Text Entry Research - the last 5 years (2018-2022)

Andreas Komninos
Ioulia Simou*
akomninos@ceid.upatras.gr
simo@ceid.upatras.gr
University of Patras
Rio, Greece

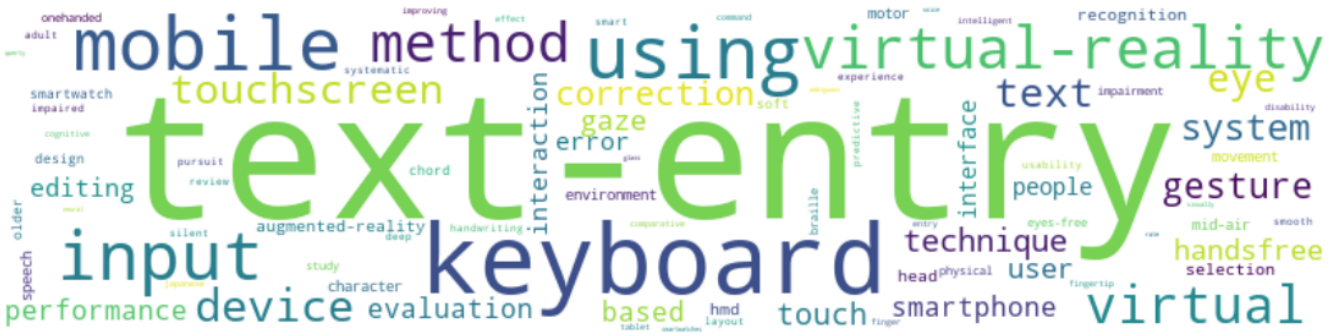

Figure 1: Wordcloud of the keywords in text entry publication titles between 2018-2022

ABSTRACT

In this paper, we present a bibliometric study of text entry research between 2018 and mid-2022. From this analysis, we identify the performance and relative challenges facing the text entry community, on the road to the next decade of research. We identify the community's low collaboration practices as a main barrier to progress, and propose several ideas on how to overcome this challenge.

CCS CONCEPTS

• **General and reference** → **Surveys and overviews**; • **Human-centered computing** → **Text input**; *Keyboards*; *Touch screens.*

KEYWORDS

text entry, bibliometrics

ACM Reference Format:
Andreas Komninos and Ioulia Simou. 2022. Text Entry Research - the last 5 years (2018-2022). In *TEXT2030: MobileHCI'22 Workshop on Shaping Text Entry Research in 2030, October 1, 2022, Vancouver, Canada.* ACM, New York, NY, USA, 7 pages.

1 INTRODUCTION

Text entry research underpins the most fundamental aspect of human-computer interaction, namely the ability to provide input to systems and applications in the form of text. This input can be used in order to provide instructions or operational parameters to various applications (e.g. entering a URL, an email address, responding to prompts), or to communicate with other humans and, increasingly, artificially intelligent agents. In their seminal book, Ian Scott MacKenzie and Kumiko Tanaka-Ishii describe text entry research in modern times as timely and of paramount importance, given the variety of devices, modalities and machine intelligent tools at our disposal [8]. The diversity of users in terms of physical and cognitive abilities, devices, situations, contexts and languages that characterises contemporary text entry practice and use, poses significant challenges for the research community.

There have been important advances in the commercial deployment of text entry methods (e.g. the use of federated learning alongside statistical modelling in Google's GBoard), which aim to provide a better user experience towards inviscid text entry (the point where user creativity, instead of the text entry method, becomes a performance bottleneck [5]). These advances are aimed to improve experience for mainstream use. Yet, the diversity of text entry usage contexts is such that many of these contexts remain relatively unexplored. New and emerging technologies and their integration in society (e.g. AR/VR, IoT) also introduce new contexts and challenges at a pace which is more rapid than the research community's ability to keep up and thoroughly investigate known contexts. Text entry research is difficult, but nevertheless will remain important for as long as humans need to interact with computers of any kind. Unfortunately, despite important technological innovations arising from research efforts, few of these advances seem to make it to the market, and therefore have a measurable impact in society.

