IT RESEARCH - MINOR THESIS

Project: A Systematic Analysis of CHI Conferences

By Arunava Munshi (Student Id: 29453232)

Supervisor: Jarrod Knibbe

Master of Data Science, Monash University

Faculty of Information Technology

Semester 1, 2020

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the work of others has been acknowledged.

Signed by ......Arunava Munshi

Name ARUNAVA MUNSHI
Date 30/JUN/2020

ACKNOWLEDGEMENT

I am sincerely thankful to Dr. Jarrod Knibbe of the Faculty of Information Technology at Monash University and appreciate his contribution to my work. He has always been helpful whenever I was in trouble or had queries about the content of my research or thesis writing. During lockdown period, I received full cooperation from my supervisor even after office hours. While I continued working on this thesis, he guided me in the appropriate direction whenever he felt I would need it. His comments and valuable feedback on this thesis have continued to keep my motivation towards this work. I am also tankful to Dr. KASPER HORNBÆK for providing us the essential datasets for this research. I wish to show my sincere gratitude to the thesis advisors Pari Delir Haghighi and Reuben Kirkham for advising me about the grading criteria, plagiarism rules and best research practices. This accomplishment would not have been possible without their continuous support and encouragement throughout all the years of my coursework including researching and documenting this thesis. Lastly, I am thankful to all my family members and friends who have provided me unconditional mental support in my final year. Thank you.

Best Regards,
Arunava Munshi

TABLE OF CONTENTS

1. CHAPTER 1: INTRODUCTION

1.1 Abstract

HCI or Human Computer Interaction has been a diversified field (Nicola Dell, 2016), a growing area in research and an inter-discipline of computer science, behavioral/cognitive linguistic science and usability/human factors aimed at designing user friendly computer technology by improving the interactions between human being and computer systems. Though started exploration since 80s, HCI research has grown predominantly (Kostakos, 2014) since the commencement of 21st century because of the popularity of mobile devices, social networks and internet technology. However, with all these fame, CHI conferences came under criticism for not developing motor themes systematically (Kostakos, 2015), the heart of any discipline, and was under the question of whether HCI could qualify to become a scientific discipline. Our study, therefore, seeks to provide a complete systematic analysis of CHI proceedings happened between 1982 and 2018 to understand certain features of HCI research such as (1) how CHI conferences have emerged with new trends and techniques, (2) what are the shortfalls or gaps (if at all) in HCI research that remained unsolved over the years and (3) was there any opportunity created to broaden the field or to solve any gaps in terms of notable contributions or knowledge additions. In doing so, our work follows a unique approach to utilize the full text corpus of these proceedings to develop an unsupervised machine learning model, namely topic model, and visualize the generated topics in human interpretable forms in order to summarize the findings such as trends, contributions or gaps. Using this methodology, we observed that CHI conferences notably grew with new trends and techniques, broadening the field significantly for the past 37 years; however, at the same time, they also experienced significant shifts in research focus, leaving behind certain knowledge gaps. Finally, our work not only confines itself into a systematic study of the CHI proceedings but also offers a new method to study any discipline using data science, not biased by expert language or prior knowledge.

1.2 The Research Context and Background

1.2.1 Definition of HCI and its role in Modern Computing and Future Research

HCI (Dix, 2009) is a sub-field of computer science broadly combined with cognitive science, behavioral science and usability/human-factors that studies the way "computer technology" influences the activities of its users. This well-known terminology "computer technology" includes, but not limited to, mobile technology, car navigation systems, internet of things, automatic appliances used in households, embedded actuators & sensors etc. HCI has a predominant association with Interaction Design between a computer and its users, which is also popularly known as User-Centered Design, concentrated to design easy handling of computer technology. However, it is also important how the users feel about the computer systems as they start using them; this is also called "personalized view of the computer system", a key area of focus for HCI researchers. So, in this era of computing, the success of a software product in the consumer market highly depends on its user experience, a crucial part of HCI. Unfortunately, a great technological innovation on the backend is often discredited for a poor design of frontend interface, while an excellent interface often carries a product despite of its shortcomings inside.

Furthermore, it has never been a good idea to separate a product's interface from itself because, from a user's perspective, it is just one system, making the interface design a state of art in this modern era. Therefore, HCI encompasses user centric product design that literally includes everything starting from the concept development to the actual delivery of the product. Needless to say, HCI essentially swallowed up the whole application development lifecycle process. However, user-centered design has few elements that stand separate from the rest of the engineering process, needing particular skills areas in all phases of product development. This makes HCI skills strong in demand in industry and according to the computer science division of U.C. Berkeley, HCI, while getting the most attention of industry experts, is one of most important fields of research in computer science and has potential scope for future research and innovations (Canny, 2006).

1.2.2 Gaps in HCI Research Over Time

For years HCI Conferences have significantly rich history and outstanding reputation. As on 2014, 3152 publications released within a gap of 20 years that had shaped HCI as a strong multidisciplinary avenue (Kostakos, 2014). However, fame also comes with criticism, so is the case for HCI. In 2015 CHI conference, a paper named "The Big Hole in HCI Research" (Kostakos, 2015) made a claim that HCI research systematically lacked "Motor Themes", also recognized as the heart and soul of any discipline, creating a big hole of knowledge into the HCI research. According to the researcher, the scarcity of these motor themes is a very worrying aspect of any scientific discipline, so is the case for HCI and he raised a question towards the credibility of HCI to become a scientific field. The research showed that CHI conferences did not put enough efforts to develop certain mainstream topics, rather they just rolled new topics year after year, making enough room for the so-called knowledge gaps. A similar research from the same group of researchers published an analysis on CHI conferences from 1994 to 2013 (Kostakos, 2014) and their observation was – HCI grew considerably higher in the period 2004 – 2013 than the period 1994 – 2003. This observed growth was because of the introduction of more and more new topics during 2004 - 2013. But at the same time the researchers also observed that almost half of the topics, which were introduced during 1994 – 2003, surprisingly disappeared during 2004 – 2013. Furthermore, the researchers also plotted the popular topics of these two eras on a two-dimensional plane against the metrics called destiny and centrality. The observation was that there was almost no topic with the feature of high destiny and high centrality, the key feature of the motor themes. These two observations led the researchers to conclude that HCI lacks the driving themes, which are essential for the discipline to grow with knowledge to be potentially considered as a strong scientific field.

1.2.3 Our Research approach and its Significance

"The Big Hole in HCI Research" (Kostakos, 2015) was based on "Co-Word Analysis" (Michael, 1983), which essentially considered the keywords appearing together on research papers and the variance of their interrelations over time. In that paper the researcher made a very strong claim that Co-Word analysis potentially maps the inherent knowledge of a discipline with the consideration of the linkage between the concepts. Co-word analysis has also been a proven technique in other fields such as psychology (Marc, 1991), software engineering (Coulter N. S., 1998) and stem cell research (An X. a., 2011). However, none of those studies considered full-text corpus of the research papers in order to map knowledge of a scientific field. Furthermore, the whole analysis was also based on an assumption that keywords at the top of the papers constitute an adequate description of their content. While this could be an adequate assumption, at the same time, the reliability of the research outcome is highly reliant on the accuracy of the keywords used to represent those papers. Finally, because of the nature of research, Kostakos and his team could only use 16035 keywords from 3152 CHI papers, making this knowledge mapping based only on an average of 5.09 words per paper.

In our view, the knowledge related to any scientific field is reliant on its published papers or conference proceedings, while the specific knowledge of any individual paper, in turn, is highly concentrated within its whole content, which is nothing but the full-text corpus of the paper and we observed that, even though there has been many research done on HCI (discussed in detail in literature review section) to understand its trends and gaps, none of them actually took the opportunity to do it by exploring the full-text corpus. There is only one research paper (KASPER HORNBÆK, 2019) (which also inspired our work and is our source of data) that explored full-text corpus of CHI conferences, but it was only focused to find a very specific trend in HCI research such as the meaning of the word "interaction" in HCI. However, we could not find an existing study that explored the full-text of CHI proceedings to do the complete trend and gap analysis of CHI conferences.

Our work, which is quite unique in this front, creates an opportunity to explore the full-text corpus of 37 years of CHI proceedings to observe the trends and potential gaps appeared over time. We are also interested to see whether we can map knowledge of any discipline (in this case, it is HCI) with its whole research content produced over time that, in this case, essentially is the full-text corpus from the CHI proceedings. Please note that the purpose of our research is not to understand how meticulous HCI conferences were in developing knowledge, nor are we interested to know if HCI has emerged as a healthy discipline over the years. But we are offering a completely new way to do the systematic trends and gaps analysis using an unsupervised ML technique called topic model derived from the underlined text corpus and we would furthermore see an opportunity for this method to be used in doing the field studies for other disciplines.

This research follows a very streamlined approach. Firstly, we have used the data (the full-text corpus of CHI proceedings) prepared by KASPER HORNBÆK to make them in specific format using a few "Natural Language Processing" (NLP) techniques; we have essentially used '*.json' file format to store the formatted proceedings. Then we applied "Latent Dirichlet Allocation (LDA)" (David M. Blei, 2003), also considered a "Generative Probabilistic Approach" used to model collections of discrete text corpora, in order to build the topic models. After building the topics, we used LDAvis technique (Carson Sievert, 2014), a combination of R and D3 based interactive visualization technique using LDA in the backend, to visualize the relevant topics generated from the text corpus into a two dimensional topic space. This topic space became the heart of our research as we use it as a baseline to discuss our findings. Based on these visualizations, we classified each topic under one or more pre-defined 16 HCI research areas (Biplab Ketan Chakraborty, 2017) in order to generate insights. The specialty of this method is that it gave us an opportunity to map the raw text corpus into one or more topics using unsupervised machine learning technique, called LDA and can visualize these topics within a topic space. The gap between two topics are essentially the distance between them in terms of how distinct they are from each other based on their underlined content. So essentially, this method gives us a notion about how the knowledge of an underlying content is scattered into a two-dimensional space, which in turn gives us an idea of the gaps or holes in knowledge within its backend material.

1.3 The Research Questions

Our study is aimed to develop a holistic view of the CHI proceedings spread across 37 years, including their successes and pain areas. We would wish to discuss the success of these proceedings in terms of the evolution of this discipline with its emergence in knowledge, trends, topics, notable works, developments and its contributions. On the other hand, we will also discuss the pain areas of the same CHI proceedings in terms of holes or gaps of knowledge produced over time and how well they were addressed. We seek to answer the below questions in order to explain our research outcome.

1.3.1 What did CHI conferences offer between 1982 and 2018?

The scope of this question is to answer the trends in CHI conferences from 1982 to 2018 in terms of their trends, topics, notable works, developments and contributions. In order to do this, we divided this span of 37 years into 4 consecutive decades (also called topic eras below) such as 1982 – 1991, 1992 – 2000, 2001 – 2010 and 2011 – 2018. Now, for each decade, we will answer the below sub-questions.

1.3.1.1 What were the trends and techniques in CHI conferences during 1982 - 2018?

The scope of this question is only to understand the evolution of CHI proceedings for the past 37 years. We would do this by firstly identifying the major topics discussed during this time frame for each of the four decades or topic eras. Then we will classify the prevalence of these topics towards 16 major pre-defined HCI research areas (discussed in 3.6.1 in detail) such as UI/UX designs, human factors, cognitive science, usability studies, interaction design etc. By this, we would like to understand which areas of HCI research have been discussed so far and which have not been under discussion, which areas were popular and during what time period, which areas were not so popular during the initial days but received attention in later years, whether there are similar or dissimilar trends during two consecutive decades etc.

1.3.1.2 What were the notable contributions and addition of knowledge in CHI conferences happened during 1982 - 2018?

The second inseparable part is to find the notable contributions happened in HCI discipline during the same time period. By "Notable Contributions" in our research, we wish to showcase those topics or areas of HCI research that have received enough attention & developments and became sufficiently matured over the years. As part of this task, we are interested to observe if we can find one or more HCI research areas that sufficiently accumulated knowledge and subsequently grew over the four decades and if they are, we would showcase this contribution in terms of a defined metric (discussed in 3.6.2) in order to understand the amount of work those research areas have presented to broaden HCI as a scientific discipline.

1.3.1.2.1 Can we offer a quantitative metric to calculate the addition or loss of knowledge within a corpus?

The sub-part of the previous question is to deep dive further in respect to the contributions happened in CHI conferences. In answering the previous question, we will showcase the relative prevalence of each topic within 16 major HCI research areas (discussed in 3.6.1). However, that does not provide a clear picture of what is the exact amount of knowledge that the topic added into HCI research spectrum. One way to estimate this is by quantifying the amount of information that has been gained or lost due to the addition of the topic to or the subtraction of the topic from the entire topic system. As part of this work, we would try to evaluate if we can offer any quantitative metric to estimate this amount of information gain or loss. Please note that we have not aimed to implement this metric as part our study; this will be propagated as part of our future work.

