Institute of Architecture of Application Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Bachelorarbeit

Who Cites Whom and Who Cites What in AI Planning

Lukas Steinegger

Course of Study:	Data Science
Examiner:	Prof. Dr. Marco Aiello
Supervisor:	Dr. Ilche Georgievski
Commenced:	April 12, 2021
Completed:	October 11, 2021

Abstract

The field of AI planning is important for many individuals and institutions, from researchers to industries. It has many application domains ranging from small drones to planning flight schedules of airports. In this work, we want to take a closer look at the academic side of AI planning by analysing publications from two major AI conferences, the ICAPS and the IJCAI conference. In particular, we want to tackle the question of how strong the relationship between theoretical and practical papers is in the academic field. To this end, we formulate an annotation task and analyse the data with well-known text processing methods as well as network analysis methods. We found that the practical side of AI planning is underrepresented, yet it is well connected and growing. Thus, the practical side of AI planning has a good foundation in this field.

Kurzfassung

Das Feld der KI-Planung hat viele Einsatzbereiche; diese reichen von der Kontrolle kleiner Drohnen bis hin zur planung von Flugplänen an Flughäfen. In dieser Arbeit wollen wir dieses Feld in der wissenschaftlichen Literatur genauer untersuchen, dazu untersuchen wir die Veröffentlichungen von zwei etablierten KI-Konferenzen, der ICAPS und der IJCAI Konferenz. Vor allem interessiert uns die Frage, wie stark das Verhältnis von der theoretischen Seite zur praktischen Seite innerhalb der wissenschaftlichen Literatur ist. Dazu analysieren wir die Daten mit Hilfe bekannten Textverarbeitungsmethoden sowie mit Netzwerkanalysemethoden, des Weiteren haben wir die Artikel manuell annotiert. Unsere Analysen zeigen, dass die praktische Seite zwar unterrepräsentiert ist, aber diese jedoch gut vernetzt und am Wachsen ist.

Contents

List of Figures

List of Tables

Acronyms

AAAI Association for the Advancement of Artificial Intelligence. 15

AI Artificial Intelligence. 13

BOW Bag-of-Words. 19

DBLP Digital Bibliography & Library Project. 17

GA Genetic Algorithms. 15

GPU Graphics Processing Unit. 18

ICAPS International Conference on Automated Planning and Scheduling. 7

IJCAI International Joint Conference on Artificial Intelligence. 7

IPC International Planning Competition. 18

IR information retrieval. 20

MAV Micro Aerial Vehicle. 13

MBD medical big data. 15

POS Part of Speech. 19

UAV Unmanned Aerial Vehicle. 19

1 Introduction

Artificial Intelligence (AI) planning is a subfield of the growing field of AI [NTX+16]. There are a lot of possible domains where AI planning can be adopted [GNT04]. There are multiple reasons to deal with AI planning, from designing automated machines to affordable planning tools free for use by any professionals. For example, it can be used in Micro Aerial Vehicles (MAVs) for surveillance missions, as demonstrated by Bernardini et al. [BFL14]. In the research of Geißer et al. [GPT+20], the authors dived into flight planning under uncertainty using AI planning.

The goal of AI planning is to obtain a feasible plan, which is not necessarily an optimal solution. In a very simple sense, AI planning is just a search of trajectories towards a set of goal states (nodes) in a graph. Since there is a range of relaxations, like an infinite set of states or a non-deterministic system, this simple view is not always applicable [GNT04].

(aipsftrw00), for instance, showed in 2000 that there are applied AI planning domains by providing a collection of examples. In this work, we want to find out how well the practical side of AI planning is connected to the theoretical side as this can help researchers to find research topics or to give them a different view on their topic. As for the industry, AI planning can give the inspiration to try academically founded solutions for industrial problems.

Our goal for this work is to elaborate on the relationship between theory and practice in the academic literature of AI Planning. We assume that we can use the references and topics discussed in the publications to gain insight into the relationship. We aim to solve this problem with the help of information retrieval and network analysis methods. Due to the scope, we concentrate on two major conferences, the ICAPS and the IJCAI conferences.

Our work is structured as follows; we start with an overview of past research on related topics. After that, we dive into our used approach, explain the annotations, and explain key components used in the analysis. We also go into details of how we collect the publications and how we preprocess them. For our analysis, besides relevant statistics, we created word clouds based with regard to the labels, a bibliographic coupling network, and a co-citation network. Then we show our findings and discuss them afterwards. We finish with a conclusion, and we take a look at some possible shortcomings of our work.

2 Related Work

Fitzpatrick et al. [FHM+18] tackled a similar question as we do. In their work, they analysed the relationship of theory and empiry in the field of ecology and evolutionary biology. To this end, they looked at seven hypotheses of citation patterns in their citation network. Contrary to a previously published survey, the authors found a stronger connection than expected, but this connection could still be improved.

Dawson et al. [DGSJ14] used a co-author network and a citation network to gain an overview of the current state of the learning analytic field using articles from two major conferences. They discovered that currently, a fragmentation in the disciplines seems to emerge.

In the citation analysis of Liao et al. [LTL+18], they looked at the contributions of authors, journals, and countries in the field of medical big data (MBD). They created a citation network using entries from the Web of Science database and developed a co-citation network and co-author networks. Worth mentioning is the creation of a journal co-author and an institute co-author network. With their analysis, they determined the most important publications, authors, countries, and journals in this field. Besides, they found that the topic is shifting from disease-centred research to patient-centred research.

Fang et al. [FJZ+21] made a bibliometric analysis in the field of AI using the Web of Science database. They looked into the most impactful countries and institutions and found that the most important contributions come from China and the USA. While also conducting a topic analysis, they found that algorithms have a long life span. And finally, they identified five noteworthy topics, namely *Perception intelligence*, *Human mind simulated intelligence*, *Classical model based machine learning*, *Bio-inspired intelligence*, and *Big-data based intelligence*.

In contrary to the so far mentioned work, Dao et al. [DAM17] did not use a citation network to make a bibliometric analysis. For their analysis, they examined the development of publications associated with Genetic Algorithms (GA) from the Elsevier Scopus. They explored the influence of institutes, journals, countries, and authors and found that the number of publications has increased rapidly since 1992. They identified the two top journals: *Expert Systems with Application* and *Bioinformatics*. As well as the top contributing countries: *China* and the *United States*, followed by *India*, *Japan* and the *United Kingdom*. Respectively the top institution is the *Tsinghua University* followed by the *Shanghai Jiaotong University*.

