Aalborg Universitet

The Impact of CHIIR Publications

A Study of Eight Years of CHIIR

Gaede, Maria; Bogers, Toine; Hall, Mark Michael; Koolen, Marijn; Petras, Vivien; Larsen, Birger

The Impact of CHIIR: A Study of Eight Years of CHIIR Publications

Toine Bogers
Department of Computer Science
IT University of Copenhagen
Copenhagen, Denmark
tobo@itu.dk

Maria Gäde
Berlin School of Library and
Information Science
Humboldt-Universität zu Berlin
Berlin, Germany
maria.gaede@ibi.hu-berlin.de

Mark M Hall
School of Computing &
Communications
The Open University
Milton Keynes, United Kingdom

Marijn Koolen
DHLab, KNAW Humanities Cluster
Amsterdam, Netherlands
marijn.koolen@di.huc.knaw.nl

Vivien Petras
Berlin School of Library and
Information Science
Humboldt-Universität zu Berlin
Berlin, Germany
vivien.petras@ibi.hu-berlin.de

Birger Larsen
Department of Communication &
Psychology
Aalborg University Copenhagen
Copenhagen, Denmark
birger@ikp.aau.dk

ABSTRACT

Across all scientific fields, there is an increased focus on the impact of scientific research: what academic and societal benefits does it provide? This question has spurred the development of a variety of different approaches to impact assessment, each appropriate in different circumstances. In this paper, we study the academic impact of the CHIIR community through a comprehensive analysis of the work published in the 2016-2023 CHIIR conference series. We collect citation counts, citing documents, and altmetrics scores for all CHIIR publications to determine their academic impact across a variety of different attributes of the CHIIR publications. In addition, we analyze a subset of citation contexts in the papers that have cited CHIIR publications to analyze *how* they are being used and what that means for their potential impact. Finally, we attempt to predict which properties of CHIIR publications are most predictive of future impact.

CCS CONCEPTS

• **Information systems** → **Presentation of retrieval results**; **Test collections**; **Users and interactive retrieval**.

KEYWORDS

interactive information retrieval, sharing, re-use, citation analysis, impact

1 INTRODUCTION

Across all scientific fields, there is an increased focus on the impact of scientific research: which academic, health, and socio-economic benefits does it provide? This question has spurred the development of a variety of different approaches to impact assessment, each appropriate in different circumstances [13]. Estimating the *academic* impact of scientific research may be the most mature of these three types of impact assessment, as demonstrated by the abundance of different indicators of academic impact [36, 38]. However, which metrics are most appropriate can be dependent on the research field, as are the factors that influence whether a paper has impact [13]. In this paper, we study the academic impact of the CHIIR community through a comprehensive analysis of the work published in the 2016-2023 CHIIR conference series. We collect citation counts, citing documents, and altmetrics scores for all CHIIR publications to determine their academic impact in a variety of across different attributes of the CHIIR publications. In addition, we analyze a substantial subset of citation contexts in the papers that have cited CHIIR publications to analyze *how* they are being used and what that means for their potential impact. Finally, we attempt to predict which properties of CHIIR publications are most predictive of future impact. Here, we build upon earlier work on describing and prescribing the practices for documentation, re-use, and sharing of research resources in the CHIIR community to analyze how this may have influenced their impact.[1]

In this paper, we take the Conference on Human Information Interaction and Retrieval (CHIIR) conference as our object of study, which is one of the premier publication venues and communities for Information Interaction and Retrieval (II&R) research. We present the results of an analysis of eight years of CHIIR publications and address the following research questions.

The main question addressed in this work is: What is the impact of CHIIR work? This requires a definition of impact as well as an operationalization of it. Therefore, this paper addresses the following more specific research questions:

---

[1]The code to extract citation contexts and link them to CHIIR paper metadata is available at https://github.com/biirrr/chiir-2024-citations. We published the datasets on Zenodo: https://zenodo.org/records/10450344

RQ1 How do we operationalize impact in the CHIIR community?

RQ2 What is the reception of CHIIR work inside and outside the CHIIR community?

RQ3 In what context are CHIIR papers cited?

RQ4 Do empirical, resource and theoretical papers have different impacts?

RQ5 Does resource sharing and re-use have an influence on impact?

The remainder of this paper is structured as follows: In Section 2 we present related work on traditional analyses of citation purpose and context. The influence of alternative measures such as altmetrics and open access is highlighted. We then describe our research design in Section 3 and present the analyzed CHIIR paper characteristics in Section 4. This is followed by an evaluation of citations, citation contexts and altmetrics scores as indicators for publication impact in Section 5. The paper concludes with a discussion on possible correlations of paper types and their (predictable) impact, limitations of this approach and suggestions for future work.

2 RELATED WORK

2.1 Citation Purpose & Context

Citation-based factors continue to be the most important and established indicators for the evaluation of research impact. Starting with Merton [24], the purposes of citations in research publications remain an often discussed topic. While a normative theory of science states that references in publications point to prior scholarly work that influenced, impacted or supported the work described in a publication, many more purposes for citations have been identified [28]. In line with the continued interest in citation purposes, the number of studies identifying and measuring citation purposes in different fields of research is large.

In 2006, Teufel et al. [32] presented a citation function categorisation with four top-level categories and 12 sub-categories. Since then, this taxonomy has been used widely. Citation purposes and contexts have been either studied from the authors' perspective and motivations [11], based on actual citation discourse or context [37] or through a combination of both [39].

A large number of other categorization schemes have also been developed over time, which are reviewed and compared by meta-studies [22]. The review reveals that limited access and documentation of individual data sets as well as different annotation schemes and procedures are major challenges in citation classification work and generalization of results [22]. For example, there is no clear consensus on the optimal window size for the citation context (i.e. the number of words or sentences around a reference) [9, 22]. Automatic procedures for the identification of citations contexts have been proposed to overcome limitations of earlier manual annotations methods [12, 33].

Our particular research interest in documentation, re-use and sharing of research resources in CHIIR guided our review of classification schemes and category choices for this study. Our main selection criteria were parsimony (the taxonomy should cover only what we need and at the level of detail that is useful to us) and the inclusion of documentation and re-use purposes (the taxonomy

should have a category that covers re-use). Based on these requirements, the proposed taxonomy by Pride and Knoth [27] serves as a basis for the annotation scheme used in this paper which will be further introduced and discussed in the section on research design. Pride and Knoth [27] propose six main categories: (1) Background, (2) Uses, (3) Compare/Contrast (with three sub-categories: similarities-differences-disagreement), (4) Motivation, (5) Extension and (6) Future Work. In their study, from 11,233 annotations provided by more than 800 authors, 54.61% belong to the category Background as part of the literature review or background section in a paper. The category Uses (the citing paper uses the methodology or tools created by the cited paper) is particularly interesting for this work when it comes to the identification of re-use cases and will be extended to distinguish different types of re-use.

2.2 Measuring Research Impact through Alternative Factors

Traditional citation-based measures have been criticized for years with accompanying suggestions for alternative or complementary indicators for research impact evaluation. So called altmetrics suggest alternative factors that include societal and cultural impact of research output. Altmetrics refer to a very broad group of metrics, capturing various parts of impact a publication can have beyond citation-based impact. Publication of research output as open or closed access is also hypothesized to have an effect on research impact.