1.3.2 What were the holes or gaps in knowledge produced in CHI conferences during 1982 - 2018?

This part of this research will try to focus into the gaps or downfalls in the CHI conferences from 1982 – 2018. However, our gap analysis, as opposed to the study done by Kostakos, will be performed using the topic modeling on the full-text corpus of the CHI proceedings, not just on the "Co-words". In answering this question, we seek to build the topic visualization models for four consecutive decades that map the topics onto a two-dimensional plane. Then we would examine the gaps between the topics, which are essentially the knowledge gaps, based on their relative distances (defined under certain probabilistic metric, called relevance metric) on that two-dimensional plane. We would, further seek to find out the trends of these gaps over time based on whether the topics become closer to or distant from each other. We would also like to examine the trends of the popular terms over the decades. In doing so, we would be investigating whether the popular terms from CHI conferences were carried forward from one decade to another, or they disappeared, a good indication of whether possible knowledge gaps persisted or not. Once these are examined, we will do some deeper level of gap analysis by answering the below questions.

1.3.2.1 What were the problems in CHI conferences left unsolved over time and the possible reasons behind?

This is a critical part of our gap analysis. Here our aim is to understand if there are any deprived research areas or topics in CHI conferences that systematically lost attention year after year. We would further study if we can find any research topics which were in trend in one or more decades but have lost their focus in the subsequent decades. We would further try to understand the amount of attentions these research areas or topics have lost through their loss of prevalence over the decades based on our defined prevalence metric below (discussed in 3.6.2). We would then explain any possible reason behind this situation.

1.3.2.2 Was there any attempt made to address any persistent gaps?

Our final query to answer, as part of this study, is to know if there is any novel attempt made to fill any persistent knowledge gaps. For that, we would like to see if we could find any research areas or topics that were prevalent in one decade but lost their focus in another decade. However, if the same research area gained their predominant attention again, then we can conclude such situation as an addressal to the gaps. If, we could also see any research areas that were deprived in the initial years, but gained significant attention in later years, then it will also be another example of addressing research gaps in CHI proceeding. With this, we present our relevant literatures, research methods, study observations and outcomes in the following chapters.

2. CHAPTER 2: LITERATURE REVIEW

2.1 Scope and Methodology

Every research must be informed by the existing works within its subject area. A literature review (Rowley, 2004) seeks to identify, organize and review the key concepts of the related research works done so far and tries to justify the ongoing work in light of these previous works. Literature review may be performed under various occasions, such as, giving a background for the ongoing research work, learning the current research based on similar topics of interest, answering the base research questions in light of the existing works etc. (Okoli, 2010). As per the scope of this literature review, (1) we are focused to explore the research works done so far to observe the evolution and the developments in HCI and its emerging trends over the years. Furthermore, (2) we will also review the existing works done so far in order to identify the shortfalls or gaps in HCI research. Finally, (3) we would be interested to look at any existing works done to address any persistent gaps.

Because we are offering a new approach to do a systematic analysis of HCI research, our aim is not to understand the outcome of these above-mentioned literatures as we will present our own research outcomes in the end, but our focus is to understand the approaches these works followed to gain those outcomes. For example, if we are reviewing a literary work on HCI trends and techniques, then we would be more interested to know how the author is able to observe these trends, rather what those observations are. As part of our method, we traced back the citations or bibliographies from a few notable works that studied CHI conferences (Hu, 2009; Kostakos, 2015; Reeves, 2015) and from these works, we backtracked further. The scope of our literature review is to analyze the papers in the below areas.

1. Literatures observing CHI trends and techniques
2. Literatures observing gaps in HCI research
3. Literatures to address the gaps or unresolved problems

2.2 Literatures Observing CHI Trends and Techniques

There has been a handful of research works done to find HCI trends and techniques. However, most of the works followed three distinct techniques to do the research, which are – 1) Manual revisions of literatures, 2) Studying CHI Publications and 3) Bibliometric Analysis.

2.2.1 Manual Revisions of Literatures

In the absence of any automatic approach in earlier days, many literatures attempted to find CHI trends by manually reviewing the papers. While this may sound a bit unmethodical, these research approaches offered few nice and comprehensive ways to find the HCI trends. One study named “The five foci of interface development” (Grudin, 1990) tried to review the HCI trends by examining its chronological evolutions and could come up with five different stages of HCI developments named as “five foci”. The approach here was to methodically separate the boundaries of inventions happened in HCI and review the notable works within that sphere. One more literature tried to find the chronological developments in HCI (Myers, 1998) by exploring two overlaid research dimensions - 1) Various HCI interaction mediums and 2) works done for each of these mediums in different types of institutions. The interactions mediums have been subdivided further and the related works from institutions like universities, corporate industries etc. on those subcategories of mediums were reviewed to understand the chronology of CHI research. Finally, we reviewed another study that attempted to explore CHI trends (Grudin, 2005) through the identification of different HCI threads from the cultural, historical and conceptual contexts and identified how the meaning these contexts changed over the period of time.

2.2.2 Studying CHI Publications

Analysis of publications was another way to find the HCI trends and techniques. A research approach analyzed the publications (Barkhuus, 2007) from 24 years of CHI proceedings in order to observe the evolutions of the discipline. This methodology considered picking samples of papers within a gap of 5 to 6 years and broadly classified them into 1) Empirical and Analytic category or b) Quantitative and Qualitative category. There was a nice overlap found between the quantitative and empirical categories, while the same overlap could also be observed between qualitative and analytical categories, using which the researchers could nicely identify the CHI trends.

2.2.3 Bibliometric Analysis

Another methodology to see the CHI trends were “Bibliometric Analysis”. Using this technique, a quantitative method (Hu, 2009) offered an analysis on the countries and institutes that consistently produced high quality materials for HCI conferences for as many as 26 years. The bibliometric analysis was applied on every publication in order to understand the systematic growth of Human Computer Interaction as an independent discipline of computer science during the same period. A key assumption of the study was that the quality of research has a vital role to play towards its success along with the country and institution where from it has been originated. This study invented a ranking system named “h-index Ranking” that helped to rank these institutes and countries based on their key contributions to CHI conferences. With this ranking method, the historical significance of different empirical studies was also presented.

2.2.4 Gaps in Existing Research

We presented few approaches to find the CHI trends. Each of these approaches is unique from its perspective and has its individual advantages. While all these methods could collectively discover different HCI themes chronologically appeared over time, they suffered from some serious disadvantages. Narrative approach was based on the evaluation of selected papers of the given timeframes, while analysis of publications was dependent on choosing paper samples. Any biasness in these methods of selections could potentially risk the quality of research outcomes and might lead in missing a trend or theme completely. Bibliometric analysis, by considering all research papers within a given time period, posed lower risk of missing any potential theme, but this approach did not utilize the full-text corpus of these research papers, thus risking information loss while trying to build those themes. Our analysis, on contrary, does not pick samples and uses the full-text corpus of all research papers under consideration, so it does not pose the risk of losing a theme entirely and possess very low probability in losing information during topic or theme building process.

2.3 Literatures Observing gaps in HCI research

We would discuss three main approaches used to observe the pitfalls in HCI research, which are 1) Surveying the CHI Taxonomies, 2) Co-Word Analysis and 3) The analysis of Citation Networks.

2.3.1 Surveying the CHI Taxonomies

A study (Quinn, 2011) observed CHI taxonomy survey to dig out the knowledge gaps in HCI research happened over the years. Initially, the main motto of this paper was to understand where the knowledge overlaps of other disciplines with HCI are. These disciplines included data mining, crowd-sourcing and social computing. Once these overlaps are confirmed, the study attempted to generate some features of HCI, called “Taxonomies”, closely lined with these disciplines. These taxonomies could be broadly classified as “Aggregation”, “Motivation” and “Human Skills”, which were divided further into low-level features, also called “Classification”. Finally, each taxonomy is systematically studied and compared with the others to understand the gaps among them, which are nothing but considered the knowledge gaps in HCI research.

2.3.2 Co-Word Analysis

We have already discussed Co-Word analysis, a widely proven techique (Callon, 1983; Coulter N. S., 1998; An X. a., 2011) to map knowledge of a scientific field by drawing the relationships among the topics or themes. Kostakos (2015), in his study, used this method to observe the gaps in CHI proceedings by understanding the gradual change in relationships among these keywords over time based on their co-appearances on research papers. The study defined the ‘Themes’ or ‘Topics’ as the cluster of these keywords and based on that, it sought to estimate two metrics of the topics called “Centrality” and “Density”. A theme with high centrality and density has been attrubuted as “Motor Theme”, which is also the heart and soul of a discipline. Based on the analysis, CHI themes, starting with lower centrality, progressed towards topic communities in order to become more central. As they matured, they attemepted to proceed towards the heart of the discipline, but, through the abrupt loss of centrality, they could not meet that milestone. Hence, most of the themes could never become “Motor Themes”. So the study concluded that CHI conferences experienced failure in systematic development of motor themes, a really worrying prospect of HCI as a scientific discipline.

2.3.3 The Analysis of Citation Networks

A method named “Citation Network Analysis” (Clement Lee, 2019) on the publications could also help us to find out the shortfalls of HCI research. This method initially got access to the text references and referenced research papers within a library where the proceedings were kept. A “Citation Network Graph” was then generated that was linked across the edges of citations within the network graph in order to identify the related “Topic communities”. Furthermore, using “Community Detection” algorithm, a clustering technique, all papers were put into separate clusters. Finally, by systematically studying generated clusters, it was concluded that CHI proceedings emerged (though infrequently) with new research topics; however, they failed to develop appropriate themes, leading to the gaps in HCI research.

2.3.4 Gaps in Existing Research

We observed different methods to find gaps or shortfalls in CHI conferences with their individual uniqueness and advantages. However, these techniques suffer from certain disadvantages. CHI Taxonomy Survey built the properties or taxonomies of HCI discipline based on some other disciplines, a technique which is at risk in doing potential misclassification of these properties because the general assumption here is that HCI is closely related to the other disciplines for this classification of properties, which may not be the case. Co-word analysis, on the other hand, works only with the keywords appearing together on different papers, which is why it never considers full-text corpus (study says that the method considered only 16035 keywords from a list of total 3152 papers with an assumption that the keywords can sufficiently capture the context of a research paper) to build the themes and hence, the technique might suffer from potential loss of information. Citation Network Analysis also has the shortcomings that are similar to the disadvantages of the Co-word analysis. Our method, on the other hand, utilizes full-text corpus in topic modeling, not just the keywords. Thus, our technique offers an opportunity to take a much deeper look into the topics and their interrelations with respect to the underlying whole content, providing a much broader context to examine the knowledge gaps in CHI conferences appeared over the years.

2.4 Literatures to Address the Gaps or Unresolved Problems

Finally, once the knowledge holes of CHI proceedings have been identified, it is the high time to understand if any such persistent gaps have adequately been addressed. In order to examine that, we have found two approaches that tried to address these holes, which are 1) Reference Task Adoption and 2) Making counter reasons against the validity of Big Hole Theory.

2.4.1 Reference Task Adoption

A research (Whittaker, 2000) understood the deep-rooted problems in CHI conferences, which are nothing but their overemphasis in making “Radical Invention” and lower focus into the mainstream topics, considered to be the main reasons for the pitfalls in HCI research. The solution given is the adoption of ‘Reference Tasks’ in order to build those themes. These tasks proposed changes in the ways HCI researches have been conducted and the CHI community has worked. The study proposed that the technical research in HCI discipline should focus more into utilizing crucial user tasks, operationalizing these reference tasks and doing field trials. Furthermore, the study conveyed that the research community must also consider those ‘Reference Tasks’ the “Idea Work”.

2.4.2 Making Counter Reasons against the Validity of Big Hole Theory

A literature (Reeves, 2015) directly questioned the validity of “The Big Hole in HCI research” theory (Kostakos, 2015) where HCI has been overemphasized as a scientific discipline. The study firstly has put its concerns that scientists have excessively attempted to prove HCI as a scientific discipline, whereas, as per its early orientation, HCI should be an interdisciplinary field of psychology and cognitive science and hence, finding the gaps in HCI would be truly unjustified based on some predefined set of attributes of scientific disciplines considered to be the ultimate “Gold-Standards”. Furthermore, the study also reminds that HCI possesses its own “phenomenal particulars”, as does every other field. Therefore, the comparison of other fields such as Software Engineering, Stem Cells etc. with HCI and the applications of their respective techniques to find gaps in CHI conferences might be irrelevant and unjustified.