In a recent work of Sharma [Sha21], the author conducted a citation analysis using articles published in the IJCAI, ICAPS, and Association for the Advancement of Artificial Intelligence (AAAI) conferences and additionally including metadata from Google scholar. In total, they considered 100 highly cited articles from 2002 to 2020. Sharma [Sha21] focused on grasping the current state of the research field and further determining gaps and emerging topics. They used three well-established journals as a source for their analysis. Further, the author focused on a manual evaluation of the articles, which results in a deep understanding of the selected articles. However, the thesis of Sharma [Sha21] has some possible limitations regarding the sample size. The author collected 100 articles from the past 19 years, which averages to 5.3 articles per year. This can be quite limiting for an in-depth analysis. In contrast, the related work mentioned above in Section 2 are considering datasets containing from around 988 [LTL+18] (27 years) up to 124.799 [DAM17] (43 years) papers. Our work includes 1442 publications over 18 years.

3 Methodology

In this chapter, we dive into the methodology of our work. We start by explaining how we retrieve the data set. Then we demonstrate how to annotate each paper and provide a list of the used labels with a corresponding description. In Section 3.3, we go into details of how to preprocess the collected papers. There we show some key components of text processing we use in our further analysis. In the last section, Section 3.4, we look into the network analysis methods we further used for our analysis.

3.1 Corpus

The collected data for our analysis are from two major AI conferences: The ICAPS[1] and the IJCAI[2] conference. The majority (~1000 papers) of the total 1442 papers are from the ICAPS conference. The ICAPS conference has a strong focus on AI planning. Since the IJCAI conference does not have a sole focus on AI planning, we used the retrieved ICAPS papers as an entry point for the IJCAI conference. We determine the most important terms from the ICAPS conference papers and use those to search for relevant IJCAI papers, using the $tf - idf$ score as described in Section 3.3.1. We further only consider papers in the period from 2003 to 2020, since the retrieved pdf files from before 2002 have an increasing number of scanned publications containing only images of text, which we can not automatically analyse without further preprocessing.

To collect additional metadata like the authors and the title, we use the Digital Bibliography & Library Project (DBLP)[3] database. We also parse the downloaded pdf files to complete eventually missing entries on a few files, where we looked up the information manually. One key point is to determine the references from each publication and then detect all references to the same work. This is accomplished by using string similarity metrics, which we will discuss later in this Chapter 3.3.3.

3.2 Labels

We now label our data to explore the relationship between theoretical and practical papers in AI planning. Ideally, we want a clear classification between both extremes. However, such a direct and unambiguous separation is hard to accomplish. Therefore, we use an approach where each paper gets annotated with one or more labels. We annotate the papers manually and use an initial set of

[1]https://www.icaps-conference.org/
[2]https://www.ijcai.org/
[3]https://dblp.org/

five labels (Foundations, Applied, Improvements, Experimental, Benchmark), which we determined after reading a few samples from our collection. To categorize each paper into our given spectrum of Theory and Practice, we add a new label if an important aspect could not be covered by an already existing one. To give an overview of the utilized labels, we include the following list with a corresponding definition or examples.

1. Foundation:
   Work containing one or multiple proofs; for example, proof of the complexity of a problem or algorithm.
2. Analysis:
   Work that looks into a single approach or multiple approaches from a theoretical point of view. For instance, a comparison of the stability of an algorithm under some specific assumptions is a single approach. In turn, work that analyses different algorithms over a reference problem would be considered a multiple approach analysis.
3. Study:
   Every paper claiming to be either an empirical study, a user study, or a survey of any kind.
4. Recap:
   Short summary papers from an existing paper. Reports and summaries on past research.
5. Optimization:
   A paper that goes into detail on how to optimize an algorithm, such as an implementation of an algorithm with an optimized data structure. A concrete example would be an implementation of the A* algorithm using CUDA[4], so that it can run more efficiently using Graphics Processing Units (GPUs).
6. Problem proposition:
   The paper has a focus on elaborating a specific problem/domain. It has a detailed description of the problem, like in the label *improvement*. Often, the problem is proposed as a new benchmarking problem, and in general, no solving algorithm is given. The problem is analysed regarding a property (e.g. a high branching factor in the search space).
7. Improvement:
   Papers that describe a new, improved or existing algorithm and often also describe a new, changed or existing domain/problem, where the introduced algorithm is evaluated. In contrary to *problem*, the focus lies on the solving algorithm and not on the problem description.
8. Experimental:
   Work that evaluates their approach with their own evaluation criterion,often with the goal to show that their approach does work.
9. Benchmark:
   The evaluation of the proposed approach is done by a well-known problem/domain, like for example an evaluation on one of the past International Planning Competition (IPC) problems. The goal is to achieve a comparison to other state-of-the-art algorithms.

[4]A toolkit from NVIDIA to program their hardware. For more information visit https://developer.nvidia.com/cuda-toolkit

10. Simulation:
    An approach is evaluated/tested on computationally more expensive tests. The simulations have the goal to imitate a realistic environment. Besides simulations, evaluations using complex games like chess, go, and video games, are also counted as simulations. Simple games like Pong do not count.

11. Realistic datasets:
    In his case, the evaluation is done by using a real or realistic data set. This category implies that the authors want to tackle a problem directly inspired by a non-theoretical problem.

12. Software:
    The researchers are not focused on solving a problem in a certain domain, but publishing a tool to solve a wider range of domains.

13. Lab:
    The research is tested on a physical robot (e.g. a Unmanned Aerial Vehicle (UAV), a drone or a robot-soccer bot), but in a controlled environment (e.g. the lab).

14. Applied:
    At the time of the publication, the approach will be in use in the foreseeable future, or it is already in use. An example could be a mars rover that uses a planning algorithm to schedule its task, which will be in use by the time of the publication.

3.3 Working with Documents

This section covers some well-known approaches for text processing. Firstly, we preprocess the text. Secondly, we go through document representations using tf and $tf-idf$ values. And finally, we look into string similarity metrics.

Before we can do any text analysis, we need to preprocess our document collection. To this end, we first need to tokenize each document into a smaller set of terms by parsing the text. Tokenizing the text can be as simple as just splitting the document at any whitespace and special character. Most of the time, this is sufficient for English data. In the next step, the tokens are further processed by applying specific rules (Stemming) or by reducing a word to its root with their Part of Speech (POS) information (Lemmatisation) [AHN19].

With our preprocessed collection, we represent the documents using a Bag-of-Words (BOW) model and a $tfi-df$ model, depending on the analysis task.

3.3.1 tf-idf

To determine important terms in a given document of our collection, we use the $tf-idf$ weights. To calculate this, we first of all need the term frequency $tf_{t,d}$ for a given term t and a given document d. The $tf_{t,d}$ value is determined by simply counting the occurrences of the term t in the document d. This is also known as a bag-of-word representation.