2.2.1 Altmetrics. The idea and usage of altmetrics aims at complementing traditional citation metrics with a broad set of indicators for measuring the impact of scientific publications in social media contexts. While there is no clearly agreed or standardized set of measures or indicators, providers like Altmetrics.com[2] offer aggregated scores based on different sources.

The Altmetrics score is a weighted count of the amount of attention that has been picked up for a research output. Mentions in some sources, such as news, blogs, policy documents or Wikipedia contribute a relatively high score, whereas mentions from other sources, such as social media and YouTube, contribute relatively little.[3] Altmetrics scores can decrease over time due to deleted posts or due to down-weighting of X (formerly known as Twitter) accounts in case of high bias or excessive focus on scholarly content.[4]

Using these scores, an increasing body of research has investigated the relationship between altmetrics and citation data with the result of limited correlation between both [10, 16]. Similarly, Costas et al. [10] found that through the usage of altmetrics, highly-cited publications could be identified with a higher precision rate but with lower recall than journal scores. This suggests that traditional citations still provide a better overview of related research while altmetrics allow a quicker and alternative entry point into research.

While citation contexts have been studied in extensive detail, this is only partly done for social media contexts. To better understand how altmetrics scores are impacted, Haustein et al. [16]

---

[2]https://www.altmetric.com/
[3]https://help.altmetric.com/support/solutions/articles/6000233311-how-is-the-altmetric-attention-score-calculated-
[4]https://help.altmetric.com/support/solutions/articles/6000241983-faq-why-has-the-altmetric-attention-score-for-my-paper-gone-down-

investigated if and how publication properties and collaboration factors (discipline, author number and background, paper type, length, references) have an effect on altmetrics scores. In contrast to citation counts that are typically affected by longer papers and higher numbers of references, this could not be observed for social media outlets, which prefer document types such as editorials or short reports [16]. Other studies have focused on specific platforms such as Mendeley or X, finding a higher correlation between citations and Mendeley reads than for citations and tweet mentions [17].

One of the main challenges for alternative research impact indicators is that altmetrics scores are still only available for a low percentage of papers. Disciplinary differences can be observed with social sciences and humanities at the top of altmetrics and computer science and natural sciences with the lowest scores [10, 16].

In summary, altmetrics seem not an alternative but rather provide complementary perspectives to citation information.

2.2.2 Open Access. The relationship between publishing research as Open Access (OA) and the impact on science, for example through citation counts, remains unclear [23]. While some studies report an Open Access citation advantage (OACA) [25], the majority of analyses report limitations or warnings against selection biases. Differences between studies based on the selected samples and different disciplines do not allow a generalization with respect to a correlation between OA publications and citation rates [19]. Some also argue that OA publication impact might be better or complementary represented by altmetrics such as downloads, mentions, likes etc [23].

As shown above, research impact is a multi-faceted concept that has been studied from different perspectives. For this project we investigate the impact of CHIIR papers based on three main factors: (1) citation counts, (2) citation contexts in which the papers have been used and (3) altmetrics scores.

3 RESEARCH DESIGN

In this study, the CHIIR conference papers serve as a proxy for a research community. CHIIR is an interesting research community to study for research impact and citation contexts as it represents the intersection of several fields of research, in particular information retrieval and human computer interaction while forming a nucleus around which the community shapes itself. It is expected that CHIIR research reaches both the information retrieval and human computer interaction research communities, which could be seen in citations from outside CHIIR. However, since the first CHIIR conference in 2016, the interactive information retrieval research community has also grown, so that we may see CHIIR papers referencing other CHIIR papers. It remains to be seen how these different impacts appear and whether this changes over time.

3.1 Data Collection

The basis of the analysis was Bogers et al.'s 2023 dataset [6], which we extended to include the CHIIR 2023 papers. The additional papers were annotated and categorized by research type, method, focus, sharing, and re-use using the same methodology as for the original dataset. The resulting, extended dataset contains 407 full, short, perspective, demonstration, and resource papers from the

CHIIR conference series between 2016-2023. This allowed us to analyze whether sharing, re-use, or paper type have an influence on the paper's impact.

Citation data. To analyze academic impact in terms of citation counts, we collected the citations of each CHIIR paper from Google Scholar using SerpApi.[5] Computer science fields in general—and conferences in particular—are poorly covered in traditional journal-based citation indexes such as Web of Science and Scopus, which led us to use Google Scholar, even though we are aware of the heterogeneous data quality. We used SerpApi to query Google Scholar for each of the 407 CHIIR papers by their DOIs. Next, we downloaded the snippet information for each of the 3,816 citing publications using SerpApi and extracted the relevant paper metadata, including publication year (if available) and their citation counts. All citation data was collected on August 22, 2023.

Altmetrics. Using the Altmetric API,[6] we were able to collect altmetrics data for 33.4% of the 407 CHIIR papers. According to Banshal et al. [4], coverage of Altmetrics varies considerably by discipline. CHIIR can be seen as a combination of the information sciences (33.5%) and arts & humanities (27.3%) examined by Banshal et al. [4], which is similar to the 33.4% of CHIIR papers that could be matched against the Altmetric API.

Additional data. We used the OpenAlex API[7] to collect additional data about the CHIIR papers and the publications citing them. OpenAlex is an open catalog of scholarly papers, authors, institutions, venues, and concepts with a free API service. We were able to match all 407 CHIIR papers to their corresponding OpenAlex records using their DOIs. For the citing publications collected in our Google Scholar crawl, we searched for their matching OpenAlex records by title, as Google Scholar and SerpAPI do not return DOIs. For each matching publication, we collected their Open Access status, detailed information about author affiliations and scientific concepts assigned by an automatic classifier.[8] We were able to match 2,656 out of 3,816 total citing documents (69.6%) using OpenAlex. When we present results using OpenAlex data for citing documents, they were calculated on this subset.

3.2 Data Annotation of the CHIIR Conference Papers

For this part of the study, we re-used the process developed by [6] to extend their dataset with the CHIIR 2023 papers. For clarity, we briefly recap the process here.

For *sharing and re-use*, the annotation schema distinguished between the three types of research resources defined by [14]: data ("any data that has been collected, observed, generated or created during or as a result of the research process"), design ("the methods and techniques used to collect and analyze empirical data"), and infrastructure (technical infrastructure providing "access to an IR system as well as the application of the data collection techniques"). Re-use is defined as any use after the initial publication and sharing

---

[5] https://www.serpapi.com/
[6] https://www.altmetric.com/solutions/altmetric-api/
[7] https://docs.openalex.org/
[8] https://docs.openalex.org/api-entities/concepts

is defined as providing access to any of the resource types after the initial publication.

For the *research types* there are three categories, which were based on and adapted from Kelly's definitions of exploratory, descriptive and explanatory research [21]: empirical, theoretical and resource. While empirical research commonly gathers, processes and analyzes some form(s) of data to confirm a hypothesis or answer a research question, theoretical research uses deductive and literature-based approaches to introduce new research questions, define research concepts or summarize the state-of-the-art. Resource papers present or demonstrate a dataset or tool.

For the *research methods*, the existing annotation codes were re-used and where necessary, we extended them with new or more detailed codes. For the analysis, the annotated methods were consolidated into larger categories, e.g. expert interviews, focus groups and user interviews were consolidated into a single category interview / focus groups.