2.4.3 Gaps in Existing Research

We found two methods addressing the gaps in HCI research. However, these approaches have some serious disadvantages. Reference task adoption is more of a prevention technique that finds a way to prevent any research from being systematically deviated from developing “Motor Themes”. However, the technique never tried to identify the gaps itself. The counter reasoning theory argued the validity of the research method in the “The Big Hole in HCI research” (2015) and questioned the traditional view of research communities towards HCI as a scientific discipline. The technique also questioned the overemphasis of the scientific communities to believe HCI as a scientific discipline and the irrelevant comparison between HCI and other fields like Software Engineering, Stem Cells etc. in order to find its gaps. In that way, the study found enough grounds to counter the validity of co-word analysis from Kostakos research team that overly tried to fit HCI into scientific discipline. However, this direct contradiction does not necessarily disregard the traditional and long-standing views of the scientific communities. So, the big hole theory does not necessarily become falsified with this approach. On contrary, our method tries to apply sophisticated machine learning technique to actually find the gaps in HCI generated over the years by trying to project different CHI topics on a two-dimensional plane and evaluate their relationships. So, our method is more of a direct approach to find the actual gaps rather than falsifying any existing technique.

3. CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

This chapter presents our research methods in detail. As discussed earlier, in this research, we aimed to do the key trends and gaps analysis of all CHI proceedings between 1982 and 2018. In order to do so, we sought to follow below six systematic steps to achieve our research outcomes. Firstly, we have introduced our methodology with the brief description of our dataset and tools and technologies used. Then we followed a list of data pre-processing steps to prepare the data for topic modeling. After the data preparation, we sought to choose our topic modeling technique, which is LDA in our case, and the appropriate topic visualization technique, which is LDAvis technique. Once these techniques are finalized, we would be interested to choose any appropriate model parameters for the purpose of better topic interpretations. Finally, we are all set to actually build those models and interpret them in a human understandable way, which we have done in our final step of our research method.

- Description of Dataset and Tools & Technologies used
- Data Pre-Processing
- Choosing Appropriate Topic Modeling Technique
- Choosing Appropriate Topic Visualization Technique
- Setting Up Model Parameters

- Topic Modeling, Visualization and Generating Insights

3.1 Description of Dataset and Tools & Technologies Used

We set to begin our research methodology with a brief description of our dataset. We have used the full-text corpus of 37 years of CHI proceedings prepared by KASPER HORNBÆK (2019) for his research. The data was provided in '*.pkl' file format for each year; so, we received 37 of such files for the proceedings between 1982 and 2018. Each file is a Python list of all published papers in CHI proceeding for the respective year and each element of these Python lists is in form of a dictionary where details of each paper has been captured, which includes publication year, raw paper text, paper title and paper DOI. As part of this study, we have only included publication year and raw paper text for our analysis.

Now, because the data have been provided in Python list format, we have decided to use Python 3.8 as our programming platform. We have used NLTK Python package to perform all data cleansing and text processing tasks (we discussed in the next section) as part of our data pre-processing part. Then we have used Scikit Learn package in Python, a popular library for Machine Learning, for our LDA Clustering (discussed later) purpose, our main method of topic modeling. Finally, we have used pyLDAvis Python visualization package as our main library for topic visualization. Additionally, for faster data processing, we have used Python parallel processing technique embedded through multiprocessing package. Our topic visualizations were done in simple HTML files in an interactive way, which provides easy handling and better interpretability. Furthermore, we have used MS Excel for collecting and managing insights from these HTML files and created more graphs and charts (provided in the next chapter) to interpret our findings. Finally, we have used Python "wordcloud" package to show word-clouds.

3.2 Data Pre-Processing

Pre-processing of data is the initial, but vital, step for any data modeling technique. A good pre-processing leads to a better data modeling technique. In this case, we have got the complete text corpus the CHI proceedings between 1982 and 2018. So, our data pre-processing is basically doing text pre-processing, which we have done using sophisticated Natural Language Processing Techniques (K. Nigam, 2000).

3.2.1 Identification of Topic Era

Before we did any text pre-processing, we had decided to show the research trends in CHI Proceedings over four consecutive decades spanning from 1982 – 2018. Because we have the data for total 37 years, we have decided to divide the decades in the following way (could not be divided in evenly manner) and have named each decade a "Topic Era" –

- Decade 1: 1982 – 1991
- Decade 2: 1992 – 2000
- Decade 3: 2001 – 2010
- Decade 4: 2011 – 2018

As we have the data for each proceeding year in a separate '*.pkl' file format, we have merged those files of respective proceedings years into respective one file for each topic era. For example, the proceedings between 1982-1991 have been merged into one '*.pkl' file and so on for the other decades. So finally, we have got four big '*.pkl' files of each of the above decades or topic eras. Now for each of these '*.pkl' files we have followed the following text pre-processing steps.

3.2.2 Removal of Enter, Tabs and Extra Spaces

At the very 1st stage of data pre-processing, we considered removal of Enter, Tab and any extra Spaces as they are not considered to have any semantic value within a corpus. On the other side, if we had kept them, they would have been used in our machine learning algorithm (discussed later) to train the model, not only making it more complex, but also challenging its interpretability. However, at this stage we have decided to keep at least one space between two consecutive word tokens to separate them based on that space in the later stage.

3.2.3 Removal of Punctuations

Our next step was to remove the punctuations as they don't bear any lexical significance within a corpus as well as in our data modeling. However, there are two special punctuation types called "Hyphen" and "Apostrophe" that may be used for various purposes. However, every situation of "Hyphen" and "Apostrophe" could be handled by just replacing them with a null character

except for the cases of "word contractions" because, in the later stage, we have removed the stop-words and, in order to do that, the word contractions should be converted into appropriate stop-word pairs. So, at this stage we have removed all other punctuations except Apostrophes and then we handled Apostrophes separately after word contractions have been taken care of.

3.2.4 Case Normalization

A word token can contain a letter in uppercase or in lowercase depending on the type of word. A noun usually starts with an uppercase character and so as the case for a commencing word in a sentence. Capitalization is useful in a text document from a reader's perspective because it helps a reader to differentiate between a proper and a common noun. However, in most of the cases uppercase words have no different lexical value within a corpus. For example, the 'Data/Einstein' has no different meaning from the word 'data/einstein'. So, converting uppercases into lowercases is considered a common practice for simplicity, so we did the same as part of our data pre-processing.

3.2.5 Expand Word Contractions and Apostrophe Removal

Continuing from punctuation removal, the next step after case normalization is the expansion of contracted words such as "don't" and "we'll' into proper forms such as "do not" and "we will" so that they can be perfectly removed during the stop-word removal process (discussed later). After the contracted words have been expanded, we could finally remove Apostrophe, which was left in the punctuation removal step. In order to handle word contractions, we have used a pre-defined set of word contractions from Stack Overflow (Stackoverflow, 2020) frequently used in English text. Whenever, we encountered any of these pre-defined contractions we replaced them with corresponding word pairs. After the word contractions have been handled, we finally had the chance to remove Apostrophes.

3.2.6 Word Tokenization and Digit Removal

All the above data pre-processing steps we followed so far were on the removal of special characters. Now, we move on to the next step of actual word processing. In NLP, words are considered tokens, which we can also call the backbone of the entire system. So, we separated the word tokens from each paper into a separate Python list, also called "word tokenization". After the tokenization step, we had list of the list of words where the inner list represented a single proceeding and the outer list corresponded to accumulation of all papers for a particular topic era e.g. Decade 1: 1982 – 1991. After the word tokenization, our next step was to remove numeric values from each list, as they don't provide much significance in modeling topics.

3.2.7 Stop-Word Removal

Stop-words are generally extremely common in any English text and they barely carry any lexical value within the text corpus. So, it is a common practice in language processing to remove such words in order to better use of main memory and speed up the text processing. Some examples of stop words are 'is', 'are', 'the', 'not' etc. NLTK package provides a list of stop-words itself, which we have used for this purpose.

3.2.8 Word Lemmatization

The next step of our text pre-processing was to synchronize various word forms within the corpus. A word can be in different forms within a corpus; e.g. a word 'education' may have various forms such as 'educate', 'educating', 'educates', 'educated' etc., but they barely convey any different linguistic properties. We have two well defined processes named "Stemming" and "Lemmatization" (Retrieval, 2009) for doing the grouping of the same words in different forms. In Stemming process, the prefixes, suffixes and word pluralization are identified and then removed using a heuristic process, leaving only the stem of the word. But this process seriously suffers from not representing the word context. Lemmatization, being an advanced technique, considers the surrounding context of a word, such as its vocabulary, morphological interpretation, other grammatical information (such as parts of speech) etc. in order to extract its basic dictionary meaning for its grouping. Lemmatization, unlike stemming, can also consider the context of usage of a word within a particular sentence. So, for the purpose of our work, we chose lemmatization using four predominant parts of speech (Noun, Verb, Adjective and Adverb) tagging or POS tagging technique using NLTK library for grouping on the wrangled text corpus.

3.2.9 Creating N-Grams

The next step of our text pre-processing was the creation of N-Grams, which is also an important part in our text pre-processing. N-Grams are nothing but the contagious word sequence useful to capture the context of using multiple words in a certain sequence. For example, if we consider a sentence "Arunava is a Data Scientist." to be converted into N-Grams of N=1,2,3 respectively that we also call Unigram, Bigram and Trigram, then the output will be something like this –

Unigram Conversion	Bigram Conversion	Trigram Conversion
Arunava is a data scientist	Arunava is is a a data data scientist	Arunava is a is a data a data scientist

(Inspired from FIT5149 Tutorial)

So, it clear that, with the increasing number of N, the tokens become more contextual and complex. But, with the bias-variance optimization concept, we need to select the optimum N for this project that we chose N = 1 and 2 for corpus level analysis and N = 2 for topic level analysis, which we have discussed later.

3.2.10 Removing Least Frequent Tokens

The tokens that appear least frequently within the corpus or in a document possess minimum literary value and higher level of noise because they are potentially the outliers. So, here we have removed those least relevant items from our data. For each publication with unigram, we have removed all words or tokens that have term frequencies less than 4 and for each paper with bigram and trigram, we have removed all tokens that have term frequencies less than 1. In doing so, we ensured better topic modeling developed only on the adequately relevant terms.

3.2.11 Document Vector Creation

The next and final step of our text pre-processing is creation of document vectors which are in turn fed into our LDA machine learning model (discussed later). This approach reduced each document corpus into a vector of real numbers. As part of our research, we used Count Vectorizer (Kong, 2018), one of the basic and popular vector space model, representing the text in form of vectors, in which each word token is represented with its corresponding occurrence within the corpus, also called Term-Frequency. This method of vector representation, also called Vector-Space-Model, assumes that the order of word tokens is not important, but the words are, and the order of tokens should therefore be disregarded. This concept is also called Bag-Of-Words (A. Alahmadi, 2013). For example, if there are three documents with following text –

Document 1: Machine learning is important. Document 2: Data analysis is as important in machine learning. Document 3: Data Science includes data analysis and machine learning.

The above three documents have 20 different words including 10 unique words with respective term frequencies, which is represented in the following Count Vectorizer Model.

	machine	learning	is	important	data	analysis	as	in	mining	includes	and
Doc 1	1	1	1	1	0	0	0	0	0	0	0
Doc 2	1	1	1	1	1	1	1	1	0	0	0
Doc 3	1	1	0	0	2	1	0	0	1	1	1

(Inspired from FIT5149 Tutorial)

The above vectorized representation of the documents for each topic era is supplied to our topic model discussed in detail in the very next section.

3.3 Choosing Appropriate Topic Modeling Technique

Topic modeling is our next important step, which has been done on preprocessed data. In NLP, topic modeling qualifies as a special category of clustering or unsupervised machine learning technique that can discover the hidden abstract topics within a document collection. Since modeling the topics is the predominant segment of our research design, we had to study different available approaches of topic modeling and choose the modeling approach that will go best with our requirement.

3.3.1 Different Topic Modeling Approaches

We have several topic modeling approaches, out of which 'tf-idf' scheme (McGill, 1983), also a document vector approach, results in a 'Term-by-Document Matrix' that has several columns containing 'tf-idf' measures where each column represents a single document within the corpus, thus reducing documents of varied length into fixed length vectors. While this method looks appealing, it suffers from higher dimensionality problem for very large number of documents and shows poor intra-document and inter-document statistical structure. LSI (S. Deerwester, 1990) addressed this dimension issue by "singular value decomposition" technique to find a liner sub-space within the X-matrix of features developed in tf-idf technique, significantly compressing large feature collections while capturing the most important aspects of linguistic notions. But this approach did not provide much advantage over Bayesian method, which is known to resort much simpler "Maximum Likelihood" approach, raising a big question as to why someone should use LSI for topic modeling. The shortcomings of LSI have been tried to solve using pLSI or Probabilistic LSI (Hofmann, 1999) technique that introduced probabilistic nature in LSI using a concept called "Mixture Modeling"; however, pLSI suffers from severe model overfitting issue, which is why we did not prefer this to use as our technique.

