journal={ACM Transactions on Graphics (TOG)},
  volume={15},
  number={2},
  pages={121--140},
  year={1996},
  publisher={ACM}
}

@inproceedings{huang2017gesture,
  title={A gesture system for graph visualization in virtual reality
environments},
  author={Huang, Yi-Jheng and Fujiwara, Takanori and Lin, Yun-Xuan and Lin,
Wen-Chieh and Ma, Kwan-Liu},
  booktitle={2017 IEEE Pacific Visualization Symposium (PacificVis)},
  pages={41--45},
  year={2017},
  organization={IEEE}
}

@inproceedings{greffard2011visual,
  title={Visual community detection: An evaluation of 2d, 3d perspective and
3d stereoscopic displays},
  author={Greffard, Nicolas and Picarougne, Fabien and Kuntz, Pascale},
  booktitle={International Symposium on Graph Drawing},
  pages={215--225},
  year={2011},
  organization={Springer}
}

@inproceedings{drogemuller2018evaluating,
  title={Evaluating Navigation Techniques for 3D Graph Visualizations in
Virtual Reality},
  author={Drogemuller, Adam and Cunningham, Andrew and Walsh, James and
Cordeil, Maxime and Ross, William and Thomas, Bruce},
  booktitle={2018 International Symposium on Big Data Visual and Immersive
Analytics (BDVA)},
  pages={1--10},
  year={2018},
  organization={IEEE}
}

@inproceedings{mine1997moving,
  title={Moving objects in space: exploiting proprioception in virtual-
environment interaction.},
  author={Mine, Mark R and Brooks Jr, Frederick P and Sequin, Carlo H},

```

```

    booktitle={SIGGRAPH},
    volume={97},
    pages={19--26},
    year={1997}
}

@inproceedings{brath20143d,
  title={3D InfoVis is here to stay: Deal with it},
  author={Brath, Richard},
  booktitle={2014 IEEE VIS International Workshop on 3DVis (3DVis)},
  pages={25--31},
  year={2014},
  organization={IEEE}
}

@inproceedings{halpin2008exploring,
  title={Exploring semantic social networks using virtual reality},
  author={Halpin, Harry and Zielinski, David J and Brady, Rachael and Kelly, Glenda},
  booktitle={International Semantic Web Conference},
  pages={599--614},
  year={2008},
  organization={Springer}
}

@article{buschel2019augmented,
  title={Augmented Reality Graph Visualizations},
  author={B{"u}schel, Wolfgang and Vogt, Stefan and Dachsel, Raimund},
  journal={IEEE computer graphics and applications},
  volume={39},
  number={3},
  pages={29--40},
  year={2019},
  publisher={IEEE}
}

@article{crutcher1995managing,
  title={Managing networks through a virtual world},
  author={Crutcher, Laurence A and Lazar, Aurel A and Feiner, Steven K and Zhou, Michelle X},
  journal={IEEE Parallel \& Distributed Technology: Systems \& Applications},
  volume={3},
  number={2},
  pages={4--13},
  year={1995},
  publisher={IEEE}
}

@inproceedings{osawa2000immersive,
  title={Immersive graph navigation using direct manipulation and gestures},
  author={Osawa, Noritaka and Asai, Kikuo and Sugimoto, Yuji Y},
  booktitle={Proceedings of the ACM symposium on Virtual reality software and technology},

```

```

    pages={147--152},
    year={2000},
    organization={ACM}
}

@inproceedings{forsberg1996aperture,
  title={Aperture based selection for immersive virtual environments},
  author={Forsberg, Andrew and Herndon, Kenneth and Zeleznik, Robert},
  booktitle={ACM Symposium on User Interface Software and Technology},
  pages={95--96},
  year={1996},
  organization={Citeseer}
}

@inproceedings{langbehn2018evaluation,
  title={Evaluation of locomotion techniques for room-scale vr: Joystick,
    teleportation, and redirected walking},
  author={Langbehn, Eike and Lubos, Paul and Steinicke, Frank},
  booktitle={Proceedings of the Virtual Reality International Conference-
    Laval Virtual},
  pages={4},
  year={2018},
  organization={ACM}
}