3.3 Data Annotation of the Citation Contexts for CHIIR Papers

3.3.1 Data Collection and Extraction. We used the SerpAPI to download all available PDF versions of publications that cite CHIIR papers (citing publications). In total, 2,799 PDFs were downloaded for 1,465 citing publications (38% of all 3,816 citing papers). The distribution is skewed, with a mean of 1.91 ($Md = 1, Sd = 1.26$) PDF versions per publication and a maximum of 9 versions.

We used GROBID [1] for transforming the PDFs to TEI (Text Encoding Initiative), with XML markup for section, paragraph and sentence boundaries, and for extracting references and the text citations to these references. Some PDFs could not be parsed by GROBID, resulting in 2,163 TEI versions of 1,147 citing publications, with a total of 109,644 citations in 1,125 citing publications (for 22 publications, GROBID identified no citations). For publications with multiple PDF versions, we selected the version with the highest number of extracted citations. There are several drawbacks to using these automated steps to identify citation contexts and motivations. First, errors were present in each of the steps. The publications metadata (e.g. title and author) was incorrect for a handful of publications, which leads to incorrectly identified links between citing and cited paper. The same goes for the citation matches, i.e. Google sometimes incorrectly identified a CHIIR paper as being cited by a citing publication. Second, GROBID extracted incorrect publication metadata for some publications, and missed some citations in the text, as well as references in the reference list. Third, GROBID only extracts explicit citations, but there are also indirect and implicit citations [3, 18]. For instance, the references to GROBID in the previous two sentences and in this one are implicit references. In this paper we limit ourselves to the explicit references identified by GROBID, but note that the implicit references can have additional motivations or meanings.

To identify which citations in the TEI versions of citing publications reference a CHIIR publication, we compared the titles in the reference lists of citing publications to the titles of the CHIIR papers that the SerpAPI linked to the citing publications. To measure similarity between the title of a publication in the reference list t_r and the title of the linked CHIIR paper t_C, we used edit distance

between the two titles and computed similarity as the inverse ratio of the edit distance $ed(t_r, t_C)$ to the length $|t_C|$ of the CHIIR publication title in number of characters: $sim(t_r, t_C) = \frac{|t_C| - ed(t_r, t_C)}{|t_C|}$. The similarity distribution between all pairs showed that scores were either above 0.6 or below 0.35. A manual check found that all scores above 0.6 were correct pairs, while a random sample of title pairs with a score below 0.35 contained only non-matching pairs. As a result, the similarity threshold for equivalence was set at $sim(t_r, t_C) \geq 0.6$.

We were able to extract 2,698 citation contexts for 264 cited CHIIR papers (65% of all CHIIR papers and 74% of cited CHIIR papers according to Google Scholar) from 952 citing publications (25% of all citing papers, and 85% of citing papers for which GROBID identified at least one citation).

Earlier work on citation context classification found that in most cases, the sentence containing the citation is enough to classify its motivation [35, 40], but in some cases the surrounding sentences are needed [2]. We use two types of representations of citation contexts, 1) the sentence containing the citation, and 2) the citing sentence with the preceding and following sentence, when preceding and following sentence were part of the same paragraph as the citing sentence. Annotators were given both types of contexts.

3.3.2 Citation Context Annotation. Through the review of existing annotation schemes it appeared that their granularity is mostly too detailed for this analysis, while at the same time generally being under-defined with respect to our interest in re-use. To address this, a classification system based on [27] was developed, with extensions for re-use of research data, design, and infrastructure (Table 1).

In total, two annotators categorized 1,492 citation contexts, with each citation context processed by one annotator. Annotators were instructed to annotate each citation context with all matching categories. The annotators initially used the sentence within which the citation occurs as the justification for the categorisation. For 214 citations, the citation sentence did not provide enough context and for these the annotators also considered the sentences before and after the citation sentence. For each citation context, the annotators were asked to indicate their confidence as low, middle or high.

A third annotator checked all citations that had initially been marked as low confidence, as well as all citations that had been marked as unclear and added their annotation to confirm or reject a category. The validation process identified a total of 32 citations that were identified incorrectly by GROBID. These were excluded from the analysis.

One limitation of the citation context analysis is that, in the interface the annotators used to classify, the citations were grouped by paper, but the papers were in no specific order. Due to artefacts in the ordering and the fact that only approximately half of the contexts were annotated, demonstration and short papers are significantly under-represented in the annotated data (chi^2 contingency with Bonferroni correction at $p < 0.001$). In particular, this impacts the analysis by paper type, where resource papers are under-represented.

Table 1: Citation context categories for CHIIR papers.

Category	Description (adapted from [27] & [14])
Background	The cited paper provides relevant background information or is part of the body of literature.
Uses	The citing paper re-uses resources created by the cited paper.
Design	The citing paper re-uses methods and techniques, including research environment, data collection protocols as well as data analysis methods and measures.
Data	The citing paper re-uses any data that has been collected, observed, generated or created during or as a result of the research process.
Infrastructure	The citing paper re-used the IR system (i.e. software, interfaces, collections) or applied data collection tools for user and tasks management, questionnaires,interaction logging, etc.
Compare/Contrast	The citing paper expresses a relationship to the cited paper.
Similarities	The citing paper uses a similar approach or shows similar results.
Differences	The citing paper contrasts or differs from the cited paper in any aspect.
Disagreement	The citing paper disagrees with the cited paper or with parts of it.
Motivation	The citing paper or research question is directly motivated by the cited paper.
Extension	The citing paper builds upon and extends the methods, tools or data etc. of the cited paper.
Future Work	The cited paper may be a potential avenue for future work.

Figure 1: Frequency of research types. Papers can adopt multiple research types, so total count exceeds the total number of papers (N = 407).

Figure 2: Frequency of data collection methods. Papers can adopt multiple data collection methods, so total count exceeds the total number of papers (N = 407).
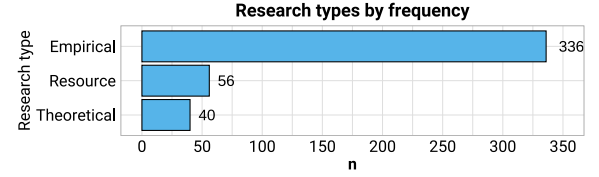
4 CHARACTERIZING THE CHIIR PAPERS

This section presents a brief overview of the characteristics of the 2016-2023 CHIIR conference proceedings collection. While the overall distributions are comparable to the Bogers et al. study [6], resource sharing and re-use fluctuates more widely because of the small frequencies observed.
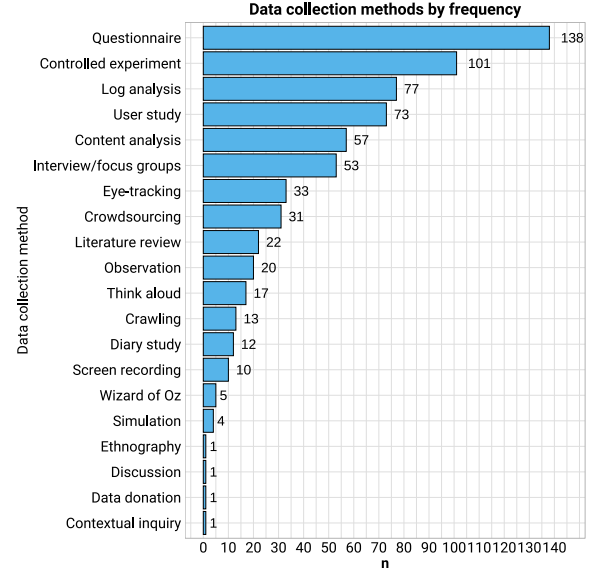
4.1 Research Type and Research Method

The CHIIR proceedings contain 185 short papers, 169, long papers, 30 demo papers, 20 perspective papers and 3 resource papers.