3.3.2 Choosing Latent Dirichlet Allocation (LDA)

When comparing all above topic modelling approaches, we encountered a common idea of "Exchangeability" behind the dimension reduction, which is essentially derived from "Bag-of-Words" concept – that the topics of any document does not depend on the ordering of its words and that any specific order of these words can be neglected. So, all the above approaches essentially assume the document corpus to be exchangeable. However, a "Classic Representation Theorem" (Finetti., 1990) represents that, when a "Bag-of-Words" concept is applied to model the topics, exchangeability concept should not be restricted to the words, but it should also be extended to the document level and corpus level. So, we chose "Latent Dirichlet Allocation (LDA)" (David M. Blei, 2003) as our topic modeling technique that is based on this "Classic Representation Theorem", essentially capturing intra-document and inter-document statistical structures using the concept of mixture modelling distribution.

LDA (David M. Blei, 2003), an unsupervised machine learning or clustering technique, which is also a "Probabilistic Generative Model", is used for modeling topics from discrete data e.g. text corpus. This technique follows a Bayesian approach with three level hierarchy that accepts each item from the collection of text and represents it as a finite mixture model, which is built on a latent or hidden set of topics. Then, each of these topics is considered an infinite mixture model on top of a finite topic probability set. While doing 'text modeling', these estimated topic probabilities potentially offer the explicit representation of the documents from the underlined text corpora. As part of this project, because we have just used LDA as our topic modeling technique, it is not relevant to go into details of how LDA topic modeling works and hence, it is not provided.

3.4 Choosing Appropriate Topic Visualization Technique

The next part of our research method is topic visualization after the LDA model has been generated. LDA is a probabilistic model, which suffers from interpretability (Daniel Ramage, 2009). Unless the generated topics from a text corpus is understandable and represented in a human interpretable form, it is very difficult for the researchers to do the field studies in order to find trends and gaps like the one we are doing. A handful of works on topic interpretation through visualization have been done in recent years and, from these techniques, we have selected our preferred method, which are discussed below.

3.4.1 Different Topic Visualization Approaches

A number of topic visualization techniques (Matthew J. Gardner, 2010; Blei., 2012; Justin Snyder, 2013) were developed to let the users browse the documents, their underlying topics and the related terms that constitute those topics. But all these techniques include the browser summarizing the topics using the lists of most probable words/terms within those topics and visualization techniques were limited to the word clouds and charts such as bar graphs, line charts etc. Seeking to develop a more compact topic visualization technique, a tool named "Termite" (Jason Chuang, 2012) was developed that could visualize the term-topic distributions measured in LDA in a matrix format. The tool is based on its interpretation of two measures named "Distinctiveness" and "Saliency". By calculating "Kullback-Liebler divergence" among the topic distributions and computing the marginal distributions of different topics, those two quantities represent a term/word in terms of the amount of information it coveys about a topic and hence, in many cases Termite was considered a comprehensive topic visualization tool by its users.

3.4.2 Choosing LDAvis Topic Modelling Over Others

Termite has a serious disadvantage as it only includes those terms/words that are qualified as high "Saliency" and "Distinctiveness" because the technique finds these two measures as underlined global properties of the terms. But, by doing so, Termite certainly restricts the global view of the topic model, potentially missing the topic specific ordering of words. So, in our research, we have used LDAvis (Carson Sievert, 2014) as our topic visualization technique that is capable of providing a complete global view of topics within a text corpora and the way these topics differ from each other, while providing an opportunity to exercise a deep inspection on different terms/words highly relevant to these individual topics.

3.4.3 LDAvis Topic Visualization

LDAvis (Carson Sievert, 2014) is an interactive web-based topic visualization technique, developed in combination of D3 and R and using LDA as the underlying topic modeling approach. While, backend LDA model is not quite interpretable, LDAvis attempts to interpret this fitted LDA model by answering below three questions –

- What does it mean by a topic i.e. how a topic is interpreted?
- What is the prevalence of each topic within the text corpora?
- How are these topics related to each other and how much are they related?

The visualization system has been accommodated into two panels. The left panel of the system allows the global view of different topics within the corpus. In this global view, the topics are represented as circles of varying sizes on a two-dimensional topic space using multidimensional scaling on the first two "Principal Components" (Robert Tibshirani, 2013) to project the "Inter-Topic distances" on a two-dimensional coordinate. **The centers of these circles are determined by estimating the distances within the topics, which can also be interpreted as the knowledge differences or gaps among these topics. The areas of these circles signify the proportion of relative prevalence of these topics within the corpus.**

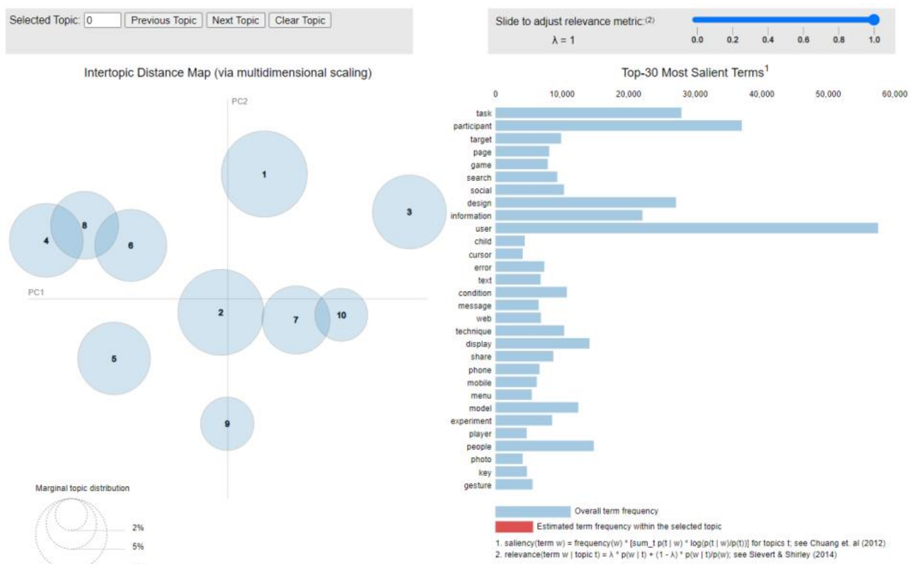

Fig 1: Global view of topics in LDAvis system for 10 topics

Figure 1 above shows an example of the global view of 10 topics for a corpus and their relative prevalence. We can see that topics 1, 2 and 3 are the three most prevalent topics taking up more than 25% of the entire topic space, whereas topic 9 and 10 are the two least prevalent topics taking around 10% of the entire topic space. This percentage of prevalence could be identified from sizes of dotted circles at the bottom left and their respective percentage of prevalence within the corpus. We could also see that topic 2 and topic 5 are close to each other, indicating that they are more similar in context than topic 3 and topic 4 which are distant from each other, indicating that they are different in context and have a gap in knowledge more than that of topic 2 and topic 5. We can also see that topic 4 and topic 8 have overlaps, which lets us presume that these two topics have some overlaps in knowledge and context, which is indicated by the common topic space between them.

The right panel shows a horizontal bar graph, in which each horizontal bar represents a term most useful in the currently selected topic. We can also see a pair of overlaid bars in sky-blue and red in colors, representing the corpus-wide term frequency (in sky-blue) and topic-wide term frequency (in red) for that term on the given scale. So, when a topic from the left panel is selected, the most relevant terms of the topic is shown in the bar chart on the right panel with their topic-wide term frequencies in descending order, whereas the remaining portions of the horizontal bars in sky blue are the unexplained term frequencies in respect to the topics, which are nothing but the term frequency of that word in the corpus. When the topic from the left panel is deselected, then we could see that the red overlaps of the horizontal bars disappear, leaving the term frequencies on the most relevant words within the corpus. Figure 2 below shows an example of this where topic 1 has been selected. We can see that "user", "display" and "interface" are the three most relevant terms for topic 1 with their respective topic-wise term frequencies (in red) approximately 10,000, 8,000 and 5,000 against their corpus-wide term frequencies (in sky blue) approximately 57,000, 13,000 and 38,000. Similarly, "application", "condition" and "information" are three least relevant terms of the topic

1. When we deselect the topic 1 from left, we again go back to Figure 1 with only relevant terms with the corpus and their respective corpus-wide term frequencies and we see that the red bars disappear.

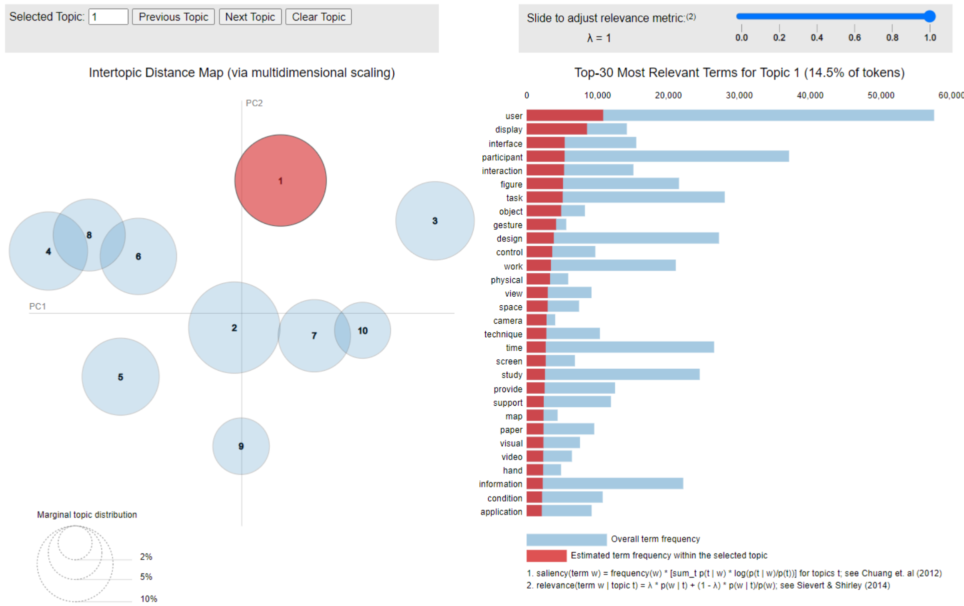

Fig 2: View of topic 1 and its relevant terms in respect to the global view

A key innovation of the above visualization system is, finding a nice way to explain the term-topic relationships. LDAvis does it by proposing a measure "Relevance" to interpret this relationship by ranking terms against the topics.

Definition of Relevance:

If φ_{ij} is the probability of a term $w \in \{1,2,3,4, \dots \dots, T\}$ within a topic $t \in \{1,2,3,4, \dots \dots, K\}$, where T = the number of words/terms within a lexicon and P_w = the "Marginal Probability" of the word/term w within the text corpora, the "Relevance" of the word/term w with the topic t can be defined below.

$$R(w,t|\lambda) = \lambda \log_e(\varphi_{ij}) + (1-\lambda)\log_e(\frac{\varphi_{ij}}{P_w})$$
$$Where, \lambda = is\ a\ weight\ paramer\ between\ [0,1]$$

The weight λ provides an intelligent way to associate the terms and the topics on the neighboring panels. When we set $\lambda \approx 1$, the words/terms are ranked in the descending order of their topic-wide probabilities and in the ascending order of the corpus-wide probability, whereas setting $\lambda \approx 0$, does the opposite, i.e., the words/terms are ranked solely based on their close association with the selected topic. In Figure 3, a comparison of such topic-term relevance for two different values of λ has been shown. The main challenge here is to find the correct value of λ. We have discussed our approach in the next chapter to find a way for the effective use of λ. Finally, as part of this project, we have used pyLDAvis package in Python to implement this topic visualization.