There are multiple variations of the $tf_{t,d}$ value, please see [RW08] for more information on this. We use the plain tf value instead of a more complicated approach, like taking the logarithm.

The next step is to calculate the inverse document frequency idf_t for a given term t. This is done by calculating the document frequency df_t for a given term t, which is the number of documents that contain the term t. We then take the logarithm from N divided through df_t, where N is the total number of documents in our collection. Thus, the equation for the idf_t score is 3.1.

$$idf_t = log\left(\frac{N}{df_t}\right) \tag{3.1}$$

Now we need to combine the $tf_{t,d}$ with the idf_t value. For that, we multiply these values with one another. The resulting equation is shown in 3.2. We calculate the $tf - idf_{t,d}$ for every word in our collection.

$$tfidf_{t,d} = tf_{t,d} \cdot log\left(\frac{N}{df_t}\right) \tag{3.2}$$

Very frequent words like the so-called stop-words (e.g. the, good, a, ...) are given a lower value by the idf_t part. Terms with a low df_t value are increasing the value of $\frac{N}{df_t}$ and thus also the value of $log\left(\frac{N}{df_t}\right)$. A word that only occurs in a few documents is probably more informative for a document than a word that appears in every document. So to speak, the idf_t value describes the informativeness of a term t in a collection. Usually, the collection of documents is quite large and therefore the $idf_{t,d}$ value is dampened with a logarithm.

The $tf_{t,d}$ also takes into account the document the term occurs in. Here, a word that only occurs a few times is probably not as interesting as a word that occurs multiple times.

The $tf - idf_{t,d}$ score is still widely used in information retrieval (IR) [MRS08].

3.3.2 Pointwise Mutual Information

Pointwise mutual information is a part of the mutual information algorithm. We use it to find common words for a given label. The pmi formula is given in Equation 3.3 [AK12]

$$pmi(X,Y) = log\left(\frac{P(X,Y)}{P(X) \cdot P(Y)}\right) \tag{3.3}$$

Two random variables X and Y are independent when $P(X,Y) = P(X) \cdot P(Y)$ holds true. From that follows that $\frac{P(X,Y)}{P(X) \cdot P(Y)} = 1$ and so the pmi of two independent variables is $log\,(1) = 0$. For any other case the value is non-zero. The pmi value is not symmetric and is not limited by any value [Hen13].

To calculate one pmi value, we determine the probability of the occurrence of a specific label $P(X)$, the normalized document frequency of a given term $P(Y)$ and the joined probability of them occurring at the same time. Negative values show a negative correlation and vice versa. A positive value shows a positive correlation. The zero value shows that the two events are independent [AK12; Hen13; MRS08].

	ε	f	a	s	t
ε	0	1	2	3	4
c	1	1	2	3	4
a	2	2	1	2	3
t	3	3	2	2	2

Table 3.1: An example of the table used to calculate the Levenshtein distance on the words "fast" and "cat"

3.3.3 String similarity

A part for our analysis is to find similar strings. We use this to determine whether papers reference the same work, even when they use different citation styles. If two references are similar enough (above a certain threshold) we regard them as equal. For this, we use a combination of two string distances, the Levenshtein distance and the Jaro Winkler Distance. In the implementation we use the library jellyfish[5] for Python. We first look into the Levenshtein distance and then into the Jaro-Winkler distance.

Levenshtein Distance

The Levenshtein distance is calculated based on the edit distance. The available operations are: insert, delete and replace. One string can be transformed into the other string by applying those operations in a particular order. The Levenshtein distance specifies the minimal amount of operations to get from the source string s to the target string t [ZS20].

The implementation is usually done by a dynamic programming approach using a table. An example of such a table is shown in Table 3.1. The table is initialized by numbering the first row and column. The entries with ϵ are marked with a grey background in Table 3.1. From there on, we can iteratively apply the rules described in the equation 3.4 [BAA+17; ZS20].

$$Levenshtein(current\ cell) = min \begin{cases} \begin{cases} diagonal & if\ letters\ are\ the\ same \\ diagonal + 1 & otherwise \end{cases} \\ above + 1 \\ left + 1 \end{cases} \tag{3.4}$$

The final value in the lower right cell is the Levenshtein distance.

[5] https://github.com/jamesturk/jellyfish

Jaro-Winkler Distance

Another popular string similarity metric is the Jaro-Winkler distance. In contrast to the Levenshtein distance, it is not based on the edit distance. The Jaro-Winkler distance is a variant of the Jaro distance. We first discuss the Jaro distance and then the Jaro-Winkler variant.

For the Jaro similarity we need two strings, a source string $s = a_i, \ldots, a_k$ and a target string $t = b_1, \ldots, b_l$. In the first step we determine the length of both strings $|s| = k$ and $|t| = l$. Then we need to calculate the number of characters m both strings have in common. If one character matches a single other character from the second string, we simply count those matches in a certain range, which is defined as $H = \frac{max(|s|,|t|)}{2} - 1$ to get m. A character a_i is only matched with a character b_j when $i - H \leq j \leq i + H$ holds. The resulting common matching strings are $s' = a'_1, \ldots, a'_{k'}$ and $t' = b'_1, \ldots, b'_{l'}$ [Win90].

We also need to know the number of transpositions T. We calculate this by counting the matches of s' and t' characters at the i-th position.

With these three values, we then calculate the Jaro-similarity using the formula 3.5 [Win90].

As there exist multiple different variations, we present the one from [Win90].

$$Jaro(s,t) = \begin{cases} 0, & if\ m = 0 \\ \frac{1}{3}\left(\frac{m}{|s|} + \frac{m}{|t|} + \frac{m-T}{m}\right), & otherwise \end{cases} \tag{3.5}$$

The Jaro-Winkler modification tries to emphasize common prefixes of the given strings. The modification is shown in Equation 3.6. P is the length of the longest common prefix of s and t. And $P' = max\ (P, 4)$ is a lower threshold for the considered prefix length [Win90].

$$Jaro - Winkler(s,t) = Jaro(s,t) + \frac{P'}{10} \cdot (1 - jaro(s,t)) \tag{3.6}$$

3.4 Networks

3.4.1 Citation Networks

One central part of our analysis is to build a citation network. For that, we interpret the papers in our collection as nodes of our citation network as well as the referenced papers from those papers. The edges in the citation network are the references from one paper to another work. Citation networks are acyclic graphs since an academic work cannot reference a paper from the future or from the same time when it gets published. Further, citation networks are directed graphs. Edges point from the referencing paper to the referenced paper [New10].