Figure 1 shows the distribution of research types for the complete data set. It shows that CHIIR predominantly covers *empirical* research (n = 336, 82.6%), followed by a smaller number of *resource* papers (n = 56, 13.8%) and *theoretical* research (n = 40, 9.8%). Note that this is based on the manual annotation of the research types by [6] of the CHIIR papers until 2022 and by one author of this paper for CHIIR 2023 papers, not the CHIIR paper categories, which are not equivalent. For example, while all 3 papers categorised by

CHIIR as "Resource paper" are annotated as research type *resource*, the annotated research type *resource* contains a further 53 papers, drawn primarily from the "Short" and "Demo" paper categories. Similarly, papers with the research type *theoretical* can be found in the long, short and perspective CHIIR paper categories.

Figure 2 shows the prevalence of empirical research reflected in the data collection methods used in CHIIR papers. Questionnaires are the most commonly used method, probably because they are often included as part of other research designs such as controlled experiments, usability tests, or other mixed-methods design. Of the ten most popular methods, only one (literature review at rank 9) is not exclusively associated with empirical research.

4.2 Sharing vs. Re-use

Over the eight years of CHIIR, 28.5% of papers re-used existing data, 29.0% re-used research designs, and 12.0% re-used infrastructure. At the same time, 11.8% shared (part of) their research data, 13.5% shared elements of their research design (e.g. survey questions or task descriptions), and 9.3% shared at least some infrastructure
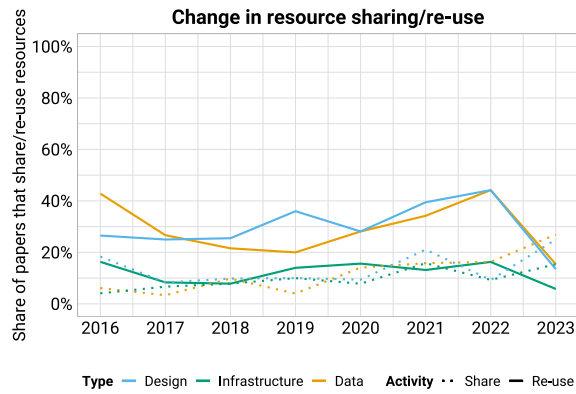
Figure 3: Change in resource sharing and re-use at CHIIR over time as a percentage of all papers per year. The dotted lines indicate sharing activity grouped by resource type, while the solid lines indicate re-use activity.

components. Figure 3 documents the change over time. While sharing and re-use numbers appear to be slowly increasing, they are still relatively low compared to other research areas [31]. A real pattern is hard to discern because of the low frequencies per year. The re-use of infrastructure appears to be particularly difficult. This is confirmed by anecdotal evidence that most infrastructure re-use appears from the same research groups that developed the infrastructure that is being re-used.

5 QUANTIFYING IMPACT

5.1 Citation impact

Using the method described in Section 3.1, we were able to identify a total of 5,706 citations to CHIIR publications on Google Scholar. These citations were produced by 3,816 unique publications, indicating that several publications each cite multiple CHIIR publications. On average, a citing document contains 1.50 references to a CHIIR paper ($Md = 1$), although this distribution is highly skewed, with only 9.6% of all citing documents citing more than 2 CHIIR papers in the same publication.

Looking at the academic impact in terms of normalized citation counts (the total number of verified citations of a publication divided by the number of years since publication) during the first eight years of CHIIR, we can see that its publications have received an average of 3.03 citations per year ($Md = 1.78$, $SD = 4.61$). 48 out of 407 CHIIR publications (or 11.8%) remained uncited at the time of crawling Google Scholar, although this distribution is highly temporally skewed. Exactly two-thirds of these publications were from the most recent (2023) edition of CHIIR and another 12.5% from 2022. Only 2.5% of CHIIR publications published in 2021 or earlier still have not accrued any citations. Figure 4(a) shows the normalized citation counts for all CHIIR papers. While an average between two to five citations per year may not seem high, it is close to the impact of the Conference on Human Factors in Computing Systems (CHI) publications during the same period [26]. It is also in line with the mean normalized number of citations that the 3,816 citing documents crawled from Google Scholar received, which

is 4.07 ($Md = 1.2$, $SD = 13.5$). The substantially greater standard deviation may be indicative of the less focused topical range of the citing documents.

The normalized citation count averages by year (Figure 4a) form a coarse summary of the academic impact of individual papers based on different characteristics. Bogers et al. [7] identified a small set of attributes of (CHIIR) publications that could have an influence on their impact. For instance, it is reasonable to assume that publications that share some type of resource—data, design or infrastructure—will accrue more citations as other researchers re-use those resources. Figure 4b shows the differences in normalized citation counts by whether a paper shares at least one resource or none at all. It does not show a clear influence of resource sharing on academic impact for all three resource types combined. Papers that simply re-use existing resources versus not re-using any resources also do not show any meaningful difference (Figure 4c).

Making a publication open access can lower the barriers for other researchers to engage with that publication, potentially leading to greater impact. In comparing closed and open access publications, some studies observe an open access citation advantage, assuming a higher interaction and impact for those publications [25]. Figure 4d shows the difference in normalized citation counts in terms of open access availability. In recent years, a larger number of CHIIR publications has been made available as open access, which could be expected to have a positive influence on their impact. However, in line with Langham-Putrow et al. [23], there does not seem to be a clear impact of open access status on citation impact.

Finally, it is reasonable to expect that different types of research have different impact. Figure 4e visualizes the difference in citation counts between the three different research types. It shows that with the exception of one year, resource papers have the lowest impact of the three types on average. Figure 4e also seems to suggests that theoretical papers have a greater impact than empirical papers, although this difference is not statistically significant according to a Mann-Whitney U-test ($U(N_{empirical} = 103, N_{theoretical} = 304)$ = 5727.5, $z = 1.53$, p = 0.13). We revisit these features and their predictiveness for academic impact in Section 7.

Citation distribution. Figure 5a shows the rank-frequency distribution of citation counts for CHIIR papers, which matches the expected long-tail distribution from bibliometrics studies [8]. While we have not collected complete citation data for other conferences in the same time period, we can approximate citation distributions outside of CHIIR by visualizing the citation count distribution of the publications that cite CHIIR papers in Figure 5b. We can assume that the citing documents will largely represent the same research field. These two distributions bear a strong resemblance, suggesting that the citation patterns for this field are similar inside and outside of CHIIR and that the two share similar characteristics.