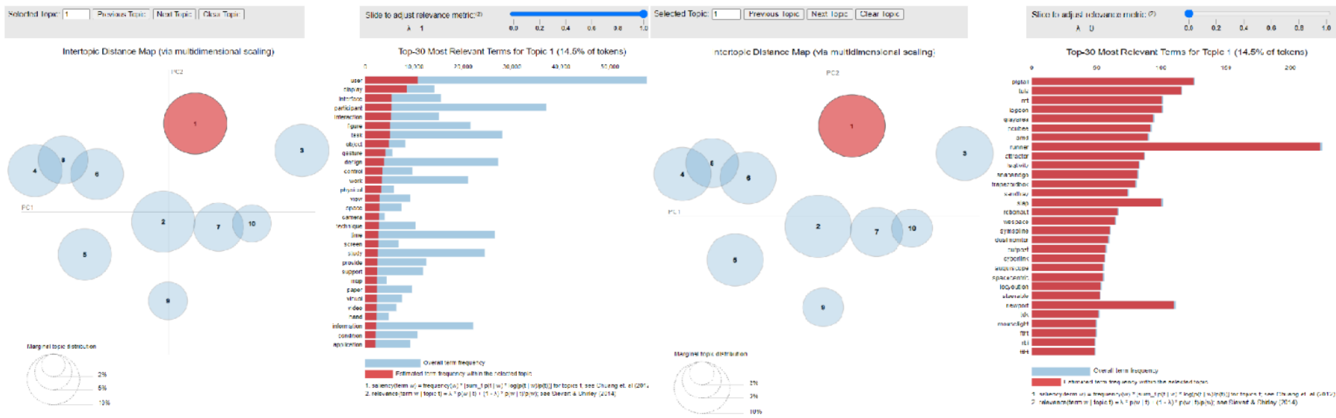

Fig 3: Comparison of Topic-Term Relevance with different values of Lambda such as 0 and 1

3.5 Setting Up Model Parameters

Once the complete text corpus is pre-processed and the topic modeling and visualization techniques are chosen, it's now the time to move on doing the actual topic modeling and present the visualization using its different features. But before that we have understood and set up appropriate model parameters.

3.5.1 Using N-Grams for LDA Topic Modeling

As part of this research, we present a new method to utilize N-Grams in LDA topic modeling and in LDAvis techniques. Because LDA is based on "Bag-of-Words" (David M. Blei, 2003; Michael Nokel, 2015) concept, the word sequences don't really matter in this technique and hence, LDA topic modeling cannot accept the bigrams/trigrams directly. In line with this LDA property, LDAvis is also incapable of handling bigrams/trigrams directly. So, we proposed a way to utilize bigrams and trigrams in LDA model under the disguise of unigrams. In our approach, we have converted a bigram (U, V) into a unigram U_V and a trigram (U, V, W) into U_V_W. In that way, we have converted a bigram/trigram in unigrams, but maintained the individual property of the respective N-Grams. For example, the model for below bigrams will look this this –

Bigrams before conversion	Unigram conversion
Arunava is	Arunava_is
is a	is_a
a data	a_data
data scientist	data_scientist

3.5.2 Choosing Appropriate N-Grams

In our research, we have decided to develop our topic models based only on unigram and bigrams. Using trigrams and above N-Grams had serious issues with the generated topic model. Since, we have removed the least frequent terms in order to get rid of those unnecessary tokens that generate outliers into the modeling system, the trigram model built in that manner suffered from huge information loss, because repetitions of tokens in trigrams are much less frequent than unigrams and bigrams. So, we ended up losing many important tokens in the process of removing the least frequent ones while using trigrams. Therefore, we have generated topic models with unigrams and bigrams for each of the four topic eras mentioned above. However, we have used unigram model (while we have used bigram model for the same analysis too) only to do corpus level analysis to understand the behavior of most relevant tokens within the corpus and how their presence changed over time. While doing the topic level analysis, we saw that bigrams can better interpret the topics than unigrams do based on most relevant pair of words. So, we have accepted unigram models (also used bigrams) while doing the analysis at corpus level, but discarded them, while doing deep level analysis of topics.

3.5.3 Choosing Number of Topics for Each Topic Model

One of the very interesting questions of our research is "**How many topics should be the best to look at in each topic model for each topic era**?". LDAvis visualization technique lets us supply the number of topics we want to generate within a topic model. Now, too many and too few numbers of topics both distort the insights and hence, we were interested to look into the number of publications within each topic era that we have decided to view. We saw that the number of publications increased significantly from the topic era 1982 – 1991 (447 publications) to 2011 – 2018 (3929 publications). So, we have decided to generate 20 topics (considered to be moderately good) for each of these topic era, where because of less number of topics in the very 1st era, we ought to experience many overlapping topics and in the successive era, we expect the topics to spread through the two dimensional topic space (which actually happened in our case).

3.5.4 Choosing Appropriate Relevance Parameters

In order to do the deep level analysis at the topic level, we need to understand intricacies of term-topic relationships that highly depend on the relevance of the terms describing these topics. As discussed earlier in LDAvis, we can find these term-topic relationships at different level of relevance defined in mathematical terms. When this relevance parameter λ takes on a value towards 1, the model shows the relevant terms more general towards the topic, while, on the other hand, setting λ towards 0 will result in ranking relevant terms more specific to the topic (See Figure 3). This feature of LDAvis helps us understand how the topic related terms and their rankings change from more general to very specific context, allowing a deep level analysis of the term-topic relationships. For every separate value of λ for each topic, the most relevant terms are re-ranked every time. As part of this research, we have set the values of $\boldsymbol{\lambda = 1, 0.6, 0.2\ and\ 0}$ for each topic analysis under each topic era (**total $\boldsymbol{4 \times 20 = 80}$ topics**) to understand different CHI themes and their relationships with the relevant terms.

3.6 Topic Modeling, Visualization and Generating Insights

The last and final step of our research method is to build topic models using LDA, visualize them using the chosen visualization system and interpret the findings through useful insights. For that we chose the below method.

3.6.1 Choosing HCI Research Areas for Topic Interpretations

LDA technique creates topics in terms of the clusters of words/tokens and so does LDAvis that generates the topics with their most relevant words ranked based on their induvial term frequencies within the corpus. This ranking of terms changes based on different values of λ. So, as part of this research, **we made an assumption that each of the relevant terms of any topic within the topic model may be classified at least one of the below HCI research areas inspired by a study** (Biplab Ketan Chakraborty, 2017) –

1. Design
   - User Interface Design/UI Design
   - User Experience Design/UX Design
   - Interaction Design
2. System/User Performance
3. User Model
4. Usability Study
5. Human Factor
6. Cognitive and Linguistic Science
7. Computer System/Science
8. Software/Application Development
9. Crowd/Collaborative Work
10. Artificial Intelligence and Data Science
11. Mobile Technology
12. Social Networking/Media
13. Information Security
14. Ethnography
15. Engineering
16. Social & Psychology

Once the topic model is generated and the LDAvis visualization is created we sought to categorize each token under one or more of these research areas. Now each topic will contain multiple such relevant words, so a topic will essentially cover a combination of one or more of these research areas. The more a topic consists of enough proportion of words/token covering each research area, the better a topic can be called as an inter-disciplinary HCI topic.

3.6.2 Metric to Measure the of Terms'/Topics' Contributions

As part of this study, we have introduced our own metric to quantify the contribution (in % scale) of a term or topic towards these major HCI areas. Because LDAvis assumes that the frequencies of the terms are proportional to their prevalence within a topic, the summation of all term frequencies for a topic is proportional and strongly correlated to the relative prevalence of that topic within a corpus. So, we can assume that the contributions of these terms and topics towards these aforesaid HCI research areas are also proportional to their respective term frequencies (for individual terms) or the summation of the term frequencies (for individual topics).

We will describe this by an example. If a topic is described by only three terms "Mark Menu", "Recommender System" and "Heuristic Evaluation" with respective term frequencies 100, 80 and 50 and if these terms can be categorized under the research areas "User Interface Design", "Artificial Intelligence and Data Science" and "Usability Study", then the prevalence of the three research areas within the topic can be calculated as –

Prevalence of "User Interface Design" within the topic $= \frac{100}{100+80+50} = 43.5\%$
Prevalence of "Artificial Intelligence and Data Science" within the topic $= \frac{80}{100+80+50} = 34.8\%$
Prevalence of "Usability Study" within the topic $= \frac{50}{100+80+50} = 21.7\%$

So, we can conclude that 43.5% of the topic content is about "User Interface Design", 34.8% of the topic content is about "Artificial Intelligence and Data Science" and 21.7% of the topic content is about "Usability Study". Therefore, the topic contributed more towards "User Interface Design" with its respective percentage value.

3.6.3 Generating Topic Models

At this stage we are all set to generate the topic models using LDA and visualize them using LDAvis. So, we have generated one topic model each using each of unigrams and bigrams techniques for all four topic eras. **So, for the 4 topic eras, we have created 4 unigram and 4 bigram LDAvis models, building total $4 + 4 = 8$ models and $(4 + 4) \times 20 = 160$ different topics.** For each of these 8 models we tried to understand two things. Firstly, what are the relative prevalence/contributions of the top 30 terms describing the corpus towards the 16 HCI research areas and secondly, what is the relative prevalence/contribution of each topic from each topic era towards the same 16 research areas. We did this through separate corpus and topic level analysis.

3.6.4 Corpus Level Analysis

For corpus level analysis we considered all 8 unigram and bigram models. From each of the 4 unigram models, we captured top 30 unigrams that best described the corpus for the respective topic era. Then we have generated a horizontal bar chart with these 30 relevant terms and their respective term frequencies for a particular topic era. And, for 4 different topic eras, we have generated 4 such bar charts (See Figure 5). From these bar charts we have done very high level of analysis on 37 years of CHI proceedings. However, we faced challenges to classify each of these unigrams against one or more aforesaid 16 HCI research areas, for which we used the bigram models. For bigram models we did the same exercise (See Figure 6), but here we went one step further by classifying each of these 120 bigrams from 4 topic decades under one or more of the 16 HCI research areas. Then we have calculated the prevalence of each of these bigrams towards one or more of these research areas and accordingly, we have shown the contribution of each topic era towards these 16 HCI research areas.

3.6.5 Topic Level Analysis

After the corpus level analysis, we sought to do even deeper analysis at each topic level. We have already observed that unigram topic model does not provide us a better understanding on the context of the topics (See section 3.5.2). So, we have decided to do the topic level analysis based on our bigram models (See Figure 8 for the bigram topic model for 1982 - 1991) for each of these above four topic eras. Now, for each topic we tried to capture all relevant bigram tokens for $\boldsymbol{\lambda} = \mathbf{1, 0.6, 0.2}$ $\boldsymbol{and}$ $\mathbf{0}$. **So essentially for each topic we have examined 30 * 4 = 120 bigrams and for each topic model we potentially analyzed 120 * 20 = 2400 bigrams. And for all topics in the 4 different topic era we have finally evaluated 2400 * 4 = 9600 bigrams individually.** We had to ignore a few bigrams, because they could not be classified under one of the above 16 HCI research areas, but essentially, we have considered most of them and classified each of them under one or more of these 16 above CHI research areas. Now, we have collected the term frequencies of each of these bigrams from each of these topics. Finally, we have calculated the relative prevalence of those 16 research areas towards each topic for each of the topic era (See Figure 9 left) and then, finally we calculated the contribution of each decade towards these major HCI research areas by averaging the relative topic prevalence (See Figure 9 right).

4. CHAPTER 4: RESEARCH OBSERVATIONS AND ANALYSIS

4.1 Introduction

This chapter broadly discusses the key research observations and their analyses. In the previous chapter we discussed our methods and approaches to build the topic models and visualize them in a human interpretable form. Once the models were developed, we experimented on few key features of our models to understand mainly two aspects such as the contribution of top 30 terms of each topic era towards 16 major HCI research areas and the contribution of 20 topics from each decade towards those major research areas. We then discussed our corpus and topic level analytical approach to understand those two aspects. We will now discuss the observations from the above-mentioned approaches in following manner.

1. General Observations of CHI Conferences
2. Observations of most relevant words/tokens at the corpus level
3. Observations of topics and the most relevant words/tokens describing them

4.2 General Observations of CHI Conferences

Before we could jump into the corpus and topic level discussions, we would like to show some general level insights from the data on CHI conferences taken place between 1982 – 2018. For that, we have plotted the number of publications in four decades and the publications took place in consecutive years.