Who are the brave researchers who engage in this type of research, and how are their efforts organised and disseminated? In this paper, we attempt a preliminary exploration of the text entry research community, through bibliographic data. From this, we aim

TEXT2030, October 1, 2022, Vancouver, Canada

to discover weaknesses and opportunities for the text entry research community, in an effort to guide practice in the next decade.

2 RELATED WORK

Bibliometric analyses are a common approach for monitoring and understanding the nuances of research activity in scientific disciplines [2]. A range of metrics, encompassing researcher performance (e.g. output counts, citation counts) and the state of disciplinary areas (e.g. co-citation, co-authorships, thematic and keyword analyses) can be used to obtain a view of a scientific discipline, negating the requirement to manually examine the voluminous related literature.

In the field of HCI, such studies have attempted to examine research practice in the discipline as a whole (e.g. [4]), within specific geographical constraints (e.g. [12, 13]), within specific outlets of HCI research (e.g. [3, 7, 9, 11]), and also in specific sub-fields (e.g. [1, 10, 14, 15]). One sub-field which has not received dedicated attention is text entry research.

In this paper, we present a preliminary analysis of the last 5 years in text entry research, using bibliometric data. Our aim is to capture the current state of research practice, and identify potential areas of improvement that might help to solidify the field in the coming decade.

3 DATA COLLECTION

We queried Google Scholar through a custom Python script, with the search term ["text entry" acm] for works with a publication year on or after 2018, excluding patents and including citations. The timeframe for our data collection therefore represents 4.5 years as the data was collected in May 2022. The query was formulated so as to discover publications that contained the bigram "text entry" and we added the term "acm" so as to capture publications either published by the ACM, or referencing publications published by the ACM. Google reported a total of approximately 3700 results, but in retrieval, the results were exhausted after 996 titles. During retrieval, we kept the publication title, year, first author name, citation count, publisher, publication URL and, where available, the URL to the publications's PDF.

We manually inspected the titles and publication excerpts produced by Google, and identified those which were likely to be relevant results (see Figure 2 for an example), excluding publications that were not in English. The total collection of relevant publications amounted to 460. For these publications, we manually corrected missing or erroneously collected data (e.g. author names, links, PDFs). Using the CrossRef API, we queried for every publication in the collection, using the title, first author name and year. This allowed us to harvest further metadata, including more accurate information about the publisher and year reported by Google, as well as new metadata including the complete author list, DOI, number of references, publication type (e.g. journal, proceedings) and outlet (e.g. name of conference or journal). Again, we manually inspected the collected data and corrected or filled in missing information, where possible.

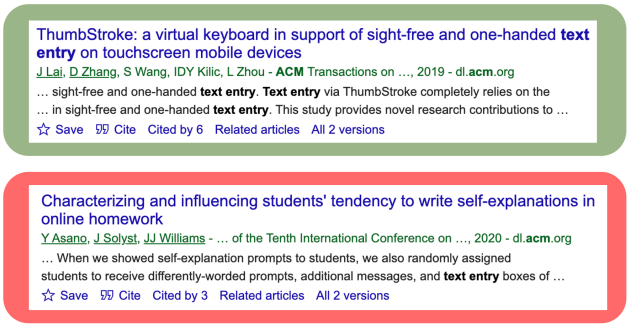

Figure 2: Example of relevant (top) and irrelevant results (bottom) from Google Scholar.

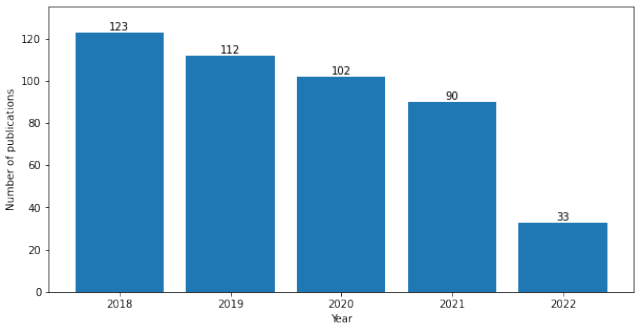

Figure 3: Text entry publications per year.