@inproceedings{usoh1999walking,
  title={Walking> walking-in-place> flying, in virtual environments},
  author={Usoh, Martin and Arthur, Kevin and Whitton, Mary C and Bastos, Rui
    and Steed, Anthony and Slater, Mel and Brooks Jr, Frederick P},
  booktitle={Proceedings of the 26th annual conference on Computer graphics
    and interactive techniques},
  pages={359--364},
  year={1999},
  organization={ACM Press/Addison-Wesley Publishing Co.}
}

@article{moghadam2018scene,
  title={Scene Transitions and Teleportation in Virtual Reality and the
    Implications for Spatial Awareness and Sickness},
  author={Moghadam, Kasra Rahimi and Banigan, Colin and Ragan, Eric D},
  journal={IEEE transactions on visualization and computer graphics},
  year={2018},
  publisher={IEEE}
}

@article{erra2019virtual,
  title={Virtual reality interfaces for interacting with three-dimensional
    graphs},
  author={Erra, Ugo and Malandrino, Delfina and Pepe, Luca},
  journal={International Journal of Human--Computer Interaction},
  volume={35},
  number={1},
  pages={75--88},

```

```

    year={2019},
    publisher={Taylor \& Francis}
}

@inproceedings{sadana2016redefining,
  title={Redefining a contribution for immersive visualization research},
  author={Sadana, Ramik and Setlur, Vidya and Stasko, John},
  booktitle={Proceedings of the 2016 ACM Companion on Interactive Surfaces
    and Spaces},
  pages={41--45},
  year={2016},
  organization={ACM}
}

@misc{webvr,
  title = {WebVR},
  howpublished = "\url{https://immersive-web.github.io/webvr/}",
  note = "Online; accessed 17-October-2019"
}

@misc{threejs,
  title = {three.js JavaScript 3D library},
  howpublished = \url{https://threejs.org/},
  note = "Online; accessed 06-September-2019"
}

@misc{aframe,
  title = {A-Frame: Hello WebVR},
  howpublished = "\url{https://aframe.io}",
  note = "Online; accessed 10-September-2019"
}

@misc{steamvr,
  title = {SteamVR},
  howpublished = "\url{https://store.steampowered.com/steamvr}",
  note = "Online; accessed 10-September-2019"
}

@article{chmiel2014spreading,
  title={Spreading of diseases through comorbidity networks across life and
    gender},
  author={Chmiel, Anna and Klimek, Peter and Thurner, Stefan},
  journal={New Journal of Physics},
  volume={16},
  number={11},
  pages={115013},
  year={2014},
  publisher={IOP Publishing}
}

@misc{forcegraph,
  title = {AFrame forcegraph component},

```

```

    howpublished = "\url{https://github.com/vasturiano/aframe-forcegraph-
    component}",
    note = "Online; accessed 07-September-2019"
}

@misc{d3force,
  title = {d3-force-3d},
  howpublished = "\url{https://github.com/vasturiano/d3-force-3d}",
  note = "Online; accessed 07-September-2019"
}

@inproceedings{belcher2003using,
  title={Using augmented reality for visualizing complex graphs in three
  dimensions},
  author={Belcher, Daniel and Billinghamurst, Mark and Hayes, SE and Stiles,
  Randy},
  booktitle={The Second IEEE and ACM International Symposium on Mixed and
  Augmented Reality, 2003. Proceedings.},
  pages={84--93},
  year={2003},
  organization={IEEE}
}

@inproceedings{yang2018maps,
  title={Maps and globes in virtual reality},
  author={Yang, Yalong and Jenny, Bernhard and Dwyer, Tim and Marriott, Kim
  and Chen, Haohui and Cordeil, Maxime},
  booktitle={Computer Graphics Forum},
  volume={37},
  number={3},
  pages={427--438},
  year={2018},
  organization={Wiley Online Library}
}