As an example, in figure 3.1 we can see multiple nodes representing a citation network. The nodes A to F, two grey nodes, one blue and one red node represent papers. The alphabetically enumerated nodes represent citing papers, and the coloured nodes represent cited papers. The coloured nodes are later used in Section 3.4.2.

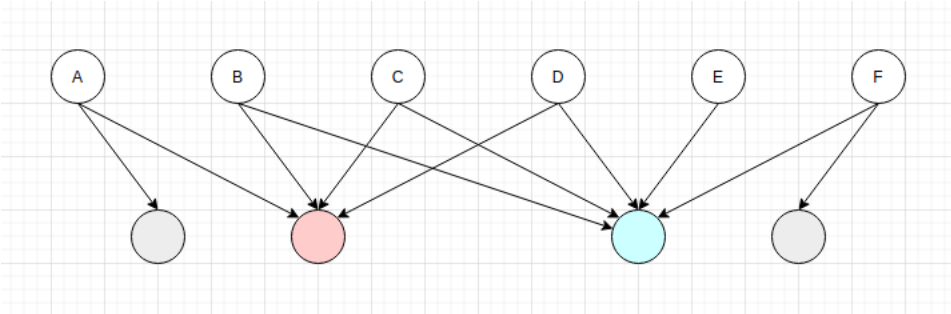


Figure 3.1: Example of a small citation network. Nodes represent papers, and edges show that one paper referenced another paper.

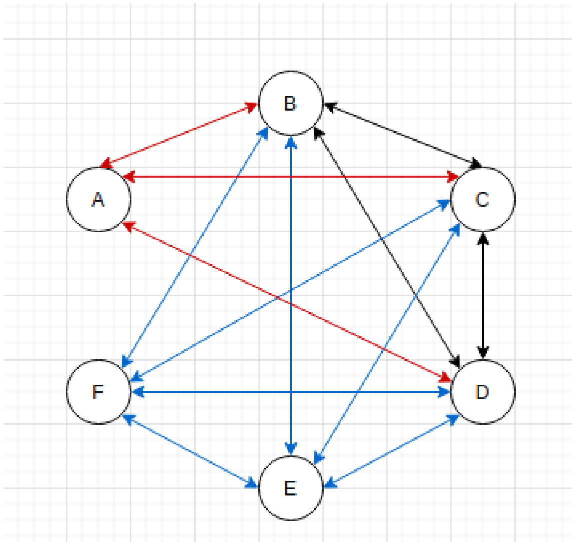


Figure 3.2: The corresponding bibliographic coupling network of 3.1. Only the nodes labelled from *A* to *F* are considered for the bibliographic coupling network. The red edges come from the nodes referencing the red node in the citation network. The blue edges are created analogously. The black edges come from the nodes referencing both, blue and red nodes.

3.4.2 Bibliographic Coupling

To create a bibliographic coupling network, we need a directed graph; in our instance, a citation network. Two papers are bibliographically coupled if both cite the same reference. In contrast to the citation network, this graph is undirected, but it could be weighted by the number of same references two nodes have. For our analysis, we use a non-weighted graph. There exists also the co-citation network, where two papers are related if both are cited from the same referenced work [New10].

Figure 3.2 shows the bibliographic coupling network corresponding to the network in Figure 3.1.

3.4.3 Co-author Network

Another network we want to take a look at is the so-called co-author network. To build this network, we need the authors and the publications they worked on. One node in this graph represents a single author. An edge is drawn between two authors if they worked together on a paper [New10].

4 Results

4.1 Analysis

The ICAPS conference has a total of 989 publications, and the IJCAI has a contribution of 453 papers. Combined there are 1442 documents. In 4.1 we visualize the number of papers published in each conference over the years, from 2003 up to 2020. It immediately stands out that the IJCAI published papers relevant to our collection every two years because it was held biennially. Since 2016 it is an annual conference [21]. Additionally, for both conferences, we can see a steady increase in the number of publications over time.

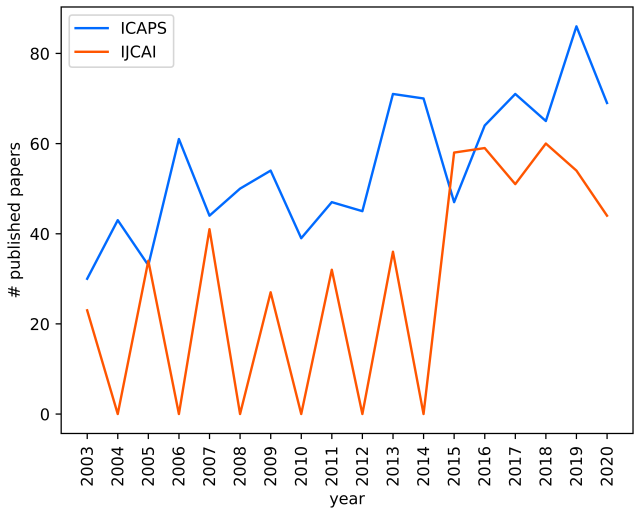


Figure 4.1: Plot with the number publication of AI-Planning paper per conference (ICAPS and IJCAI) over the period of 2003 to 2020. The x-axis shows the year and the y-axis shows the number of publications per conference.

Figure 4.2 shows a bar plot with the count of the assigned labels. We can see that the majority of the collected papers, namely 1253 out of 1442, was assigned the label *improvement*. After that, the second most assigned label is *benchmark* (674). With nearly the same count, the labels *foundations* (425) and *experimental* (421) were assigned to a third of the papers. The remaining labels ranging from *recap* to *real*, whose count sums up to 421, were assigned to 341 papers. The x-axis shows the number of publications per conference. The y-axis shows the corresponding labels.

In Figure 4.3, we look into the number of labels assigned per year. Figure 4.2 indicates that the label *improvement* was always assigned to the majority of papers over the years. The labels *benchmark* and *experimental* are roughly increasing with the number of publications. Furthermore, the label

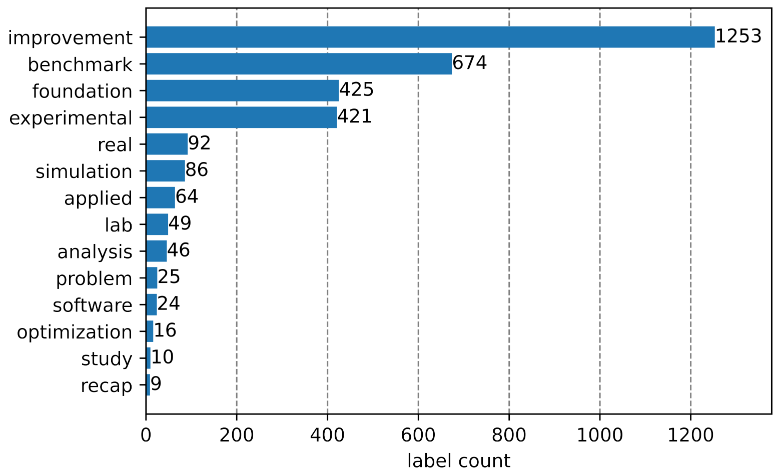


Figure 4.2: Bar-plot with the number of assignments per label. The x-axis shows the number of assignments. The y-axis shows the bars for each label.

foundation has a peak in 2017 and decreases from then on. Since 2013, the appearance of the labels *real*, *applied*, *lab*, and *simulation* has increased. As the diagram is a bit cluttered on the bottom, we include the corresponding tables in the Appendix A.1 and A.2.