Self-citations. Of the 5,706 citations to CHIIR publications, 382 (6.7%) are from other CHIIR publications while 5,324 (93.3%) come from outside the CHIIR conference series. This suggests that CHIIR has a strong impact outside of the conference itself. However, some of this impact may be over-inflated due to self-citations. Self-citation occurs when an author references another of their own publications. While self-citation can be a legitimate way to build upon earlier work, sometimes self-citations can be unduly made in attempt to
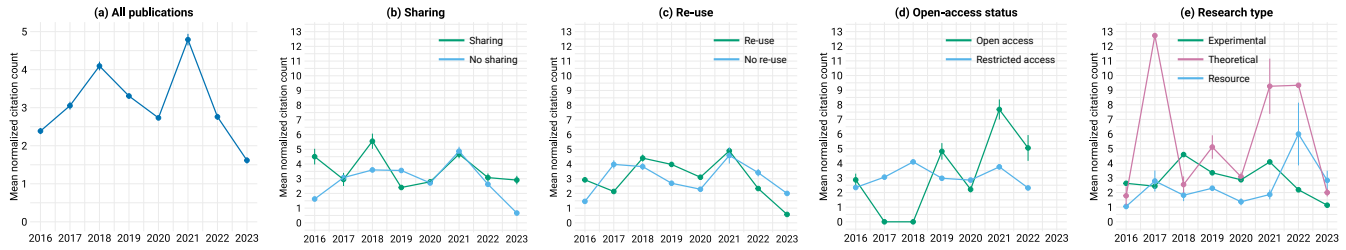
Figure 4: Mean normalized citation counts with standard error bars for (a) all CHIIR papers (N = 407) grouped per year as well as split by (b) resource sharing, (c) resource re-use, (d) open access status, (e) research type. Open access data for 2023 is left out as all 2023 were erroneously marked as open access in OpenAlex, resulting in an flawed comparison.
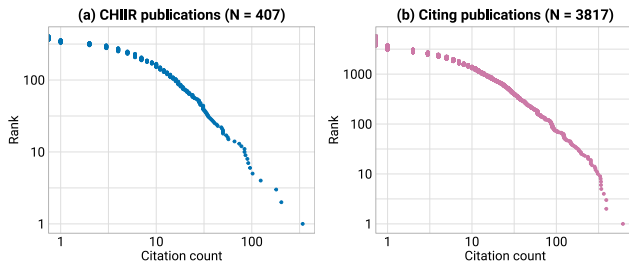
Figure 5: Rank-frequency distributions of the citation counts for (a) CHIIR papers (N = 407) and (b) all the publications that have cited CHIIR papers (N = 3,816).

Table 2: Number of CHIIR papers with a non-zero Altmetrics count for 13 features.

Feature	Mean	Sd	Md	Max	# Non-zero scores
Redditors	0.0	0.0	0	0	0
Patents	0.0	0.0	0	0	0
Videos	0.0	0.0	0	0	0
Wikipedia	0.0	0.1	0	1	2
Tweeters	0.0	0.2	0	2	3
Policies	0.0	0.1	0	1	4
Facebook Walls	0.0	0.1	0	1	5
Blogs	0.0	0.3	0	4	7
News Sources	0.2	2.2	0	30	7
Forums	1.8	8.6	0	155	129
Accounts	2.1	10.7	0	190	135
Posts	2.7	15.3	0	274	135
Altmetrics Score	3.2	23.5	0	355.9	135

inflate an individual's citation count, thereby also over-inflating the impact a publication has had. In our dataset, 745 of all 5,706 citations (or 13.1%) are self-citations by authors of CHIIR papers to their own work published at CHIIR, which corresponds to 231 out of 407 papers that were self-cited at least once. Out of all 5,706 citations, 2,294 (or 40.2%) come from citing papers that have no authors that have ever authored a CHIIR paper before. This also suggests that CHIIR papers are received outside of the authoring CHIIR community.

Another relevant question to ask is *where* these citations are coming from. Unfortunately, citing sources are hard to identify. Often, journal and conference names are abbreviated in the bibliographic data, which means that the different variants are difficult to relate to one another. Therefore, no reliable statement about top journals and conferences can be made at this time. If we disregard the self-citations and only consider the 4,951 true citations that CHIIR papers have accrued, we can, for instance, analyze the institutions of the citing researchers to determine the geographic distribution. Figure 6a shows that the geographic distribution between CHIIR papers and citing papers is largely the same with China ('CN') being a notable difference as authors with a Chinese affiliation cite CHIIR papers much more than they publish there. While CHIIR papers are concentrated among a smaller number of countries (mostly Anglo-American) citation impact from CHIIR papers is more widespread.

Figure 6b shows what type of institutions produce CHIIR papers and cite them. These distributions are virtually identical, but it does show that CHIIR's impact goes beyond academic institutions

as around twenty percent of the impact is on companies, facilities, non-profit organizations and other non-academic institutions. This suggests that CHIIR has an impact beyond purely academic research.

5.2 Altmetrics impact

The Altmetrics API provides data per publication consisting of an overall score and counts of 12 features, such as mentions on social media platforms, in blog posts or in Wikipedia articles (see Table 2). For three features—number of mentions by *Redditors*, or in *Patents* or *Videos*—all 407 papers have a count of zero, and for six other features, there are fewer than 10 papers with a positive count. Only for the three features *Forums* (count of mentions on forums and Stack Exchanges), the number of *Accounts* and *Posts* (the sum of users and citations of all other features combined) activities could be observed. The altmetrics data is thus sparse, and as a consequence, only 135 of the 407 CHIIR papers (33%) have an Altmetrics score above zero.

The mean altmetrics scores of CHIIR publications grouped by year is shown in Figure 7, which shows a peak for 2021 and 2022. This is due to one or two papers in each of those years having very high scores: two papers with scores 41.55 and 251.08 in 2021,
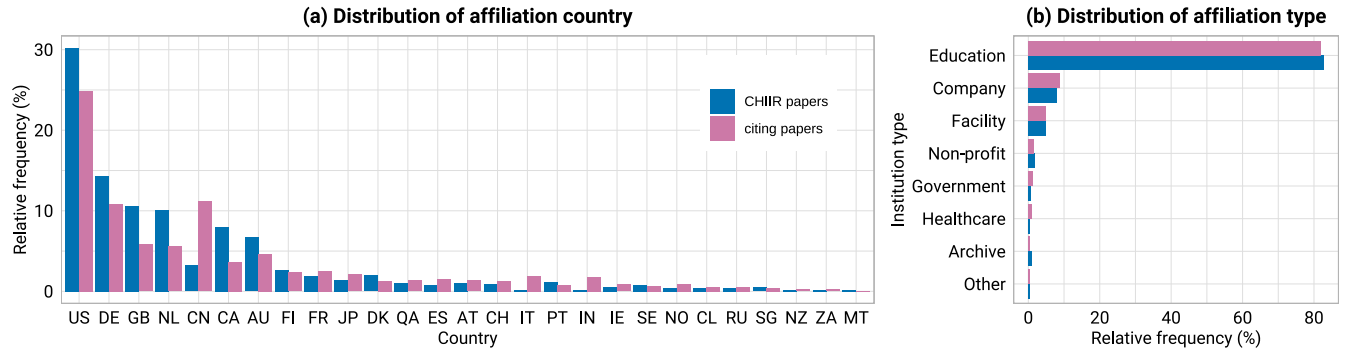
Figure 6: Distributions of the (a) author affiliation countries and (b) institution types for both the CHIIR papers ($N = 407$) and the citing publications ($N = 3,816$). Affiliation data is from OpenAlex.
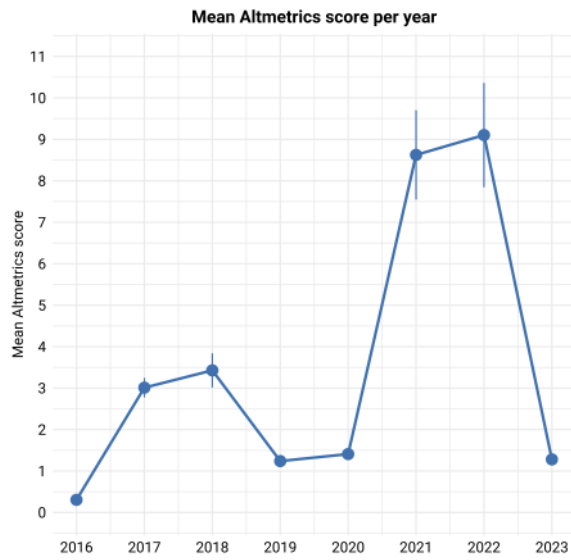
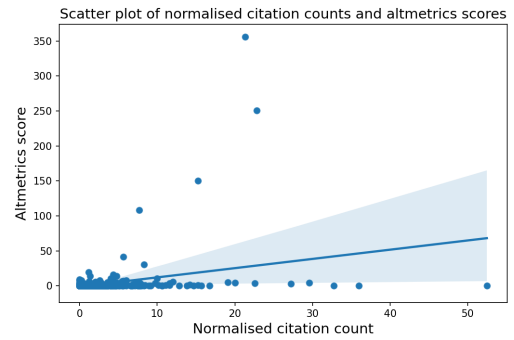