@article{bowman2007virtual,
  title={Virtual reality: how much immersion is enough?},
  author={Bowman, Doug A and McMahan, Ryan P},
  journal={Computer},
  volume={40},
  number={7},
  pages={36--43},
  year={2007},
  publisher={IEEE}
}

@inproceedings{drogemuller2017vrige,
  title={VRige: exploring social network interactions in immersive virtual
  environments},
  author={Drogemuller, Adam and Cunningham, Andrew and Walsh, James and Ross,
  William and Thomas, Bruce H},
  booktitle={Proceedings of the international symposium on big data visual
  analytics (BDVA). IEEE NJ, USA},

```

```

    year={2017}
}

@incollection{aggarwal2011introduction,
  title={An introduction to social network data analytics},
  author={Aggarwal, Charu C},
  booktitle={Social network data analytics},
  pages={1--15},
  year={2011},
  publisher={Springer}
}

@article{barabasi2011network,
  title={Network medicine: a network-based approach to human disease},
  author={Barab{\a}si, Albert-L{\a}szl{\o} and Gulbahce, Natali and
    Loscalzo, Joseph},
  journal={Nature reviews genetics},
  volume={12},
  number={1},
  pages={56},
  year={2011},
  publisher={Nature Publishing Group}
}

@inproceedings{shneiderman1996eyes,
  title={The eyes have it: A task by data type taxonomy for information
    visualizations},
  author={Shneiderman, Ben},
  booktitle={Proceedings 1996 IEEE symposium on visual languages},
  pages={336--343},
  year={1996},
  organization={IEEE}
}

@inproceedings{du2017isphere,
  title={isphere: Focus+ context sphere visualization for interactive large
    graph exploration},
  author={Du, Fan and Cao, Nan and Lin, Yu-Ru and Xu, Panpan and Tong,
    Hanghang},
  booktitle={Proceedings of the 2017 CHI Conference on Human Factors in
    Computing Systems},
  pages={2916--2927},
  year={2017},
  organization={ACM}
}

@book{furnas1986generalized,
  title={Generalized fisheye views},
  author={Furnas, George W},
  volume={17},
  number={4},
  year={1986},
}

```

```

    publisher={Bell Communications Research Morris Research and Engineering
    Center}
}

@article{sorger2019immersive,
  title={Immersive Analytics of Large Dynamic Networks via Overview and
  Detail Navigation},
  author={Sorger, Johannes and Waldner, Manuela and Knecht, Wolfgang and
  Arleo, Alessio},
  journal={arXiv preprint arXiv:1910.06825},
  year={2019}
}

@article{kennedy1993simulator,
  title={Simulator sickness questionnaire: An enhanced method for quantifying
  simulator sickness},
  author={Kennedy, Robert S and Lane, Norman E and Berbaum, Kevin S and
  Lilienthal, Michael G},
  journal={The international journal of aviation psychology},
  volume={3},
  number={3},
  pages={203--220},
  year={1993},
  publisher={Taylor & Francis}
}

@incollection{hart1988development,
  title={Development of NASA-TLX (Task Load Index): Results of empirical and
  theoretical research},
  author={Hart, Sandra G and Staveland, Lowell E},
  booktitle={Advances in psychology},
  volume={52},
  pages={139--183},
  year={1988},
  publisher={Elsevier}
}

@book{perry2018egocentric,
  title={Egocentric network analysis: Foundations, methods, and models},
  author={Perry, Brea L and Pescosolido, Bernice A and Borgatti, Stephen P},
  volume={44},
  year={2018},
  publisher={Cambridge university press}
}

@article{mcgrath2004you,
  title={Do you see what i want you to see? the effects of motion and spatial
  layout on viewers perceptions of graph structure},
  author={McGrath, Cathleen and Blythe, Jim},
  journal={Journal of Social Structure},
  volume={5},
  number={2},
  pages={2},

```

```
    year={2004}
  }

@inproceedings{blythe1995effect,
  title={The effect of graph layout on inference from social network data},
  author={Blythe, Jim and McGrath, Cathleen and Krackhardt, David},
  booktitle={International symposium on graph drawing},
  pages={40--51},
  year={1995},
  organization={Springer}
}
```

---

## **bibliography2.bib:**