4 DATA ANALYSES

4.1 How much and where do we publish?

To begin, we examine the evolution of publication frequency across the timeframe of our data (Fig. 3. We note that there appears to be a steady decline in output, excluding of course 2022 which is only half a year. The majority of publications have publisher information (425). A total of 32 publishers are reported in our dataset, with five publishers sharing 89.87% of all published works. From these, the majority is held by the ACM (52.47%), while IEEE and Springer follow with significantly smaller percentages (Fig. 4). Further, we note that the majority of the 425 publications which have type information is published in conference proceedings (57.65%). This is followed by journal articles (29.18%), book chapters (11.29%), books (0.71%), posted content (0.47%), other (0.47%) and dissertation (0.24%). In combination, conference articles published by the ACM represent 43.29% of all retrieved results, therefore indicating the community's preference and orientation towards publication outlets. Another observation is that the text entry community represents only a small fraction of work published in HCI. Taking the ACM CHI conference as a field of comparison, 2,876 papers were published in its proceedings in the timeframe between 2018-2021 [1]. Out of these, only 69 papers in our dataset (2.4%) were amongst the accepted papers.

[1] https://sigchi.org/conferences/conference-history/chi/

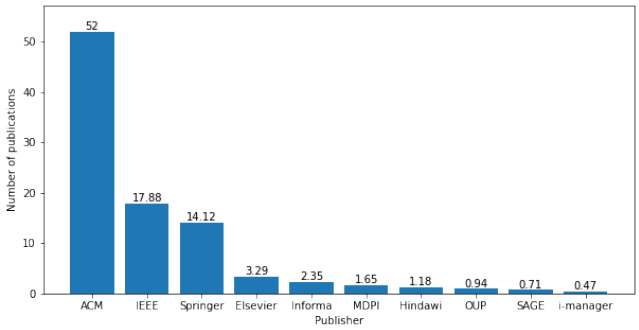

Figure 4: Percentage of publications shared by top 10 publishers.

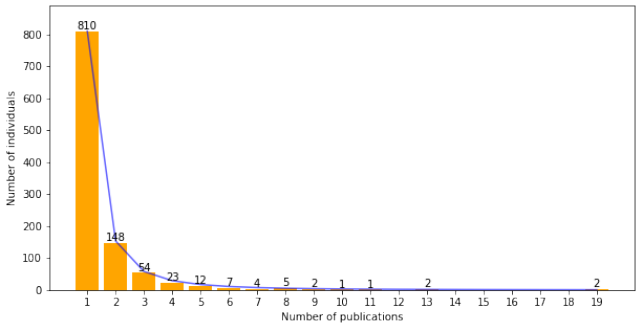

Figure 5: Number of individuals per publication frequency. The Lotka's law equation with $a = 2.4$ is shown in blue.

4.2 The text entry community

Overall, we find a total of 1071 distinct authors in our dataset. On average, each author is listed in 1.52 publications ($\sigma = 1.46$, $max = 19$). A total of 36 individuals are mentioned as co-authors in five or more publications, i.e. have at least one publication per year in text entry. This demonstrates that the size of the community which considers text entry research as one of their fundamental interest, and steadily pursues it, is very small. On the other hand, the number of authors who appear in two to four publications in our timeframe is quite significant (225 individuals). We might consider these as authors who are either at the start of their career, or authors who have some interest in text entry research, but it is not necessarily their primary (or only) research interest (Fig. 5). The community follows Lotka's law, which states that the number of authors making x contributions in a given period is a fraction of the number making a single contribution, following the formula $Y = 1/x^a$ where a nearly always equals two. In our case, the closest fit was for $a = 2.4$. Lotka's law has been empirically found to apply to the HCI community in general [6].

Further, we mapped the author names within publications to visualise the network of collaboration, as shown in Fig. 6. In this visualisation, authors are shown as nodes and are connected by an edge, if their names appear on at least one publication together (co-authors). The edge weight is relevant to the number of such co-authored publications, thus a stronger edge indicates more frequent co-authorship between two authors. The node size is relevant to the number of publications for each author. The average number of co-authors (graph degree) is 4.801, which drops to 3.54 for authors that have more than 3 publications in our dataset. The average path length is 5.608, demonstrating that the community is mostly disjoint and does not often collaborate outside a local "network". In contrast, for authors that have more than 3 publications, the average path length drops to 3.158, demonstrating a closer cooperation between these individuals and perhaps mobility of junior authors from one group, to another. The tight-knitting of the community clusters is also exemplified by the high average clustering coefficient of 0.822 for authors with more than 3 publications.