Figure 7: Mean Altmetrics scores with standard error bars for all CHIIR papers ($N = 407$) grouped per year

Figure 8: Relationship between normalized citation counts and altmetrics scores

and one with a score 335.90 in 2022, while all other papers in those years score close to zero. In other words, the per-year data is highly skewed and the (arithmetic) means are more reflective of the extremes than of the bulk of the data. For that reason, we leave out the per year means differentiated by research type, sharing, re-use and Open Access. Given that the scores provided by Altmetrics are already a weighted sum, and previous studies have used arithmetic means for comparing sets of scores [34], we use the same means calculation. One alternative is to use a different calculation of the mean, e.g. geometric mean, or to model the scores as a power-law distribution.

The relationship between normalized citation counts and altmetrics scores is shown in Figure 8. The Pearson correlation between normalized citation counts and altmetrics scores is 0.29. This shows that, particularly for publications with higher scores, there is a large discrepancy between citation counts and altmetrics. The CHIIR

papers with the highest citation counts do not have the highest altmetrics scores, although papers with high altmetrics scores also tend to have relatively high normalised citation counts. This could be partly due to citation counts having a bias towards older publications and altmetrics scores having a bias towards more recent publications, as social media attention is a more recent phenomenon [30]. But given that the social media channels that are included in altmetrics were already well established when the first CHIIR conference took place, and that we use normalized citation counts, we expect the effect of this time difference to be minimal. It is more likely that, at least on the case of CHIIR publications, citations counts and altmetrics measure different aspects of impact.

From these observations it seems that, outside of academic citations, the impact of most CHIIR papers is limited.

6 CATEGORIZING IMPACT

Publications can be cited for different reasons. To analyze the motivations for citing CHIIR publications, we annotated the contexts of citations to CHIIR publications using an adapted version of another citation context classification (see Section 3.3).

The annotated data set contains 1,460 citation contexts. Since annotators could annotate each context with more than one category, the data set contains 1,561 citation-category pairs (see Table 3). The

data clearly shows that CHIIR citations are mostly used as background (82%). Two recent large annotated data sets using (almost) the same classification schema report percentages for background of 51.8% [20] and 54.6% [27]. Due to the relatively large proportion of empirical work within CHIIR, this prompts the question why these papers are mostly cited to provide background information, and are rarely cited for comparison, motivation or because elements of the research are used by others. It is possible that this is an artefact of the annotation process, with the single-sentence citation context not providing enough detail. However, the annotators noted where they used the extended citation context and of those citations 73% are classified as background. A further indicator of annotator influence is that the annotator confidence (columns 2 and 3 in Table 3) for the background category is significantly higher than for both the uses and compare-contrast categories (chi^2 contingency with Bonferroni correction with $p < 0.001$).

However, while the limited citation context is likely to have an influence, this is unlikely to be the sole influence. Even when taking into account a wider citation context, the fraction of background citations is still around 20 percentage-points higher than in the previous literature and likely indicates a difference in how the II&R field works. There is clearly less re-use happening, with only 7.8% of citations in the uses category, compared to 18.5% and 15.5% in [20] and [27] respectively. This is possibly influenced by CHIIR papers often being subject to ethical and commercial limitations, and thus cannot be shared and re-used as much. However, it may also indicate that documentation practices could be improved to enable more re-use. The fact that compare-contrast is also lower at 7% of CHIIR citations, compared to 12.0% and 17.5%, strengthens the assumption that documentation practices could be improved, as it is difficult to cite similarities or differences, if not enough detail is shared in the original paper to make a firm statement.

Finally, the use of CHIIR papers to motivate further research represents 1% of citations, where previous work showed 10% falling into that category [27]. We hypothesise that it is the very focused nature of the primarily experimental publications in CHIIR. These tend to investigate a single phenomenon under very specific conditions. Unless the citing paper is interested in exactly the same phenomenon, the work is thus unlikely to be motivational, but will provide contextual background, leading to the observed pattern. This is supported by a similar fraction of 1% of citations representing an extension of the original work, compared to 3.7% and 6.2% in the literature.

The type of publication ("resource", "empirical", "theoretical") shows an expected and significant impact on how it is cited (all pairwise chi^2 contingency tests with Bonferroni correction with $p < 0.001$). Unsurprisingly, theoretical papers are almost exclusively cited in the background (96%), whereas empirical and resource papers are used across all categories, with one interesting exception, namely that no resource paper citation has been categorised as "Compare / Contrast - Differences". The difference between empirical and resource papers is mainly driven by the "Compare / Contrast" sub-categories containing a higher fraction of empirical papers and the "Uses" sub-categories containing a higher fraction of resource papers. However, if the analysis is repeated with just the top-level categories, then the difference between empirical and

Table 3: Number of annotations per citation purpose with High and Medium confidence.

Category	High	Medium	Total
Background	901	385	1286
Uses	60	61	121
- Design	31	37	68
- Data	20	21	41
- Infrastructure	9	3	12
Compare / Contrast	35	74	109
- Similarities	23	41	64
- Differences	12	30	42
- Disagreement	0	3	3
Motivation	13	3	16
Extension	9	11	20
Future Work	3	6	9
Total	1021	540	1561

resource papers is no longer significant (pairwise chi^2 with Bonferroni at $p = 0.06$), although this may be influenced by the fact that resource papers are slightly underrepresented in this dataset.

7 PREDICTING IMPACT

A logical follow-up to our descriptive analysis of the citation impact of CHIIR papers is to study whether it is possible to predict the future impact of CHIIR papers. For this purpose, we performed a quasi-Poisson regression to predict the normalized citation count of CHIIR publications based on different sets of variables: (1) whether they shared one of three different resource types (data, design, and infrastructure); (2) whether they re-used at least one resource type; (3) the research type of the publication in question; (4) the open access status of the paper; (5) the number of authors of the paper; and (6) the number of unique affiliations associated with the paper. The latter two were chosen to test the notion that collaboration could lead to multi-faceted papers and thereby higher impact. We chose a quasi-Poisson regression, because citation counts are known to follow a Poisson (log-normal) distribution [8, 29] and because normalized citation counts are not integer values. We did not choose to include the separate resource re-use variables into the regression model, because they are even more likely to be correlated with each other and because a Mann-Whitney U test showed no impact of resource re-use on citation counts.