```
@inproceedings{di_geronimo_surveying_2016,
  title={{Surveying personal device ecosystems with cross-device
    applications in mind}},
  booktitle={Proceedings of the 5th {ACM} {International} {Symposium} on {
    Pervasive} {Displays} - {PerDis} '16},
  author={Di Geronimo, Linda and Husmann, Maria and Norrie, Moira C.},
  year={2016},
}

@article{langner_vistiles:_2018,
  title={{VISTILES: Coordinating and Combining Co-located Mobile Devices for
    Visual Data Exploration}},
  journal={IEEE Transactions on Visualization and Computer Graphics},
  author={Langner, Ricardo and Horak, Tom and Dachsel, Raimund},
  year={2018},
}

@article{badam_visfer:_2019,
  title={{Visfer: Camera-based visual data transfer for cross-device
    visualization}},
  journal={Information Visualization},
  author={Badam, Sriram Karthik and Elmqvist, Niklas},
  year={2019},
}

@article{desolda_exploring_2019,
  title={{Exploring spatially-aware cross-device interaction techniques for
    mobile collaborative sensemaking}},
  journal={International Journal of Human-Computer Studies},
  author={Desolda, Giuseppe and Ardito, Carmelo and Jetter, Hans-Christian
    and Lanzilotti, Rosa},
  year={2019},
}

@article{herrera_ping-pong:_2013,
  title={{Ping-Pong: Using smartphones to measure distances and relative
    positions}},
  author={Herrera, Jorge and Kim, Hyung Suk},
  year={2013},
}

@article{huang_magmobile:_2012,
  title={{MagMobile: enhancing social interactions with rapid view-stitching
    games of mobile devices}},
  author={Huang, Da-Yuan and Lin, Chien-Pang and Hung, Yi-Ping and Chang,
    Tzu-Wen and Yu, Neng-Hao and Tsai, Min-Lun and Chen, Mike Y},
  year={2012},
}

@article{gronbaek_built-device_2016,
  title={{Built-In Device Orientation Sensors for Ad-Hoc Pairing and Spatial
    Awareness}},
```

```

    author={Gronbak, Jens Emil and O'Hara, Kenton},
    year={2016},
}

@inproceedings{hamilton_conductor:_2014,
  title={{Conductor: enabling and understanding cross-device interaction}},
  booktitle={Proceedings of the 32nd annual {ACM} conference on {Human}
    factors in computing systems - {CHI} '14},
  author={Hamilton, Peter and Wigdor, Daniel J.},
  year={2014},
}

@inproceedings{boring_touch_2010,
  title={{Touch projector: mobile interaction through video}},
  booktitle={Proceedings of the 28th international conference on {Human}
    factors in computing systems - {CHI} '10},
  author={Boring, Sebastian and Baur, Dominikus and Butz, Andreas and
    Gustafson, Sean and Baudisch, Patrick},
  year={2010},
}

@article{chung_visporter:_2014,
  title={{VisPorter: facilitating information sharing for collaborative
    sensemaking on multiple displays}},
  journal={Personal and Ubiquitous Computing},
  author={Chung, Haeyong and North, Chris and Self, Jessica Zeitz and Chu,
    Sharon and Quek, Francis},
  year={2014},
}

@inproceedings{wozniak_thaddeus:_2014,
  title={{Thaddeus: a dual device interaction space for exploring
    information visualisation}},
  booktitle={Proceedings of the 8th {Nordic} {Conference} on {Human}-{
    Computer} {Interaction} {Fun}, {Fast}, {Foundational} - {NordicCHI}
    '14},
  author={Wozniak, Pawel and Lischke, Lars and Schmidt, Benjamin and Zhao,
    Shengdong and Fjeld, Morten},
  year={2014},
}

@article{chung_savil:_2018,
  title={{SAViL: cross-display visual links for sensemaking in display
    ecologies}},
  journal={Personal and Ubiquitous Computing},
  author={Chung, Haeyong and North, Chris},
  year={2018},
}