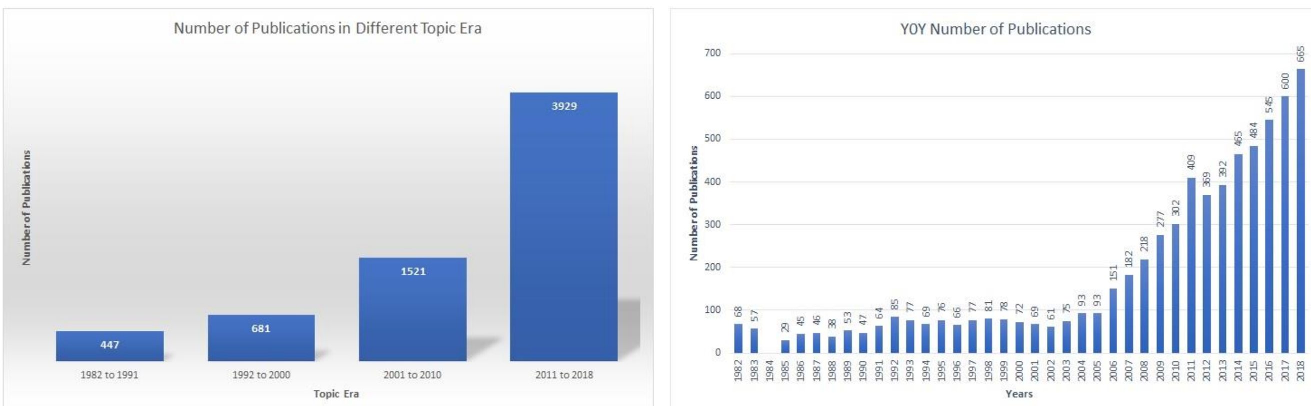

Fig 4: Comparison in the numbers of publications in four topic eras versus the numbers of publications in consecutive years

In the Figure 4, we can see that the number of conferences in consecutive decades increased significantly with the number of publications reaching from 447 (in 1982 to 1991) to 3929 (in 2011 to 2018). However, when we took a closer look into the year-wise publications (in Figure 7), we observed that the number of papers started increasing heavily since 2004 and this number grew significantly year after year till 2018, the year of highest number of publications. We also surprisingly saw that there was no CHI publication in 1984. This observation indicates that, due to technological booming in 21st century, HCI research received huge attention especially because of the inventions in mobile, internet and web technologies, which was quite scarce in the 20th century where the main focus was to build the computers and improve the interface, usability and performance.

4.3 Observations of Most Relevant Words/Tokens at the Corpus Level

After having an idea on how the numbers of publications have taken place over the years, we would like to understand how top 30 tokens from both unigram and bigram topic models represent the complete corpus for a certain topic era with respect to the 16 aforesaid HCI research areas. Because we could barely classify the unigrams for different HCI research areas (for which, we understood bigrams are more suitable), we did very high level of corpus analysis based on unigram topic models. However, bigrams with their better interpretability gave us better opportunity to analyze the corpus based on their prevalence towards those CHI research areas and we could come up with the respective contributions of each of the 4 topic eras towards 16 HCI research areas.

4.3.1 Analysis of Unigram Model

With our approach (section 3.6.4) to perform very high-level analysis of unigram models, we have observed (in Figure 5) that, in all four consecutive decades, "User", "Task", "Interface", "Design" etc. were the most frequented terms in the topics. This gives us a good reason to believe that the main focus of CHI conferences was mostly to focus into user centric interface, experience or interaction design for the period of 37 years. However, we also found that this specific focus was biggest during the second topic era (1992 - 2000), where we could see that interface/experience/interaction design consumed around 53% of the most relevant term distributions, possibly because of the huge focus of researchers to develop better GUI experience to the end-users at that time. We could also find that the first two topic eras did not have much work in mobile technology and social media/networking, but we found quite a bit of focus into these two areas during the next two topic eras because of huge technological advancements in mobile and web technology during 21st century. Also, we have found that the highest measured term frequencies for first two topic eras were 1200 and 2400, while the same are 59000 and 123000 for the next two topic eras respectively. This observation is validated by the number of papers published in four topic eras (Figure 4), where we could see that the last two decades had significantly higher number of publications than the first two decades.

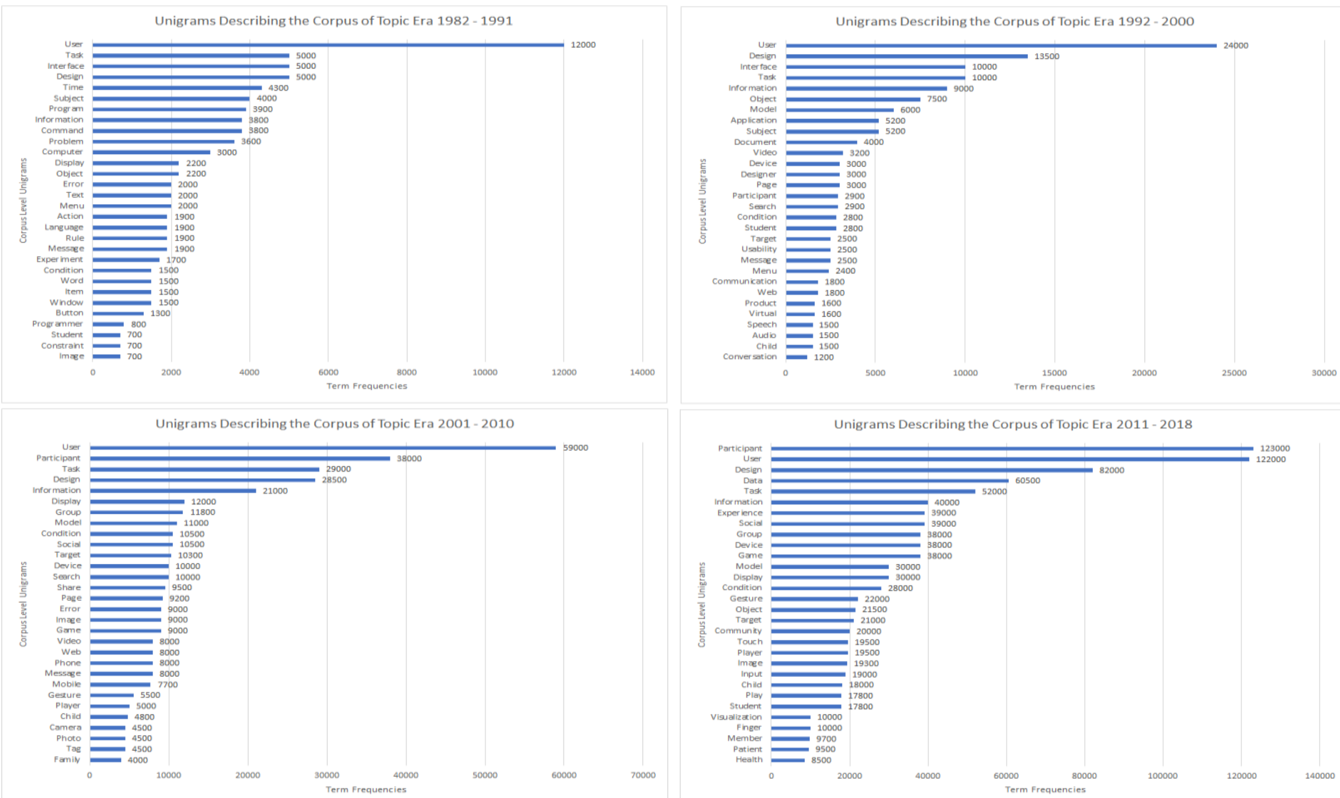

Fig 5: Term Frequencies of most relevant Unigrams in the corpus in four different topic eras

We furthermore tried to understand how these most frequent words in the corpus have changed over time. In this aspect, we could see that around 19 out of 30 (63%) most relevant words that had appeared in 1992 – 2000, disappeared in 1992 – 2000. We saw that 15 most frequent words (50%), appearing during 1992 – 2000, disappeared in 2001 – 2010 and 13 most frequent (43%) words, appearing during 2001 – 2010, disappeared during 2011 – 2018. Finally, we observed that 21 out of 30 words (70%) that had appeared during 1st decade ended up disappearing in the final decade; however, we could see a decreasing trend in vanishing of the most frequent words decade after decade. So, this indicates that, even though, during the initial years CHI conference might have lost its focus, later years somewhat tried to maintain around 50% or more of the leading descriptive terms. So, we believe that because of the technological advancements in the four consecutive topic eras, CHI proceedings immensely shifted their research focus and at the same time, they left some research areas at relatively budding stage without making significant developments to themselves.

4.3.2 Analysis of Bigram Model

With our approach (section 3.6.4) to perform a bit deeper level of analysis of the whole corpus based on the most popular 30 bigrams for each of the four consecutive topic eras, we have observed (Figure 6) that the era 1982 – 1991 started with a bit primitive level of computing. It is evident by the appearance of popular bigrams such as ‘command line’, ‘text object’, ‘load plan’, ‘pie menu’ and ‘data structure’ that are nothing but the primitive computer interfaces or experiences. We also observed the terms related to early days of usability studies such as ‘error message’ and ‘user recovery’, and early days of computer systems such as ‘input output’, ‘Windows System’, ‘control panel’ and ‘event handler. While we could also see few advanced level terms such as ‘speech recognition’ and ‘adaptive systems’, their appearance was perceived a bit random in nature and not much related to the other popular terms. So, we could safely conclude that the CHI conferences during this era were focused into building computer technology and tools for the humans to interact with them at the very basic level and the advanced technologies discussed were in planning stage and far from implementation.

Moving onto the era 1992 – 2000, we observed that this decade introduced bigrams such as ‘web page’, ‘mark menu’, ‘video display’, ‘drag icon’ etc. that are related to interface and interaction designs. We also saw few popular terms related to user experience such as ‘visual environment’, ‘visual display’, ‘audio aura’ etc. However, they mainly refer to the initial days of user interface/experience/interaction designs. We saw few terms related to usability study, human factor and cognitive science such as ‘heuristic evaluation’, ‘style guide’, ‘mental simulation’, ‘cognitive function’ etc. So, it’s apparent that HCI research started picking up different interdisciplinary factors. However, we did see that the other comprehensive HCI research factors such as social & psychology, ethnography, engineering, AI & data science, user/case studies, performance factors and information security were missing as part the research context. So, we can say that this era was speeding towards actual HCI research with its lot of focus into interface/experience/interaction designs and a bit of focus into other interdisciplinary areas such as usability study, human factor and cognitive science. This is because of advancement in computer technologies, but this era also had very less focus into other important HCI research factors.

We furthermore investigated the next topic era i.e. 2001 – 2010, where we saw a huge shift in CHI proceedings with its focus mainly into mobile and internet technology, which is evident from the most popular bigrams such as ‘mobile phone’, ‘web page’, ‘social network’ etc. We surprisingly saw that the terms like ‘user interface’, ‘interface design’, which were predominant in the previous decade, went missing from most popular word pairs during this time period. At the same time, we saw much advanced level terms related to user interface/experience/interaction designs such as ‘virtual human’, ‘hover widget’, ‘video conferencing’, ‘bubble cursor’ etc. We further observed the terms from different HCI research areas such as human factors (‘eye contact’, ‘physical activity’), usability study/user-system performance (‘large display’, ‘error rate’, ‘tactile Feedback’, ‘API Usability’), social science (‘group member’, ‘Old Adult’, ‘Young People’), computer science/technology (‘open system’) and ‘Information security’ (‘privacy policy’, ‘graphical password’). So, it can be deduced that huge technical advancement in 21st century led HCI research to radically shift its focus into mobile device and internet technology. In this era, user interface/interaction design was not limited to designing individual graphical components but spread across to design complete user experience such as virtual reality. The CHI proceedings also became more comprehensive and matured being interdisciplinary in nature, doing the research from a very holistic perspective.

This trend continued till final decade 2011 – 2018, but CHI proceedings progressed to discuss much more advanced level topics. While we saw mobile phone and social networking being the same hot topics in this era, we also observed the discussion of much more advanced level terms such as ‘social media’, ‘live streaming’, ‘news feed’ etc. We experienced the consideration of much more advanced level of experience design such as ‘shape changing interface’, ‘visual question’ etc. that further progressed to the level of AI technology such as ‘gaming’, ‘smart home’ etc. The consideration of social science advanced from studying different users in the previous decade (‘group member’, ‘Old Adult’, ‘Young People’) to the different characteristics and psychological situations of the users such as ‘mental health’, ‘dementia’, ‘care recipient’ etc. in the current era. We also found much advanced level of consideration of human factors/cognitive side such as ‘gesture set’, ‘phantom sensation’, ‘bend gesture’ etc. So, cumulatively we can say that this decade almost retained the trend of the previous decade while proceeding to much more advanced technological era.