In Fig. 7 top, high clustering signifies authors with steady collaborations (low diversification in co-authorship). Low publication counts are a mixed bag of all states between individuals who always collaborate together, and individuals who always diversify. This diverges towards the extremes as publication counts rise, while those with the highest publication counts tend to converge towards the middle of the scale - this demonstrates the existence of a few group leaders with a hierarchy of long-term (other group leaders), mid-term (e.g. Ph.D. students) and short-term collaborations (other groups or Masters / Undergraduate students). Notably, the two authors in our set with the highest publication counts (19) have an almost identical clustering coefficient of ≈ 0.27. Eigencentrality is another metric of author influence - a higher score indicates co-authorship with other high-scoring authors. Authors with lower publication counts this time tend to cluster around low eigencentrality scores. As publication count rises, a few authors begin to exhibit co-authorship with more influential authors but many still seem to carry on with less influential collaborations. At the end of the publication count scale, we see the same two individuals also characterised by their relatively low clustering coefficient, but this time we note that these persons also seem to be collaborators between themselves and also other influential (but less published) authors.

Further, we examined the affiliations of authors as presented in each publication, where these were available from CrossRef. We geocoded the complete affiliation name using the Google Maps API, in order to obtain the relevant country. For each publication, we counted the number of distinct countries present in the author list. As an example, for a publication with 3 authors from, if the affiliations were (Organisation A, Country A), (Organisation B, Country A), (Organisation C, Country B), then we counted the publication as originating from countries Country A (1) and Country B (1). Overall we discovered affiliations present in 245 publications, spread across 38 countries. Figure 9 shows that the countries represented in the majority of publications are the USA (89), China (30), Japan (23), Germany, Great Britain, Canada and South Korea (22) and Finland (13). We discovered 76 publications with collaborating authors from 27 different countries, demonstrating that a respectable 31% of publications are international efforts (Fig. 8). The most frequent collaborating country pairs are shown in Table 1. A notable aspect from this examination is that researchers from the global South are completely under-represented, as are researchers from populations that use accented latin (e.g. French, Spanish, Portuguese) or other non-latin alphabets (e.g. Cyrillic, Greek, Hebrew).

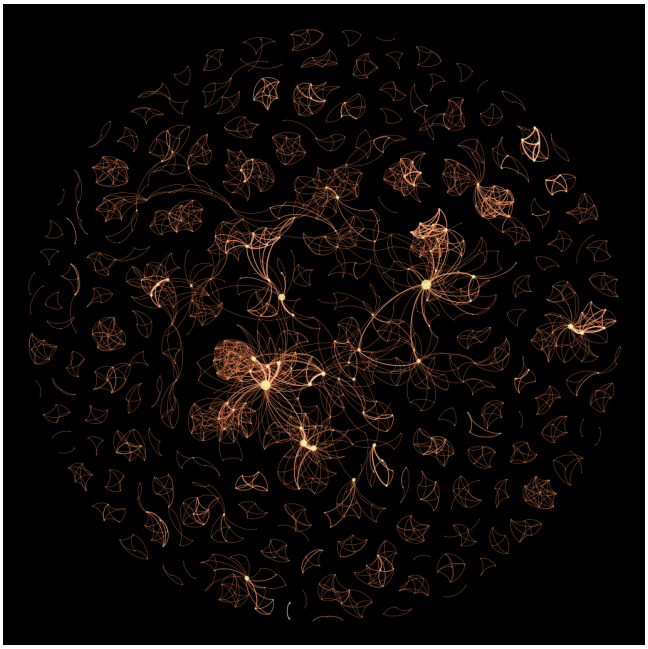

Figure 6: Network of author collaborations.