@inproceedings{lucero_collaborative_2010,
  title={{Collaborative use of mobile phones for brainstorming}},
  booktitle={Proceedings of the 12th international conference on {Human}
    computer interaction with mobile devices and services - {MobileHCI}

```

```

    '10},
    author={Lucero, Andres and Keraenen, Jaakko and Korhonen, Hannu},
    year={2010},
}

@inproceedings{schreiner_connichiwa:_2015,
  title={{Connichiwa: A Framework for Cross-Device Web Applications}},
  booktitle={Proceedings of the 33rd {Annual} {ACM} {Conference} {Extended}
    {Abstracts} on {Human} {Factors} in {Computing} {Systems} - {CHI} {EA}
    '15},
  author={Schreiner, Mario and Raedle, Roman and Jetter, Hans-Christian and
    Reiterer, Harald},
  year={2015},
}

@inproceedings{langner_demonstrating_2018,
  title={{Demonstrating vistiles: visual data exploration using mobile
    devices}},
  booktitle={Proceedings of the 2018 {International} {Conference} on {
    Advanced} {Visual} {Interfaces} - {AVI} '18},
  author={Langner, Ricardo and Horak, Tom and Dachsel, Raimund},
  year={2018},
}

@inproceedings{strohmeier_displaypointers:_2015,
  title={{DisplayPointers: seamless cross-device interactions}},
  booktitle={Proceedings of the 12th {International} {Conference} on {
    Advances} in {Computer} {Entertainment} {Technology} - {ACE} '15},
  author={Strohmeier, Paul},
  year={2015},
}

@inproceedings{ohta_dynamically_2008,
  title={{Dynamically reconfigurable multi-display environment for CG
    contents}},
  booktitle={Proceedings of the 2008 {International} {Conference} in {
    Advances} on {Computer} {Entertainment} {Technology} - {ACE} '08},
  author={Ohta, Takashi},
  year={2008},
}

@inproceedings{hardy_touch_2008,
  title={{Touch \& interact: touch-based interaction of mobile phones with
    displays}},
  booktitle={Proceedings of the 10th international conference on {Human}
    computer interaction with mobile devices and services - {MobileHCI}
    '08},
  author={Hardy, Robert and Rukzio, Enrico},
  year={2008},
}

@inproceedings{fei_peripheral_2013,
  title={{Peripheral array of tangible NFC tags: positioning portals for

```

```

    embodied trans-surface interaction}},
  booktitle={Proceedings of the 2013 {ACM} international conference on {
    Interactive} tabletops and surfaces - {ITS} '13},
  author={Fei, Shenfeng and Webb, Andrew M. and Kerne, Andruid and Qu, Yin
    and Jain, Ajit},
  year={2013},
}

@misc{gapminder2019,
  title={{Gapminder}},
  url={https://www.gapminder.org/},
}

@article{tobiasz2009lark,
  title={{Lark: Coordinating co-located collaboration with information
    visualization}},
  author={Tobiasz, Matthew and Isenberg, Petra and Carpendale, Sheelagh},
  journal={IEEE transactions on visualization and computer graphics},
  year={2009},
}

@article{sadana_designing_2016,
  title={{Designing Multiple Coordinated Visualizations for Tablets}},
  journal={Computer Graphics Forum},
  author={Sadana, R. and Stasko, J.},
  year={2016},
}

@inproceedings{goel2014surfacelink,
  title={{SurfaceLink: using inertial and acoustic sensing to enable multi-
    device interaction on a surface}},
  author={Goel, Mayank and Lee, Brendan and Islam Aumi, Md Tanvir and Patel,
    Shwetak and Borriello, Gaetano and Hibino, Stacie and Begole, Bo},
  booktitle={Proceedings of the SIGCHI Conference on Human Factors in
    Computing Systems},
  year={2014},
}