We have built two stacked bar charts (Figure 7) representing the contribution (discussed in section 3.6.2 above) of different decades towards each HCI research areas and vice versa and we saw that the first two decades of the research areas are quite disproportionately prevalent. In the first era, only 10 of total 16 research areas were covered, while in the second decade 12 research areas have been covered and they were disproportionate in nature as is the case during the 1st decade. This observation indicates same situation that CHI proceedings were not comprehensive in nature before 21st century, not covering most of the required areas of HCI. But we saw the opposite trend in the next two decades where we observed that most of the research areas have been sufficiently covered, proving our claim that in 21st century, the HCI research were comprehensive, covering most of the important HCI factors. However, no bigrams from Engineering and Ethnography could qualify within the top 30 terms in the four consecutive decades, which is possibly because CHI conferences systematically left these two research areas unexplored year after year and the research papers possibly did not consider much to cover these two areas.

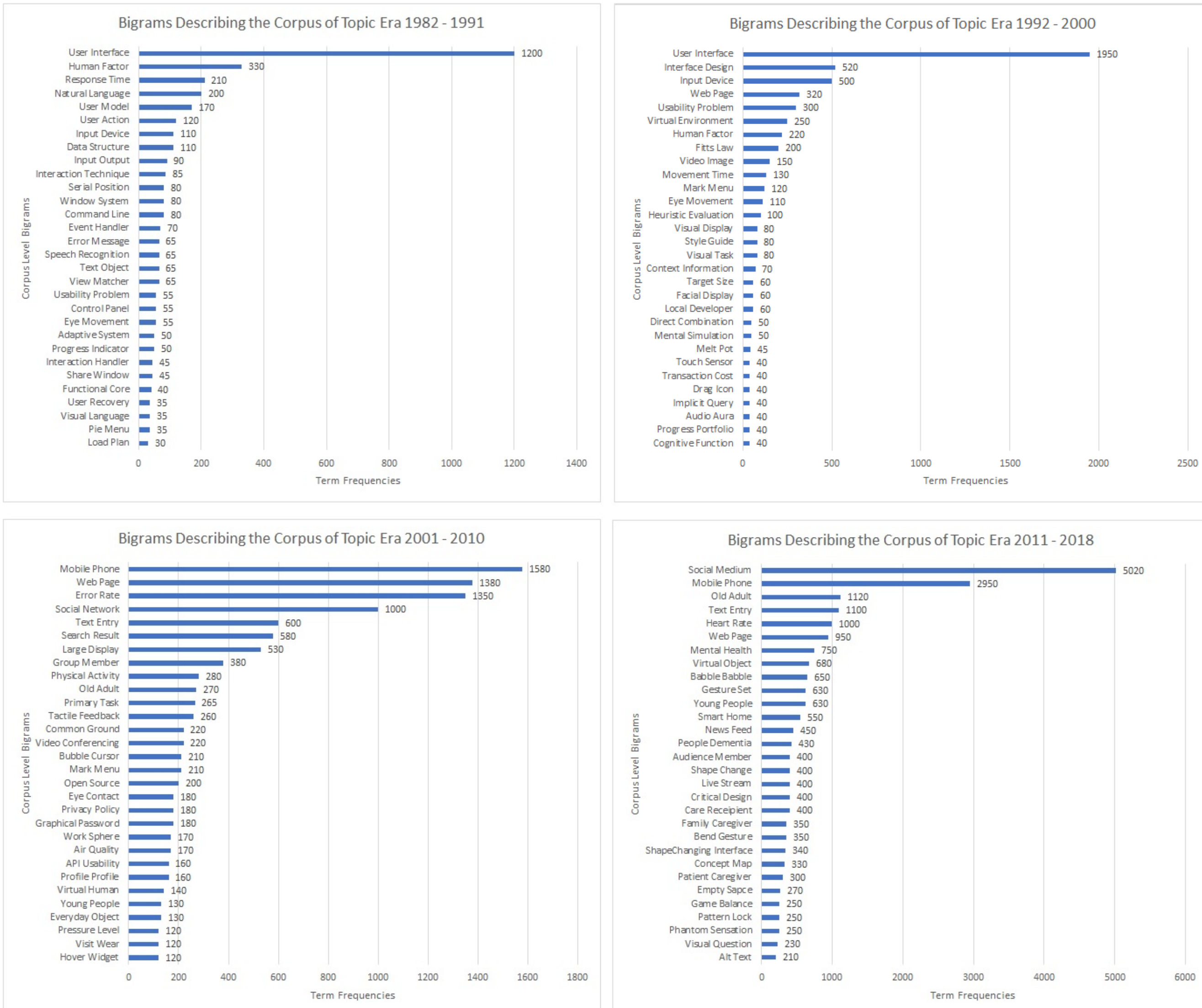

Fig 6: Term Frequencies of most relevant Bigrams in the corpus in four different topic eras

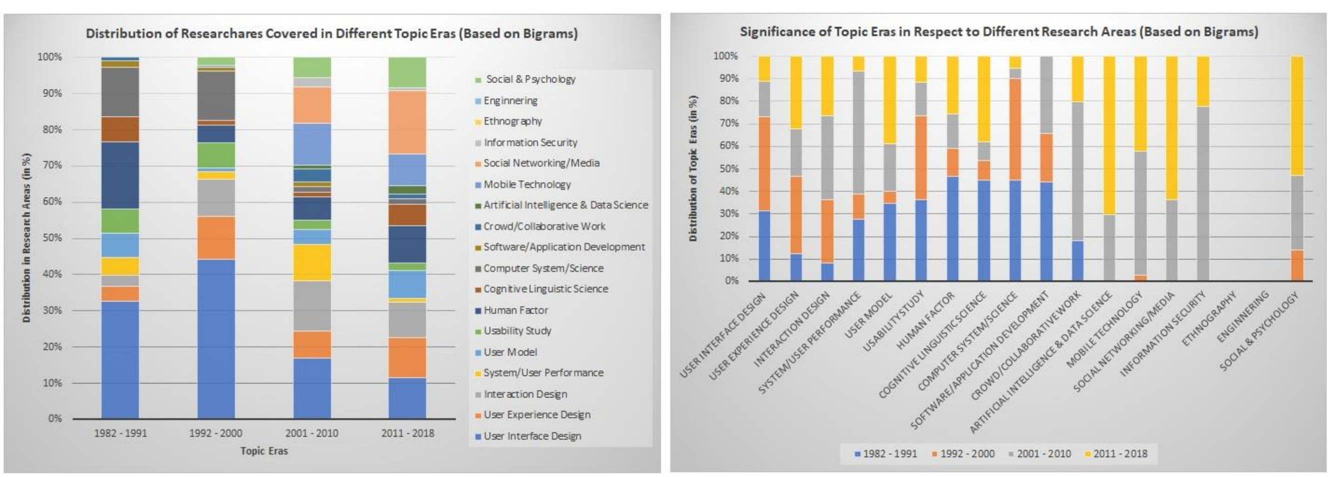

Fig 7: Contributions of different topic eras in major HCI research areas vs prevalence of major HCI research areas in different topic eras

4.4 Observations of Topics and the most Relevant Words/Tokens Describing Them

After doing the corpus level analysis we are set to begin even deeper analysis of individual topics generated in the four bigram topic models (reason for choosing bigram topic models has been discussed in section 3.5.2 above). For that we have analyzed 4 LDAvis bigram models during the four decades. The analysis is given below.

4.4.1 Analysis of Topic Model during 1982 – 1991

We have shown the bigram topic model (Figure 8) for 1st decade (1982 – 1991). We can observe two characteristics of the topics within this topic era. Firstly, the topic 1, 2, 3 and 6 are the four independent topics distant from each other with their marginal distribution of around 20% of the entire topic space, while all other topics stood 80% of the corpus. These 80% of the other topics are overlapping with each other with minimal knowledge gaps among themselves, indicating that most of the knowledge during this decade was accumulated within them and accordingly 80% of the CHI papers are very similar to each other with everyone talking about almost similar things, while topic 1, 2, 3 and 6 have discussed about different aspects with their observed knowledge gaps among each other. Secondly, we can also observe that most of the topics are immatured in nature with many of their average distribution around 2% towards the corpus. For example, when we compared topic 1 and 3 based on their most relevant terms, we saw that they are actually quite different. Topic 1 mostly talks about interface/experience designs from the perspective of different types of users and performance metrics, while topic 3 talks about different primitive computer systems and their significance within a shared working environment. We saw very similar behavior between topic 1 and 2. On the other hand, we also compared topics 15 and 19, which are overlapping with each other, and saw that they both talked about very similar types of user interface designs and their usability.

We have also plotted (Figure 9 left) the relative prevalence of each of these 20 topics towards 16 different HCI research areas based on our defined metric. Then we have plotted the average prevalence (Figure 9 right) of all these research areas for decade (1982 – 1991) that provides us more insights about the topic trends and gaps. In the left figure, we saw that for each topic, the research areas are disproportionately distributed with user interface design getting the most attention. When we found out the average of all the prevalence in the bar graph on the right-hand side, we again saw that user interface design got disproportionately higher attention than any other research areas, with its total contribution of around 26% during the whole topic era. We interestingly saw that usability study, computer systems/science and experience design received 2nd, 3rd and 4th highest attention. Interestingly, AI and data science received 5th highest attention, while interaction design ranked 11th in the list. We saw that mobile technology, social and psychology and social networking received very less attention. No keyword was found from ethnography and information security, while few keywords have been found from engineering.

These findings are in line with our understandings so far. As we already know that this was a primitive computing era and computer technologies were being developed, significant effort was done into developing computer systems and their interfaces and to understand their usability and user experiences. These computers and their primitive interfaces were not interactive in nature, so interaction design received relatively less attention during this era. The other factors such as human factors, cognitive science, social & psychology, crowd/collaborative works etc. experienced less to least attention because most of the topics were not matured during this era. However, we interestingly saw a special attention to AI and Data Science during this primitive age of computing, which indicates that the researchers were quite aware of the next generation of computing and they took this opportunity to discuss about it in their scholarly works. Finally, we have also shown (Figure 10) few notable works covered during this topic era.

PS: The relative prevalence of two topics towards research areas (Figure 9) are different from their similarity measures (Figure 8) as they are based on different metrics. Exp. Topic 10 and 15 are very similar, but their contributions towards HCI research areas are different.

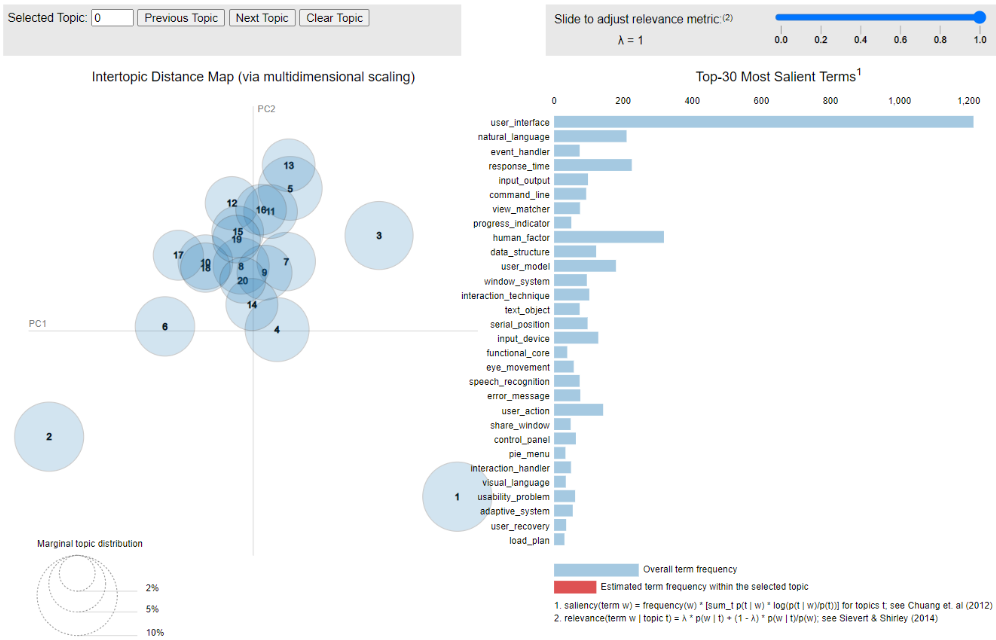

Fig 8: Bigram topic model for the topic era 1982 – 1991

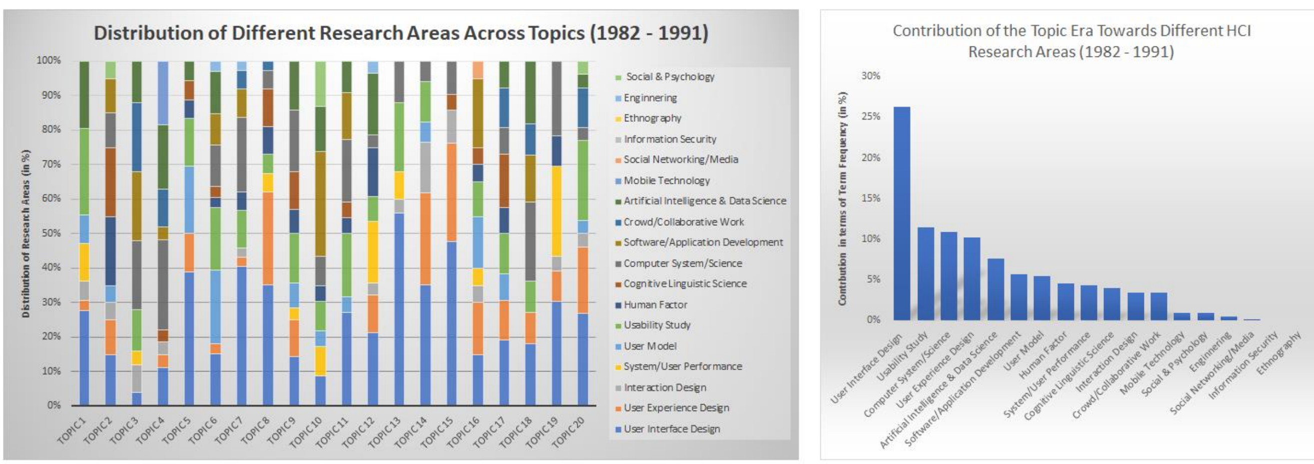

Fig 9: Prevalence of difference research areas towards different topics vs average contributions of 16 HCI research areas in topic era 1982 - 1991