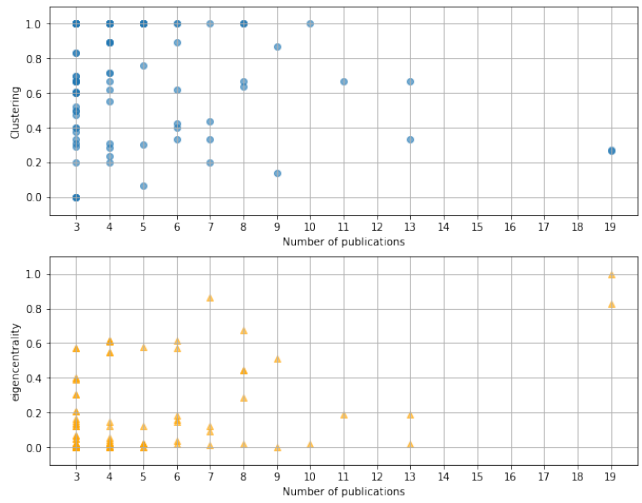

Figure 7: Clustering (top) and Eigencentrality (bottom) per number of publications.

4.3 Text Entry Topics

Finally, we took all publication titles and tokenised them, after removing English stopwords. We then manually processed the output, removing acronyms (e.g. 'KeyCube', 'MojiBoard') and replacing various synomous terms with a common representation (e.g. ['text entry', 'text input', 'typing']→'text-entry'; ['virtual reality', vr']→'virtual-reality'). We visualised these terms as a word cloud (Fig. 1) and their co-occurrence in a network diagram (Fig 10). This provides an impression of the current topics of research,

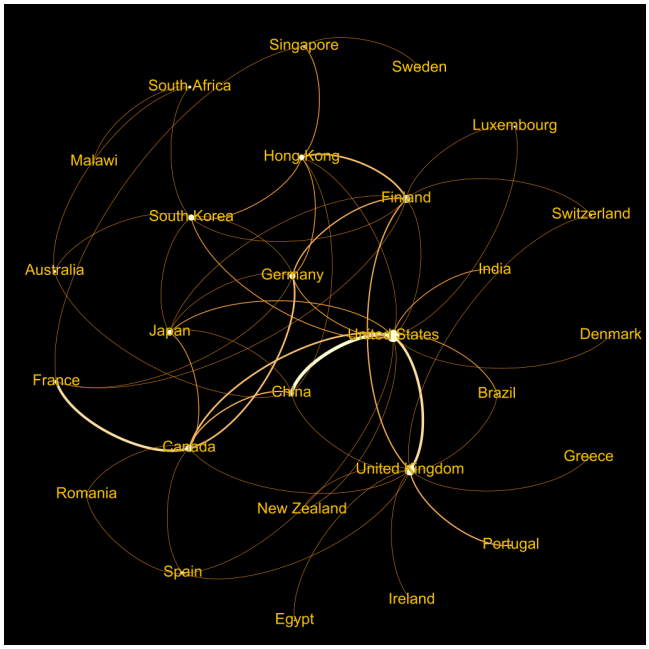

Figure 8: Collaborations across countries.

Table 1: Frequency of collaboration across countries

Country A	Country B	Frequency
China	United States	8
Canada	France	6
United States	United Kingdom	6
Canada	United States	4
Hong Kong	Finland	4
Germany	Canada	4
Germany	Finland	3
United Kingdom	Finland	3
Canada	China	3
Portugal	United Kingdom	3

where we note the under-representation of thematic areas such as accessibility, non-English language, and populations with specific characteristics. A strong focus on soft (virtual) keyboards in mobile devices and AR/VR is also seen, highlighting the lack of interest in physical keyboards.

5 DISCUSSION

Examining the bibliometric data for Nordic and Baltic HCI researchers, Sandnes makes several important observations which can serve as a basis for the discussion of our findings [13]. Firstly, as with most scientific disciplines, text entry follows Lotka's law, demonstrating that few researchers are making the most contributions in the field. We note that the level of collaboration between researchers is limited overall, but when considering these top-performing individuals, the picture is reversed. Since interest in the text entry field is found in many countries, these top-performing researchers