@inproceedings{hinckley2004stitching,
  title={{Stitching: pen gestures that span multiple displays}},
  author={Hinckley, Ken and Ramos, Gonzalo and Guimbretiere, Francois and
    Baudisch, Patrick and Smith, Marc},
  booktitle={Proceedings of the working conference on Advanced visual
    interfaces},
  year={2004},
}

@inproceedings{tandler2001connectables,
  title={{Connectables: dynamic coupling of displays for the flexible
    creation of shared workspaces}},
  author={Tandler, Peter and Prante, Thorsten and M{"u}ller-Tomfelde,
    Christian and Streitz, Norbert and Steinmetz, Ralf},
  booktitle={Proceedings of the 14th annual ACM symposium on User interface

```

```

        software and technology},
        year={2001},
    }

@inproceedings{grubert2017towards,
    title={{Towards ad hoc mobile multi-display environments on commodity
        mobile devices}},
    author={Grubert, Jens and Kr{\\"a}nz, Matthias},
    booktitle={2017 IEEE Virtual Reality (VR)},
    year={2017},
}

@inproceedings{rekimoto1999augmented,
    title={{Augmented surfaces: a spatially continuous work space for hybrid
        computing environments}},
    author={Rekimoto, Jun and Saitoh, Masanori},
    booktitle={Proceedings of the SIGCHI conference on Human Factors in
        Computing Systems},
    year={1999},
}

@article{spindler2014tangible,
    title={{Tangible displays for the masses: spatial interaction with
        handheld displays by using consumer depth cameras}},
    author={Spindler, Martin and B{\\"u}schel, Wolfgang and Winkler, Charlotte
        and Dachsel, Raimund},
    journal={Personal and Ubiquitous Computing},
    year={2014},
}

@inproceedings{grubert2017headphones,
    title={{Headphones: Ad hoc mobile multi-display environments through head
        tracking}},
    author={Grubert, Jens and Kranz, Matthias},
    booktitle={Proceedings of the 2017 CHI Conference on Human Factors in
        Computing Systems},
    year={2017},
}

@inproceedings{hincapie2015gyrowand,
    title={{GyroWand: IMU-based raycasting for augmented reality head-mounted
        displays}},
    author={Hincapie{\'}e-Ramos, Juan David and Ozacar, Kasim and Irani,
        Pourang P and Kitamura, Yoshifumi},
    booktitle={Proceedings of the 3rd ACM Symposium on Spatial User
        Interaction},
    year={2015},
}

```

---

# Bibliography

- [1] APA Style. <https://blog.apastyle.org/apastyle/2012/03/title-case-and-sentence-case-capitalization-in-apa-style.html>.
- [2] Autobib. <https://github.com/jdumas/autobib/blob/master/README.md>.
- [3] biblib documentation. <https://wgserve.de/biblib/intro.html>.
- [4] bibtex-check project site. <https://code.google.com/archive/p/bibtex-check/>.
- [5] bibtex-parser. <https://www.npmjs.com/package/bibtex-parser>.
- [6] *CrossTEX: A modern bibliography management tool*.
- [7] Express. <https://expressjs.com/>.
- [8] JabRef Documentation. <https://docs.jabref.org/>.
- [9] Mendeley. <https://www.mendeley.com/reference-management/mendeley-desktop>.
- [10] React. <https://reactjs.org/>.
- [11] RefWorks Documentation. [https://knowledge.exlibrisgroup.com/RefWorks/Product\\_Documentation/RefWorks\\_User\\_Guide](https://knowledge.exlibrisgroup.com/RefWorks/Product_Documentation/RefWorks_User_Guide).
- [12] Zotero Documentation. <https://www.zotero.org/support/>.
- [13] D. Datta. *LATEX in 24 Hours*. Springer, 2017.
- [14] J. Fenn. Managing Citations and Your Bibliography with BibTeX. *The PracTeX Journal*, 2006, No. 4, 2006.
- [15] N. Markey. Tame the BeaST. [https://texdoc.net/texmf-dist/doc/bibtex/tamethebeast/ttb\\_en.pdf](https://texdoc.net/texmf-dist/doc/bibtex/tamethebeast/ttb_en.pdf).