Table 4 shows the results of our Poisson regression analysis. Sharing data resources ($p = 0.026$) significantly and positively affects the number of citations: sharing at least one data resource increases the citation count by 1.55 extra citations. This suggests that sharing data has the greatest impact, because it is easier for researchers to re-use data than to re-use design elements or infrastructure, as also predicted by Hall [15]. A publication's research type, resource re-use, open access status and our proxy variables for collaboration do not significantly affect the citation counts.

We also performed the quasi-Poisson regression on the Altmetrics scores, resulting in no features that have a significant effect. However, the sparsity of the Altmetrics scores data means that no firm conclusions can be drawn from this.

Table 4: Regression values for resource sharing, re-use and research design type according to a quasi-Poisson regression. Significance is marked as * for $\alpha = 0.001$, ** for $\alpha = 0.01$ and * for $\alpha = 0.05$.**

Variable	β	e^{β}	SE	p-value
(Intercept)	0.977**	2.657	0.366	0.008
Shared infrastructure	-0.150	0.861	0.283	0.598
Shared data	0.437*	1.548	0.195	0.026
Shared design	0.227	1.255	0.195	0.245
Re-used at least one resource	0.177	1.194	0.149	0.236
Empirical paper	-0.481	0.618	0.348	0.168
Theoretical paper	0.399	1.490	0.353	0.260
Resource paper	-0.475	0.622	0.351	0.177
Open access	-0.219	0.803	0.147	0.137
No. of authors	0.086	1.089	0.047	0.067
No. of unique affiliations	0.089	1.094	0.083	0.282

8 DISCUSSION & CONCLUSIONS

In this paper we studied the impact of CHIIR papers via a range of measures. The first conclusion to draw from the analysis is that CHIIR is a publication venue with significant impact. Almost all CHIIR papers are cited within a few years after publication. In terms of reach, CHIIR is an international conference with a far-reaching reception in the wider academic community and other domains such as industry.

The second conclusion is that the impact of sharing resources is complex. Resource papers on their own generate significantly fewer citations than experimental and theoretical papers, with theoretical papers seemingly having the highest impact. At the same time, papers that explicitly share their data, regardless of the type of paper, are cited significantly more often. When combined with the fact that only a small fraction of papers explicitly share their data, we believe this makes a very strong case that the CHIIR community can do more to share the data that form the basis for or are the outcomes of the experiments presented in this forum, with the potential for significant impact benefits for both the authors and the community. At the same time we recognise that some research data cannot be shared for either ethical or commercial reasons. How to ensure that these groups are not negatively impacted by the push for more sharing is an important and open question.

In previous work it has been suggested that II&R research is dispersed and disconnected, with each study being essentially unique, resulting in a lower perceived value of sharing and re-use [5]. However, we are not aware of any study that has explicitly investigated related communities such as CHI or SIGIR with respect to re-use and sharing. The nature of studies presented at CHIIR may create some bias against easily and effectively re-useing, but considering the low amount that is being shared in the first place, it is impossible to determine whether this is a causal or just coincidental relationship. Analysing the reasons for citing a CHIIR paper shows that they are predominantly cited for background information, rather than for motivation, extension, use, or comparison. The difference to other areas, where background citations only make up around 50%, is so pronounced, that we believe that it is the lack of sharing

or at least documenting that is driving this difference and that this is something the community should address.

At the same time citation counts do not tell the whole story and in particular the lower level of citations for resource papers should not in themselves be used as a justification for ignoring them. A paper with a lower citation count may still have a larger impact, because it is cited by other publications that make use of some of its shared resources, rather than just as one of many background citations. Moreover, it is possible that resource papers increase the visibility and reuse of the resource, while the subsequent citation does not go to these resource papers. Instead, the reuse may come with a reference to the website from which the resource was downloaded, or to an experimental paper that mentions both the resource and related findings that are relevant to the citing paper.

While we believe that better sharing and documentation would improve the impact of the CHIIR community, we do not know whether the effort required for better sharing and documentation would justify the increased impact. We thus have to consider that the current structure is working as desired. If CHIIR publications are meant to be self-contained and the main purpose of CHIIR publications is to provide background for later studies, perhaps we should shift our efforts from encouraging sharing and documentation to better describing the findings and implications of our work.

The nature of the work means that there are some limitations. The citation context annotations cover only a small fraction of the citing publications and less than half of the cited CHIIR papers, and only the explicit citations at that. Therefore, it is possible that a larger sample would reveal different patterns. In particular there might be a significant shift in the distribution of citation categories.

Also, as mentioned above, citation counts and altmetrics scores are only proxies for impact, and as a consequence, our analysis provides only a limited view on research impact. We should therefore be conservative in drawing conclusions.

One aspect that our analysis does not cover is how impact is created through research collaborations. If researchers discover relevant connections between their work and that of others, e.g. by seeing presentations at a CHIIR conference, and decide to collaborate on joint research, they each bring their previous experiences and background knowledge to the new work. Although they may cite some of their previous publications, part of the impact may remain implicit, and is thus not reflected in citation counts, purposes or altmetrics.

Our findings and these limitations suggest directions for future work. To address the limitations, the citation context dataset should be extended to cover a larger fraction of the full citations, it should be fully annotated, and extended with the altmetrics data. Second, a deeper analysis of research collaborations as another form of impact may provide a richer picture of research impact. Third, an analysis with similar conferences would yield comparative insights into CHIIR's impact.

REFERENCES

[1] 2008–2023. GROBID. https://github.com/kermitt2/grobid. swh:1:dir:dab86b296e3c3216e2241968f0d63b68e8209d3c

[2] Amjad Abu-Jbara, Jefferson Ezra, and Dragomir Radev. 2013. Purpose and polarity of citation: Towards nlp-based bibliometrics. In *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics:*

Human language technologies. 596–606.

[3] Awais Athar and Simone Teufel. 2012. Detection of implicit citations for sentiment detection. In *Proceedings of the workshop on detecting structure in scholarly discourse.* 18–26.

[4] Sumit Kumar Banshal, Vivek Kumar Singh, Pranab K Muhuri, and Philipp Mayr. 2019. Disciplinary variations in altmetric coverage of scholarly articles. *arXiv preprint arXiv:1910.04205* (2019).

[5] Toine Bogers, Maria Gäde, Mark Hall, Luanne Freund, Marijn Koolen, Vivien Petras, and Mette Skov. 2018. Report on the Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2018). *SIGIR Forum* 52, 1 (Aug. 2018), 119–128.

[6] Toine Bogers, Maria Gäde, Mark Michael Hall, Marijn Koolen, Vivien Petras, and Birger Larsen. 2023. How We Work, Share, and Re-Use at CHIIR. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23).* Association for Computing Machinery, New York, NY, USA, 351–356. https://doi.org/10.1145/3576840.3578305

[7] Toine Bogers, Birger Larsen, Marijn Koolen, Maria Gäde, Mark M. Hall, and Vivien Petras. 2023. Collaboration Patterns and Impact of Sharing at CHIIR. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23).* Association for Computing Machinery, New York, NY, USA, 313–320. https://doi.org/10.1145/3576840.3578276

[8] Arnab Chatterjee, Asim Ghosh, and Bikas K Chakrabarti. 2016. Universality of citation distributions for academic institutions and journals. *PLoS one* 11, 1 (2016), e0146762.