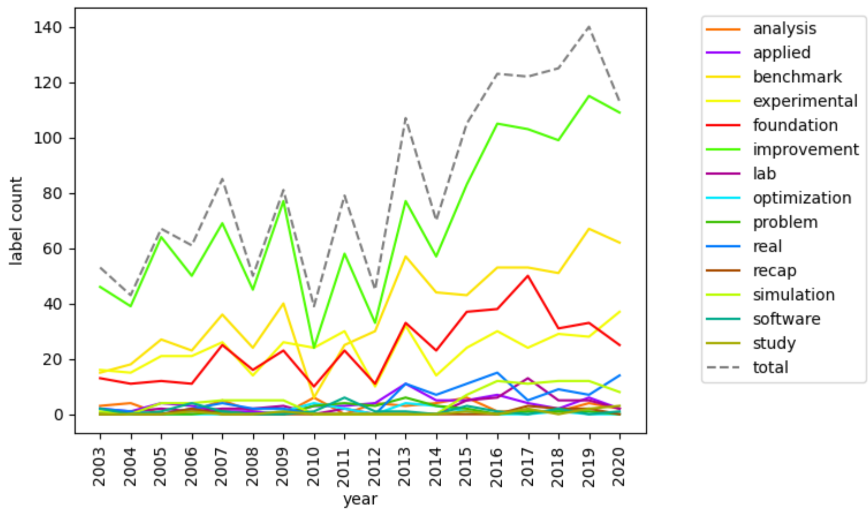


Figure 4.3: Line-plot with the assigned labels per year, from 2003 to 2020. The x-axis shows the year, and the y-axis indicates the number of assigned labels per year for all papers of that year. The grey line indicates the total amount of published papers as a reference for the assigned labels.

This section shows multiple word clouds built with the help of the *pmi* value as described in Section 3.3.2. Each word cloud displays the words with the highest positive *pmi* value regarding a given label. Since we have 14 labels, we concentrate on the labels that were assigned fewer times. With the help of the word clouds, we can get an idea of the topics discussed in these categories. For the label *lab* in Cloud 4.4d, we can see that this label represent robotic tasks, e.g. *collided*, *haptic* and *humanoids*. Also, for the label *real* we have the words *fri*, *tue* and *wed* which seem to stand for Friday, Tuesday and Wednesday.

In cases where some words do seem to have missing letters, there are two possible reasons for that. For one, the words are split in the original document, for example due to a linebreak. Or they are incomplete due to a preprocessing step, like lemmatization or stemming.

In Histogram 4.5, we can observe that the majority of authors only contributed to between one and five publications. We can also see that there are a few authors who contributed frequently to both conferences. Further, we have one author who has published 47 papers during our considered period. So on average, this author has published 2.7 times per year.

We also build a bibliographic coupling network, which we can see in Figure 4.6. A node represents a paper, and an edge is drawn when two papers cite the same reference. The colour of the node represents the conference in which it was published: the red nodes are from the ICAPS conference, and the green nodes are from the IJCAI conference. The figure shows that there are two slightly connected subgraphs. These subgraphs seem to depend on the conferences, yet they are not strictly separating the papers by conference.

Further analysis on the same graph displayed by Figure 4.7 shows that the *applied* label seems to be assigned more often to the conference ICAPS than to the IJCAI conference. We can also confirm this by taking a closer look into our data set. There are 60 papers with the label *applied* from the ICAPS conference and only four papers from the IJCAI conference.

Another correlation between the label and another subgraph in the bibliographic coupling network could not be found.

In the last step, we also built a co-author graph as shown in Figure 4.8. Here the nodes are the authors, and the edges show when authors have published a paper with someone else. Since the edges represent papers, we can assign them a colour depending on a selected label to better visualize the relationships. We can see clear clusters of authors working together. On the edge, we can observe that there are small collaborations. In the middle, there are a lot more and mixed collaborations.

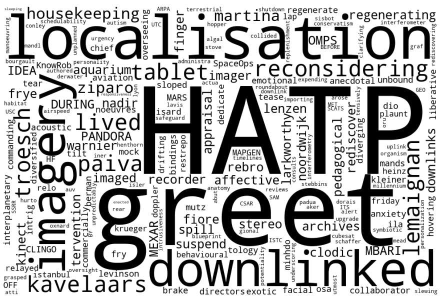
When we explore the collaborations in more detail in Figure 4.9, we see that the labels are assigned independently of the number of collaborations. Furthermore, we can detect red edges on the rim of the graph as well as in the middle. In all graphs, we can find multiple authors who have multiple collaborations within the same assigned label. This pattern can be found for every label.

The graphs shown in the Figures 4.6, 4.7, 4.8, and 4.9 were created with the tool Gephi[1].

4.2 Discussion

We now want to discuss the findings from our analysis. One of our first observations is an increasing amount of publications (4.1) alongside an increase in the number of assigned labels like *real*, *applied*, *simulation*, *lab*, and especially *foundation* (4.3). This is an indication that both fields, in

[1] https://gephi.org/



(a) applied

(b) analysis

(c) real

(d) lab

(e) simulation

(f) foundation

Figure 4.4: Word cloud for the labels 4.4a *applied*, 4.4b *analysis*, 4.4c *real*, 4.4d *lab*, 4.4e *simulation* and 4.4f *foundation*. The font size corresponds to the importance of the word determined by the *pmi* value.

particular the practical field, are increasing. On the other hand, in total, we do not have as many practical-oriented publications, as indicated by the labels *real*, *simulation*, and *applied* in Figure 4.2.

An important contribution from the word clouds is that the publications are covering topics like robotics (4.4d) and real-world-oriented datasets (4.4c).