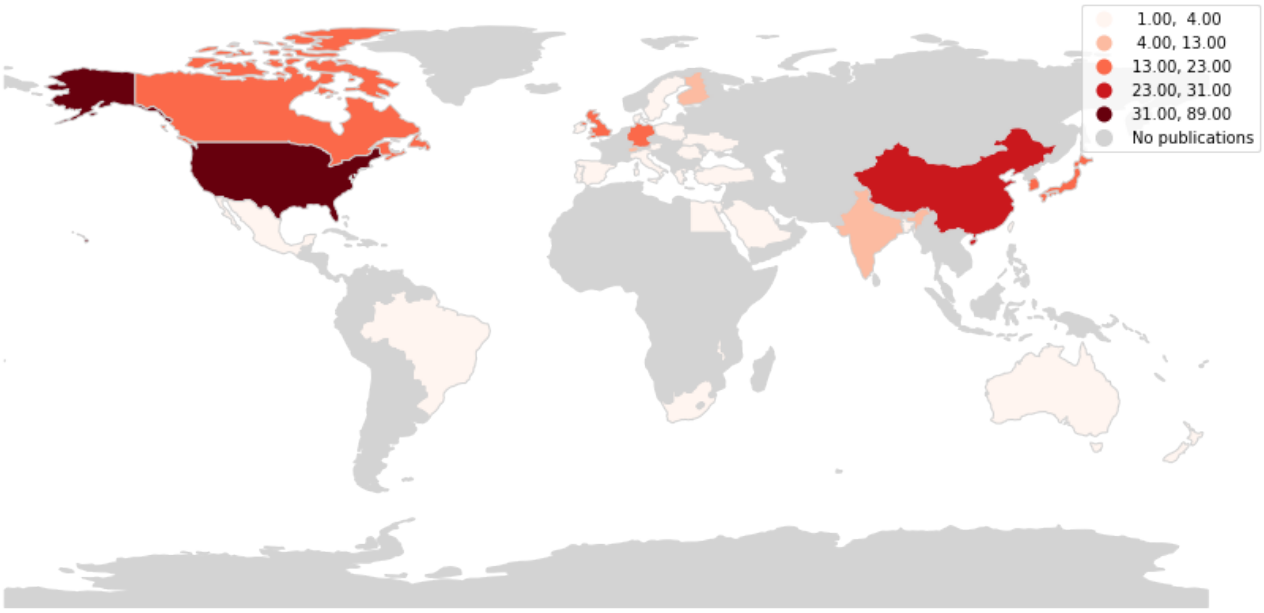

Figure 9: Publication frequency by author affiliation country.

might consider opening up collaborations with other first-time or inexperienced authors with diverse levels of experience and origins. Such collaborations might strengthen the field further, and allow pockets of excellence to emerge in other institutions. Collaboration is not just an act of solidarity towards the field. Rather, limited resources such as time, equipment, funding and access to experiment participants, pose significant challenges which may be overcome more easily. International collaboration might enable more ambitious projects addressing the needs of a global population, and enable research for which access to participants may be difficult (an example is the recruitment of low vision users). Further, new partners can become a gateway to access other transdisciplinary groups, who may provide new perspectives and opportunities for research (e.g. linguistics, psychology, or sociology).

How could the community foster higher levels of cooperation? We believe that the answer may not only be in the proactive establishment of wider collaboration networks (e.g. through direct approaches or open calls for expressions of interest). Instead, material and knowledge sharing practices may enable indirect collaborations, for example through study replications, re-examination or appropriation of open data to answer novel research questions. Software tools, such as a common directory of published research in text entry, interactive visualisations of authorships and collaborations, or a common repository for all data and materials related to the field, might allow better sharing of resources. The partial automation of a bibliometric analysis process such as described here, may also provide a useful tool for the continuous monitoring of community progress, towards the next decade.

Limitations are part of every publication and ours is no exception. We are certainly bound to have missed some of the publications out there since our search was limited to the Google Scholar engine and a simple query term. Some of our data is incomplete (e.g. author affiliations) or missing altogether (items without a DOI). Several aspects of data processing required manual intervention, which is also prone to omission or errors. Another interesting aspect which we were not able to investigate is the extent of open data and code practices, and also sample sizes which are important to assess the generalisability of findings. Yet, we believe that this investigation captures the overall state of text entry research in 2022.

6 CONCLUSIONS

In this paper, we analyse the state of text entry research in 2022, based on bibliometric data from Google Scholar. We find that the field could benefit from proactive engagement of leading scholars with other interested researchers across the globe, in order to seed new pockets of research excellence worldwide. More collaboration will enable projects of larger, even global, scope, and may help shed light in text entry for diverse populations. Openness of data and code may help indirectly in the development of the field, by encouraging authors to continue researching in this domain.