[9] Arman Cohan, Waleed Ammar, Madeleine van Zuylen, and Field Cady. 2019. Structural Scaffolds for Citation Intent Classification in Scientific Publications. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers).* Association for Computational Linguistics, Minneapolis, Minnesota, 3586–3596. https://doi.org/10.18653/v1/N19-1361

[10] Rodrigo Costas, Zohreh Zahedi, and Paul Wouters. 2015. Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology* 66, 10 (2015), 2003–2019. https://doi.org/10.1002/asi.23309 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23309

[11] Martin G Erikson and Peter Erlandson. 2014. A taxonomy of motives to cite. *Social Studies of Science* 44, 4 (2014), 625–637. https://doi.org/10.1177/0306312714522871 arXiv:https://doi.org/10.1177/0306312714522871 PMID: 25272615.

[12] Mark Garzone and Robert E Mercer. 2000. Towards an automated citation classifier. In *Advances in Artificial Intelligence: 13th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, AI 2000 Montéal, Quebec, Canada, May 14–17, 2000 Proceedings 13.* Springer, 337–346.

[13] Trisha Greenhalgh, James Raftery, Steve Hanney, and Matthew Glover. 2016. Research Impact: A Narrative Review. *BMC Medicine* 14, 1 (2016), 1–16.

[14] Maria Gäde, Marijn Koolen, Mark Hall, Toine Bogers, and Vivien Petras. 2021. A Manifesto on Resource Re-Use in Interactive Information Retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21).* Association for Computing Machinery, New York, NY, USA, 141–149. https://doi.org/10.1145/3406522.3446056

[15] Mark M. Hall. 2019. To Re-use is to Re-write: Experiences with Re-using IIR Experiment Software. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use co-located with the ACM SIGIR Conference on Human Information Interaction and Retrieval, BIIRRR@CHIIR 2019, Glasgow, UK, March 14, 2019 (CEUR Workshop Proceedings, Vol. 2337),* Toine Bogers, Samuel Dodson, Maria Gäde, Luanne Freund, Mark M. Hall, Marijn Koolen, Vivien Petras, Nils Pharo, and Mette Skov (Eds.). CEUR-WS.org, 19–23. http://ceur-ws.org/Vol-2337/paper3.pdf

[16] Stefanie Haustein, Rodrigo Costas, and Vincent Larivière. 2015. Characterizing social media metrics of scholarly papers: the effect of document properties and collaboration patterns. *PLoS ONE* 10 (01 2015). https://doi.org/10.1371/journal.pone.0127830

[17] Stefanie Haustein, Vincent Larivière, Mike Thelwall, Didier Amyot, and Isabella Peters. 2014. Tweets vs. Mendeley readers: How do these two social media metrics differ? *it - Information Technology* 56, 5 (2014), 207–215. https://doi.org/doi:10.1515/itit-2014-1048

[18] Myriam Hernández-Alvarez and José M Gomez. 2016. Survey about citation context analysis: Tasks, techniques, and resources. *Natural Language Engineering* 22, 3 (2016), 327–349.

[19] D. Hopf, S. Dellmann, C. Hauschke, and M. Tullney. 2022. *Wirkungen von Open Access. Literaturstudie über empirische Arbeiten 2010-2021.* Hannover : Technische Informationsbibliothek (TIB).

[20] David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. 2018. Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics* 6 (2018), 391–406.

[21] Diane Kelly et al. 2009. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval* 3, 1–2 (2009), 1–224. https://doi.org/10.1561/1500000012

[22] Suchetha N Kunnath, Drahomira Herrmannova, David Pride, and Petr Knoth. 2021. A meta-analysis of semantic classification of citations. *Quantitative science studies* 2, 4 (2021), 1170–1215.

[23] Allison Langham-Putrow, Caitlin Bakker, and Amy Riegelman. 2021. Is the open access citation advantage real? A systematic review of the citation of open access and subscription-based articles. *PloS one* 16, 6 (2021), e0253129.

[24] R. Merton. 1957. Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review* 22 (1957), 635. https://api.semanticscholar.org/CorpusID:147253305

[25] Heather Piwowar, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West, and Stefanie Haustein. 2018. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ* 6 (2018), e4375.

[26] Henning Pohl and Aske Mottelson. 2019. How we Guide, Write, and Cite at CHI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19).* ACM, New York, NY, USA, 1–11. https://doi.org/10.1145/3290607.3310429

[27] David Pride and Petr Knoth. 2020. An Authoritative Approach to Citation Classification. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* (Virtual Event, China) *(JCDL '20).* Association for Computing Machinery, New York, NY, USA, 337–340. https://doi.org/10.1145/3383583.3398617

[28] Cara Seitz, Marion Schmidt, Nathalie Schwichtenberg, and Theresa Velden. 2021. A Case Study of the Epistemic Function of Citations - Implications for Citation-based Science Mapping.

[29] John A. Stewart. 1994. The Poisson-lognormal model for bibliometric/scientometric distributions. *Information Processing & Management* 30, 2 (1994), 239–251.

[30] Pardeep Sud and Mike Thelwall. 2014. Evaluating altmetrics. *scientometrics* 98 (2014), 1131–1143.

[31] Leho Tedersoo, Rainer Küngas, Ester Oras, Kajar Köster, Helen Eenmaa, Äli Leijen, Margus Pedaste, Marju Raju, Anastasiya Astapova, Heli Lukner, Karin Kogermann, and Tuul Sepp. 2021. Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data* 8 (2021). https://api.semanticscholar.org/CorpusID:236461324

[32] Simone Teufel, Advaith Siddharthan, and Dan Tidhar. 2006. An annotation scheme for citation function. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue.* 80–87. https://aclanthology.org/W06-1312.pdf

[33] Simone Teufel, Advaith Siddharthan, and Dan Tidhar. 2006. Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing.* 103–110. https://aclanthology.org/W06-1613.pdf

[34] Mike Thelwall, Stefanie Haustein, Vincent Larivière, and Cassidy R Sugimoto. 2013. Do altmetrics work? Twitter and ten other social web services. *PloS one* 8, 5 (2013), e64841.

[35] Marco Valenzuela, Vu Ha, and Oren Etzioni. 2015. Identifying Meaningful Citations.. In *AAAI workshop: Scholarly big data,* Vol. 15. 13.

[36] Ludo Waltman. 2016. A Review of the Literature on Citation Impact Indicators. *Journal of Informetrics* 10, 2 (2016), 365–391.

[37] Howard D White. 2004. Citation analysis and discourse analysis revisited. *Applied linguistics* 25, 1 (2004), 89–116.

[38] Lorna Wildgaard, Jesper W Schneider, and Birger Larsen. 2014. A Review of the Characteristics of 108 Author-level Bibliometric Indicators. *Scientometrics* 101 (2014), 125–158.

[39] Peter Willett. 2013. Readers' perceptions of authors' citation behaviour. *Journal of Documentation* 69, 1 (2013), 145–156.

[40] Han Xu, Eric Martin, and Ashesh Mahidadia. 2013. Using heterogeneous features for scientific citation classification. In *Proceedings of the 13th conference of the Pacific Association for Computational Linguistics.*