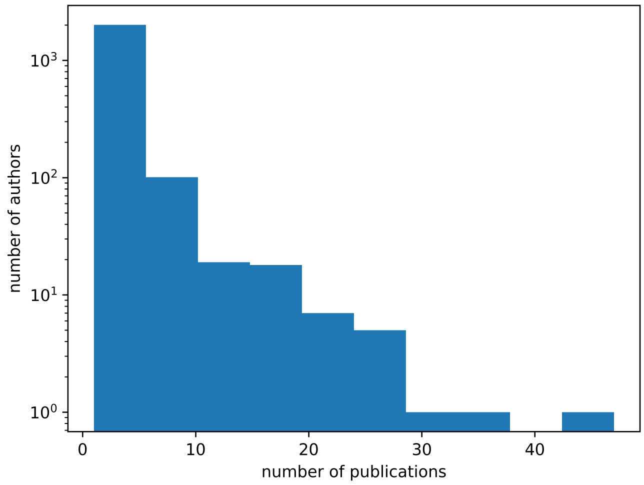


Figure 4.5: This figure shows a histogram of how many authors have contributed to publish n papers. Please note that the y-axis has a logarithmic scale. The x-axis shows the number of publications an author has. The y-axis shows how many authors have published n papers.

Except for the clear partition seen in Figure 4.6 and the weaker one in Figure 4.7, we did not find any further correlation between the subgraphs and our annotations. However, we can see that our annotated labels are represented in all these possible subtopics not covered by us. This indicates a homogenous distribution of the labelled topics and thus an interaction between the research.

From our last figures, the co-author networks, we can conclude that between the various research, researchers are engaged in different topics and are not only interested in a single subdomain. At least, this seems to be the case for researchers who contributed multiple times. Since we can see an increase in the practical topics in the last years (4.3), we can conclude that these cooperations have also been increasing in the last years. This confirms one result from [Sha21], which expected an increase in the topics like robotics. The same applies to the more theoretically oriented topics (e.g. *foundation* and *improvement*).

Another factor for the relationship between theoretical and practical research is that we have, so to speak, core contributors to the conferences (4.5). They push on collaborations and contribute to a diverse landscape. Thus, they provide a connection between theory and practice.

All in all, we can infer that there is a collaboration between theory and practice. Yet, it is not as strong as we expected in the beginning. Moreover, this relationship seems to get stronger over time, as it has grown in the past few years.

4.3 Limitations and Future Work

This work has some possible shortcomings.

Since we only considered two conferences and publications written in English, we have a possible bias in our results.

Figure 4.6: Graph of the bibliographic coupling network. Nodes are papers in our data set, and the edges show that two papers have cited the same reference. Red nodes are papers from the ICAPS conference, green nodes from the IJCAI conference.

Further, only one undergraduate student labelled the papers, whereas multiple annotators with more expertise in AI planning could have possibly reduced errors in the annotations. It would also be interesting to have a more granular labelling since there is an imbalance in the annotated labels as Figure 4.2 demonstrates.

Finally, due to the scope of this work, the number of papers considered could still bias the results, especially in contrast to past work such as Dao et Al. who analysed up to 124.799 publications [DAM17].

Future work could address these limitations as discussed above. Additionally, the scope of the work could include information about AI planning competitions like the IPC and explore whether or not the growth is continuing. Future studies on the current topic could also extract important contributions and authors in this field. To further our research, another approach could focus on practical research only and gain insight into the relationship between theoretical and practical papers via a directed look on this topic.

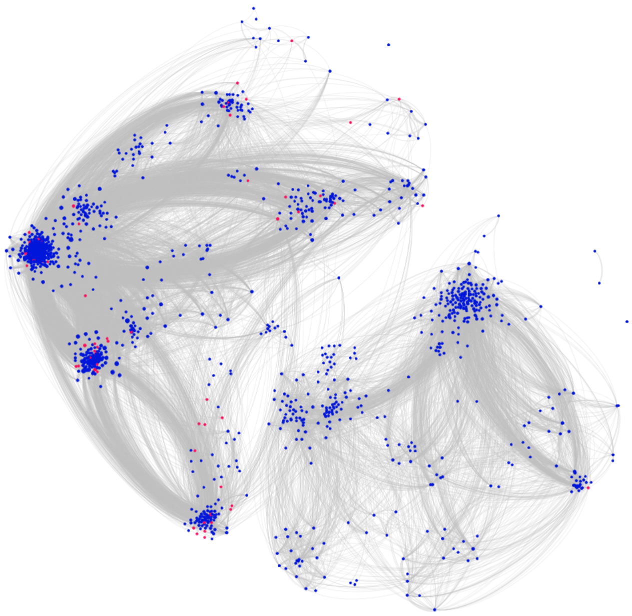


Figure 4.7: Graph of the bibliographic coupling network, with the same layout as in Figure 4.6. Red nodes have the label *applied* assigned, blue nodes do not.

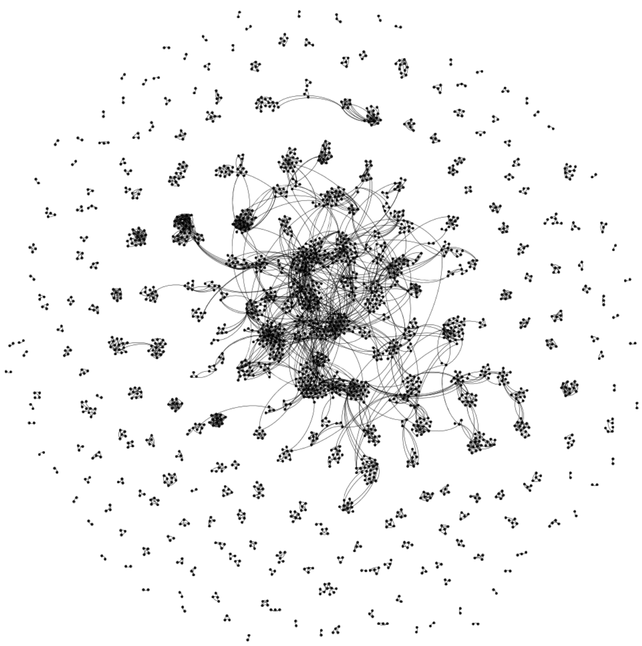
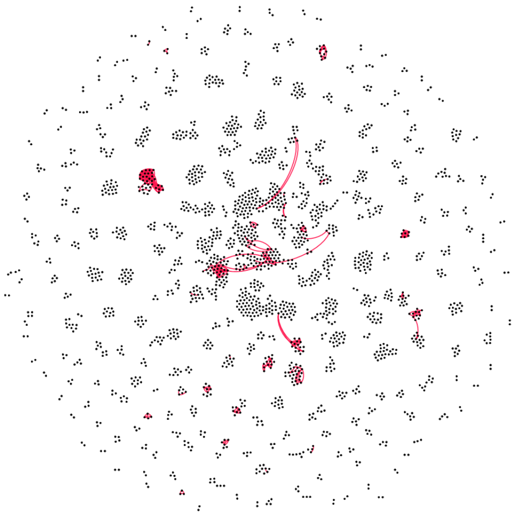
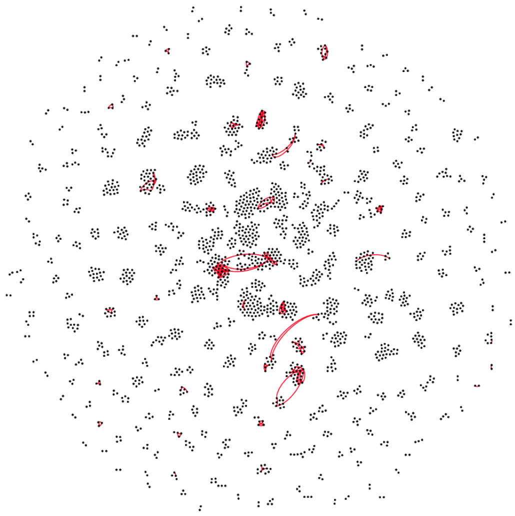