7 DATA AVAILABILITY

The data used in our paper as well as high-res versions of figures in this paper, are publicly available at https://github.com/komis1/text2030.

8 AUTHOR BIOGRAPHIES

Andreas Komninos is an Assistant Professor in the Department of Computer and Information Engineering, at the University of

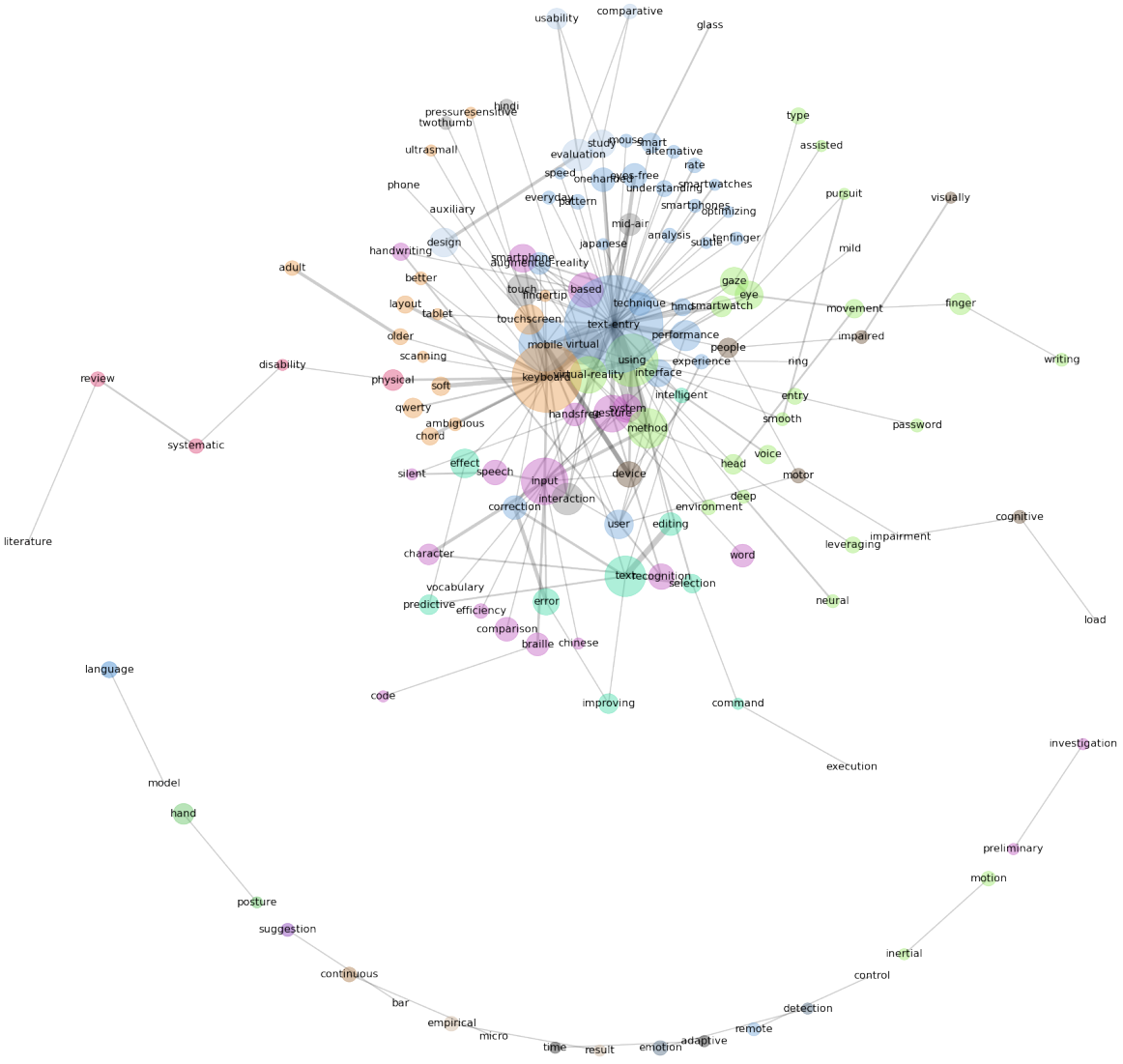

Figure 10: Co-occurrence of publication title keywords.

Patras, Greece. He has worked in the field of Mobile and Ubiquitous Computing since 2001.

Ioulia Simou is a postgraduate student in the Department of Computer and Information Engineering, at the University of Patras, Greece.

REFERENCES

[1] António Correia, Shoaib Jameel, Daniel Schneider, Benjamim Fonseca, and Hugo Paredes. 2019. The Effect of Scientific Collaboration on CSCW Research: A Scientometric Study. In *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 129–134. https://doi.org/10.1109/CSCWD.2019.8791855

[2] Naveen Donthu, Satish Kumar, Debmalya Mukherjee, Nitesh Pandey, and Weng Marc Lim. 2021. How to Conduct a Bibliometric Analysis: An Overview and Guidelines. *Journal of Business Research* 133 (Sept. 2021), 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

[3] Vassilis Kostakos. 2015. The Big Hole in HCI Research. *Interactions* 22, 2 (Feb. 2015), 48–51. https://doi.org/10.1145/2729103

[4] Konstantinos Koumaditis and Tajammal Hussain. 2017. Human Computer Interaction Research Through the Lens of a Bibliometric Analysis. In *Human-Computer*

Interaction. User Interface Design, Development and Multimodality (Lecture Notes in Computer Science), Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 23–37. https://doi.org/10.1007/978-3-319-58071-5_2

[5] Per Ola Kristensson and Keith Vertanen. 2014. The Inviscid Text Entry Rate and Its Application as a Grand Goal for Mobile Text Entry. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services (MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 335–338. https://doi.org/10.1145/2628363.2628405

[6] Suresh Kumar. 2014. Author Productivity in the Field of Human Computer Interaction (HCI) Research. *ALIS Vol.61(4) [December 2014]* (Dec. 2014). http://nopr.niscpr.res.in/handle/123456789/30335

[7] Yong Liu, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. 2014. CHI 1994-2013: Mapping Two Decades of Intellectual Progress through Co-Word Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3553–3562. https://doi.org/10.1145/2556288.2556969

[8] I. Scott MacKenzie and Kumiko Tanaka-Ishii. 2007. *Text Entry Systems: Mobility, Accessibility, Universality.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[9] Andrea Mannocci, Francesco Osborne, and Enrico Motta. 2019. The Evolution of IJHCS and CHI: A Quantitative Analysis. *International Journal of Human-Computer Studies* 131 (Nov. 2019), 23–40. https://doi.org/10.1016/j.ijhcs.2019.05.009

[10] Omar Mubin, Max Manalo, Muneeb Ahmad, and Mohammad Obaid. 2017. Scientometric Analysis of the HAI Conference. In *Proceedings of the 5th International Conference on Human Agent Interaction (HAI '17)*. Association for Computing Machinery, New York, NY, USA, 45–51. https://doi.org/10.1145/3125739.3125747

[11] Stefano Padilla, Thomas S. Methven, David W. Corne, and Mike J. Chantler. 2014. Hot Topics in CHI: Trend Maps for Visualising Research. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 815–824. https://doi.org/10.1145/2559206.2578867

[12] Daisuke Sakamoto. 2015. Asian Researchers at the CHI Conference. *Interactions* 22, 1 (Jan. 2015), 52–55. https://doi.org/10.1145/2692310

[13] Frode Eika Sandnes. 2021. A Bibliometric Study of Human–Computer Interaction Research Activity in the Nordic-Baltic Eight Countries. *Scientometrics* 126, 6 (June 2021), 4733–4767. https://doi.org/10.1007/s11192-021-03940-z

[14] Thiago Henrique Silva, Clayson S. F. de S. Celes, Vinicius Fernandes Soares Mota, and Antonio Alfredo Ferreira Loureiro. 2012. A Picture of Present Ubicomp Research Exploring Publications from Important Events in the Field. *Journal of Applied Computing Research* 2, 1 (Dec. 2012), 32–49. https://doi.org/10.4013/jacr.2012.21.04

[15] Lucy Lu Wang, Kelly Mack, Emma J McDonnell, Dhruv Jain, Leah Findlater, and Jon E. Froehlich. 2021. A Bibliometric Analysis of Citation Diversity in Accessibility and HCI Research. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3411763.3451618