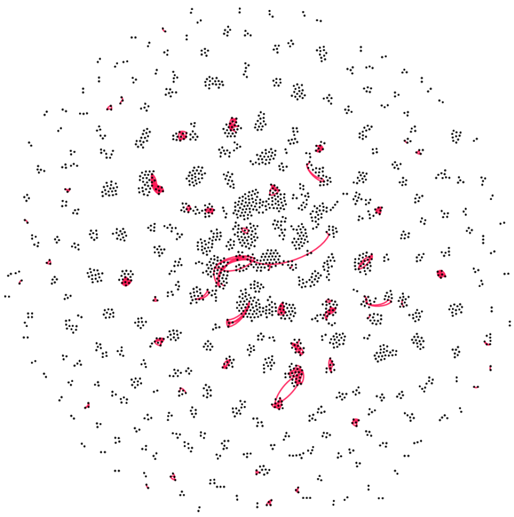
Figure 4.8: This graph shows the collaborations between authors. A node represents an author and an edge is drawn when two authors have published a paper together. Authors without any cooperation are not shown. This graph serves not only as an overview but also as a reference for the figures in 4.9.



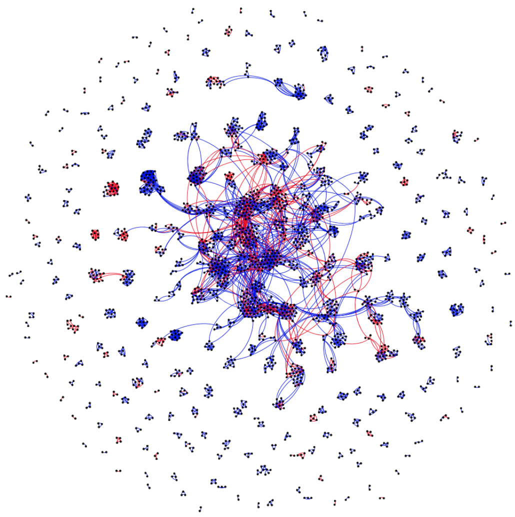
(a) applied



(b) lab



(c) simulation



(d) foundation

Figure 4.9: The same graph as 4.8. In the graphs 4.9a, 4.9b and 4.9c only the edges with the corresponding labels are shown to reduce visual clutter. The last graph 4.9d has all edges drawn. The red edges show if a collaboration was assigned with its specific label, the blue edges show the remaining collaborations.

5 Conclusion

In this work, we collected and analysed data based on well-known approaches and investigated the connection between theory and practice in academic literature in the research field of AI planning.

Taken together, our results suggest an interconnected relationship between theory and practice. However, the practical side seems to only have a small share in the academic literature.

Because of that, the relationship between theory and practice has still room to grow. Thus, it would be interesting to conduct further research in this area such as investigating the corresponding co-citation network, or, in a more general way, observing the found pattern of an interconnected relationship in the future.

Bibliography

[21] *IJCAI Conference Startpage*. 2021. URL: https://www.ijcai.org/ (visited on 09/06/2021) (cit. on p. 25).

[AHN19] M. Anandarajan, C. Hill, T. Nolan. "Text Preprocessing". In: *Practical Text Analytics: Maximizing the Value of Text Data*. Cham: Springer International Publishing, 2019. Chap. 4. ISBN: 978-3-319-95663-3. DOI: 10.1007/978-3-319-95663-3_4. URL: https://doi.org/10.1007/978-3-319-95663-3_4 (cit. on p. 19).

[AK12] S. Aji, R. Kaimal. "Document Summarization Using Positive Pointwise Mutual Information". In: *International Journal of Computer Science & Information Technology* 4 (May 2012). DOI: 10.5121/ijcsit.2012.4204 (cit. on p. 20).

[BAA+17] K. Balhaf, M. A. Alsmirat, M. Al-Ayyoub, Y. Jararweh, M. A. Shehab. "Accelerating Levenshtein and Damerau Edit Distance Algorithms Using GPU with Unified Memory". In: *2017 8TH INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION SYSTEMS (ICICS)*. International Conference on Information and Communication Systems. 8th International Conference on Information and Communication Systems (ICICS), Irbid, JORDAN, APR 04-06, 2017. Jordan Univ Sci & Technol; Atypon; gce. 2017, 7–11. ISBN: 978-1-5090-4243-2 (cit. on p. 21).

[BFL14] S. Bernardini, M. Fox, D. Long. "Planning the Behaviour of Low-Cost Quadcopters for Surveillance Missions". In: *Proceedings of the International Conference on Automated Planning and Scheduling* 24.1 (May 2014), pp. 445–453. URL: https://ojs.aaai.org/index.php/ICAPS/article/view/13670 (cit. on p. 13).

[DAM17] S. D. Dao, K. Abhary, R. Marian. "A bibliometric analysis of Genetic Algorithms throughout the history". In: *Computers & Industrial Engineering* 110 (2017), pp. 395–403. ISSN: 0360-8352. DOI: https://doi.org/10.1016/j.cie.2017.06.009. URL: https://www.sciencedirect.com/science/article/pii/S0360835217302619 (cit. on pp. 15, 16, 30).

[DGSJ14] S. Dawson, D. Gašević, G. Siemens, S. Joksimovic. "Current State and Future Trends: A Citation Network Analysis of the Learning Analytics Field". In: *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*. LAK '14. Indianapolis, Indiana, USA: Association for Computing Machinery, 2014, pp. 231–240. ISBN: 9781450326643. DOI: 10.1145/2567574.2567585. URL: https://doi.org/10.1145/2567574.2567585 (cit. on p. 15).

[FHM+18] C. L. Fitzpatrick, E. A. Hobson, T. C. Mendelson, R. L. Rodriguez, R. J. Safran, E. S. C. Scordato, M. R. Servedio, C. A. Stern, L. B. Symes, M. Kopp. "Theory Meets Empiry: A Citation Network Analysis". In: *BIOSCIENCE* 68.10 (OCT 2018), 805–812. ISSN: 0006-3568. DOI: {10.1093/biosci/biy083} (cit. on p. 15).

[FJZ+21] G. Fang, X. Jia, Z. Zhao, C.-C. Chen, F. Xu, Z. Geng, X. Song. “Bibliometric analysis on tendency and topics of artificial intelligence over last decade”. In: *Microsystem Technologies* 27 (Apr. 2021). DOI: 10.1007/s00542-019-04426-y (cit. on p. 15).

[GNT04] M. Ghallab, D. Nau, P. Traverso. *Automated Planning: Theory and Practice*. The Morgan Kaufmann Series in Artificial Intelligence. Elsevier Science, 2004. ISBN: 9781558608566. URL: https://books.google.de/books?id=eCj3cKC%5C_3ikC (cit. on p. 13).

[GPT+20] F. Geißer, G. Povéda, F. Trevizan, M. Bondouy, F. Teichteil-Königsbuch, S. Thiébaux. “Optimal and Heuristic Approaches for Constrained Flight Planning under Weather Uncertainty”. In: *Proceedings of the International Conference on Automated Planning and Scheduling* 30.1 (June 2020), pp. 384–393. URL: https://ojs.aaai.org/index.php/ICAPS/article/view/6684 (cit. on p. 13).

[Hen13] N. Henze. *Stochastik für Einsteiger. eine Einführung in die faszinierende Welt des Zufalls*. ger. 10., überarbeitete Auflage. Lehrbuch. Wiesbaden: Springer Spektrum, 2013, X, 402 Seiten. ISBN: 978-3-658-03077-3. DOI: https://doi.org/10.1007/978-3-658-03077-3 (cit. on p. 20).

[LTL+18] H. Liao, M. Tang, L. Luo, C. Li, F. Chiclana, X.-J. Zeng. “A Bibliometric Analysis and Visualization of Medical Big Data Research”. In: *Sustainability* 10.1 (2018). ISSN: 2071-1050. DOI: 10.3390/su10010166. URL: https://www.mdpi.com/2071-1050/10/1/166 (cit. on pp. 15, 16).

[MRS08] C. D. Manning, P. Raghavan, H. Schütze. *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press, 2008. ISBN: 978-0-521-86571-5. URL: http://nlp.stanford.edu/IR-book/information-retrieval-book.html (cit. on p. 20).

[New10] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010. ISBN: 978-0-19920665-0. DOI: 10.1093/ACPROF:OSO/9780199206650.001.0001. URL: https://doi.org/10.1093/ACPROF:OSO/9780199206650.001.0001 (cit. on pp. 22–24).

[NTX+16] J. Niu, W. Tang, F. Xu, X. Zhou, Y. Song. “Global Research on Artificial Intelligence from 1990-2014: Spatially-Explicit Bibliometric Analysis”. In: *ISPRS INTERNATIONAL JOURNAL OF GEO-INFORMATION* 5.5 (MAY 2016). ISSN: 2220-9964. DOI: {10.3390/ijgi5050066} (cit. on p. 13).

[RW08] T. Roelleke, J. Wang. “TF-IDF Uncovered: A Study of Theories and Probabilities”. In: *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’08. Singapore, Singapore: Association for Computing Machinery, 2008, pp. 435–442. ISBN: 9781605581644. DOI: 10.1145/1390334.1390409. URL: https://doi.org/10.1145/1390334.1390409 (cit. on p. 19).

[Sha21] B. Sharma. “Exploring the Scientific Landscape of AI Planning using Citation Network Analysis”. Bachelor’s Thesis. APL 2021 (cit. on pp. 15, 16, 29).

[Win90] W. Winkler. “String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage”. In: *Proceedings of the Section on Survey Research Methods* (Jan. 1990) (cit. on p. 22).

[ZS20] C. Zhao, S. Sahni. “Linear space string correction algorithm using the Damerau-Levenshtein distance”. In: *BMC BIOINFORMATICS* 21.1, SI (DEC 9 2020). 8th IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), Las Vegas, NV, OCT 18-20, 2018. ISSN: 1471-2105. DOI: {10.1186/s12859-019-3184-8} (cit. on p. 21).

A Tables

year	analysis	applied	benchmark	experimental	foundation	improvement
2003	3	2	15	16	13	46
2004	4	1	18	15	11	39
2005	0	4	27	21	12	64
2006	1	3	23	21	11	50
2007	5	1	36	26	25	69
2008	1	1	24	14	16	45
2009	0	0	40	26	23	77
2010	6	3	6	24	10	24
2011	0	3	25	30	23	58
2012	4	4	30	10	11	33
2013	3	11	57	32	33	77
2014	4	5	44	14	23	57
2015	6	5	43	24	37	83
2016	1	7	53	30	38	105
2017	1	4	53	24	50	103
2018	1	2	51	29	31	99
2019	4	6	67	28	33	115
2020	2	2	62	37	25	109

Table A.1: The table used to create Figure 4.3. It only shows the labels *analysis*, *applied*, *benchmark*, *experimental*, *foundation* and *improvement*, the remaining labels are shown in A.2.

year	optimization	problem	real	recap	lab	simulation	software	study
2003	0	0	2	0	0	1	2	0
2004	0	0	1	0	1	0	0	0
2005	0	0	0	0	2	4	1	0
2006	0	0	2	2	1	4	4	1
2007	0	1	4	0	2	5	1	0
2008	0	0	2	0	2	5	0	0
2009	1	0	2	0	3	5	0	1
2010	4	3	0	0	0	0	1	0
2011	2	4	0	0	2	0	6	0
2012	0	3	0	0	0	0	1	0
2013	4	6	11	0	0	0	1	0
2014	3	3	7	0	0	0	0	0
2015	1	2	11	0	5	7	3	1
2016	0	0	15	0	6	12	1	0
2017	0	1	5	3	13	11	0	2
2018	1	1	9	2	5	12	2	0
2019	0	1	7	2	5	12	0	2
2020	0	0	14	0	2	8	1	3

Table A.2: The table used to create Figure 4.3. It only shows the labels *optimization*, *problem*, *real*, *recap*, *lab*, *simulation*, *software* and *study*, the remaining labels are shown in A.1.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

place, date, signature