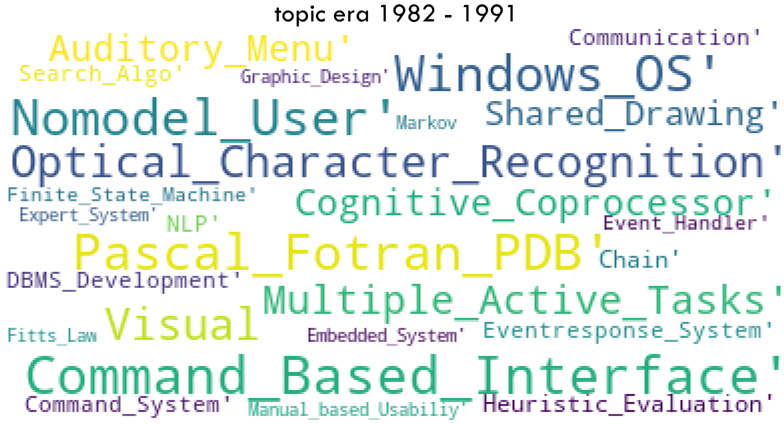

Fig 10: Few notable works during topic era (1982 - 1991)

4.4.2 Analysis of Topic Model during 1992 – 2000

We have shown the bigram topic model (Figure 11) for 2nd decade (1992 – 2000). Analyzing the characteristics of the topic model during this era, we saw that the topics in this era have somewhat spread across the topic space. We no longer see a cloud of most of the topics clubbed at one corner of the topic space as we saw in the previous decade, but saw a chain of semi-overlapping topics in the middle of the topic space. This indicates that most of the literatures were somewhat similar to each other within the overlapping chain. However, the topics residing at the two ends of the chain should be quite dissimilar in nature as they are distant from each other. We further saw that first two topics, representing relative distribution around 10% of the entire topic space, are much distant from each other and essentially have gaps of knowledge in between. Topics 6 and 10, representing around 8% distribution of the topic space, essentially shows similar behavior with relatively less knowledge gap between themselves. However, the topic chain that is in the middle (representing around 82% of distribution), is an indicator that there was an effort within this era to make a good balance between the similarities and gaps among the topics. However, we could still see immatured topics such as topic 13, 14, 18, 20 etc. with their marginal distributions around 2% each, but number of such topics are less than what we saw in the first era.

Plotting Figure 12, we observed that the term frequencies of the relevant terms from most of these topics have disproportionately higher bias on research areas such as UI design, UX design and usability studies. When looking in the average prevalence of all these research areas, we saw that user interface design received most attention with its 18% of prevalence. Experience design and usability study received 2nd and 3rd most preferences with their individual prevalence around 13%. The other important HCI research factors such as interaction design, computer system/science, social & psychology, system/user performance, human factors and cognitive science received disproportionately lower level attention with their individual prevalence below 10%. Interestingly, social & psychology that did not appear within top 30 words at the corpus level appeared at the topic level. AI & data science that had moderately good appearance in the previous decade started fading away in the current decade. Mobile technology and social networking have very less presence in this decade. Software and application development lost its rank from 5th in 1st era to 12th in this era. Ethnography have at least appeared this time, but the discussion on information security was far from reality.

This observation goes hand in hand with our understanding so far. During 1992 – 2000, computer systems started becoming more advanced with budding interface and experience design technologies. So, the concept of usability was also accounted for. But, since the technological advancements like mobile, internet and collaborative working technologies before 21st century were still underplaying, there was still less scope for interactive applications. But, HCI research had slowly but started picking up other comprehensive aspects of it. Since, information science was at the very basic stage, information and data security did not have much role to play in this era. The surprising loss of focus in software and application development can be explained in the way that HCI has not been a core field of software engineering and had more focus in human centric software/application design. So, the areas such as UI/UX/usability design covered more of software/application development aspects than the actual application development as an individual research side. Finally, areas like ethnography and engineering were still far less explored during this decade. Finally, we have also shown (Figure 13) few notable works covered during this topic era.

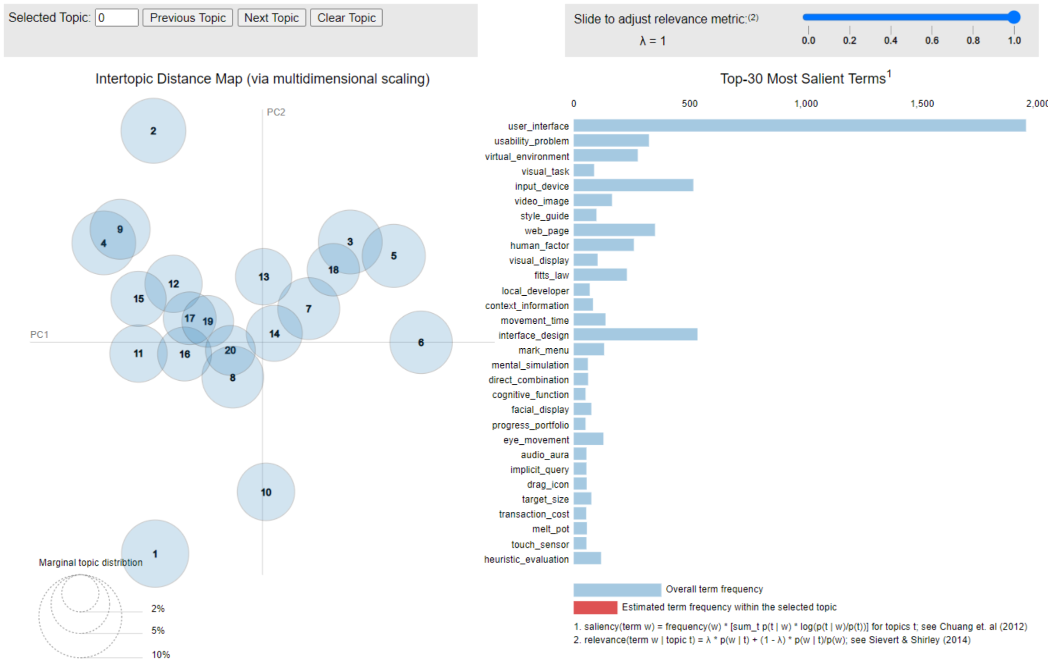

Fig 11: Bigram topic model for the topic era 1992 – 2000

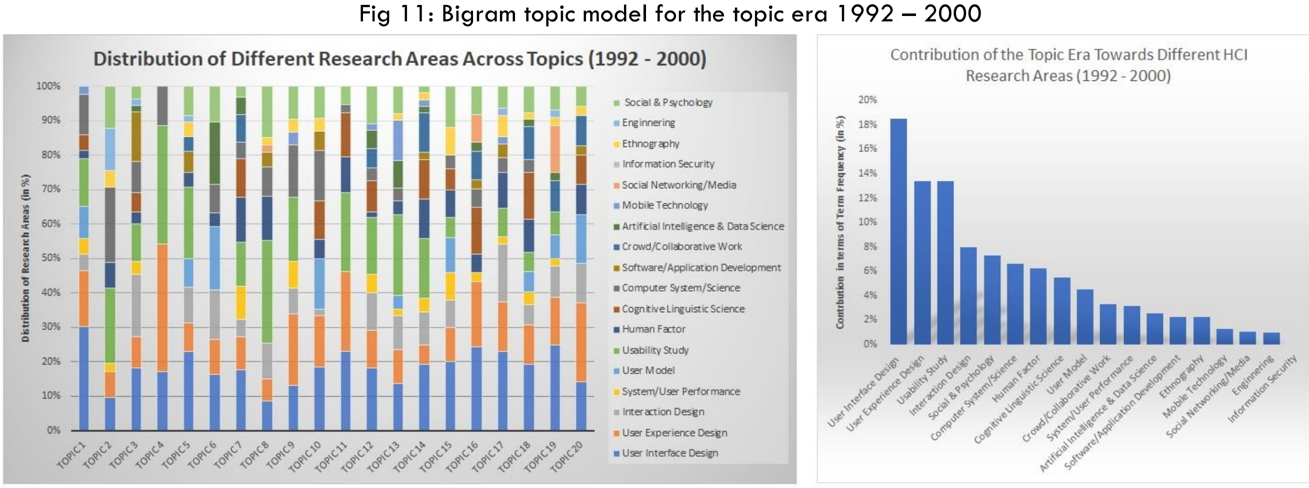

Fig 12: Prevalence of difference research areas towards different topics vs average contributions of 16 HCI research areas in topic era 1992 – 2000

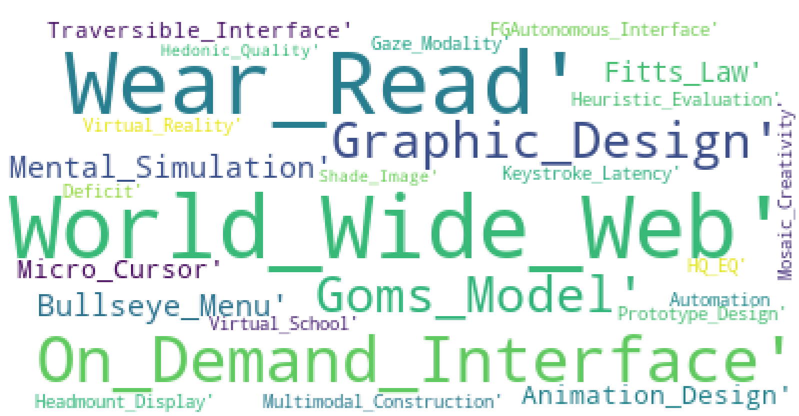

Fig 13: Few notable works during topic era (1992 - 2000)

4.4.3 Analysis of Topic Model during 2001 – 2010

We have now shown the bigram topic model (Figure 14) for 3rd decade (2001 – 2010). Analyzing the characteristics of the topic model during this era, we observed that the topics got even more scattered across the corpus as compared to the previous decade. However, we could still see a weaker chain of 13 semi-overlapping topics this time. This indicates that while 7 topics became distant from each other, creating a gap of knowledge between themselves, the other 13 topics within the topic chain somewhat tried to maintain a balance between the similarities and gaps. This era saw higher number of topics becoming matured enough to stand independently, while only 6 topics remained immature in nature. This observation leads to a conclusion that the CHI proceedings during this decade discussed more papers with random topics than that of previous decade and have lost its topic centrality in comparison. For example, we examined topics 2, 7, 10 and 13 and observed that they are quite different in nature. But, at the same time, these individual topics became matured enough with their individual distributions around 5% in a model of 20 topics (healthy average distribution should be 5%).

Plotting Figure 15, we have surprisingly observed a well proportionate prevalence of many important HCI research areas within the topics. We saw that UI/UX design, interaction design, system performance, usability study, mobile technology, social network/media, human factors, cognitive science, social & psychology etc. have got distributions between 6% - 8%. We observed that system performance gained huge priority in this era. We also saw that mobile technology and social media that were deprived in the first two decades, became excessively prevalent in this decade, a huge shift in research focus. We also found that AI and Data Science, that abruptly faded away in the previous decade, became prevalent in this decade. Engineering and ethnography continued to become the deprived research areas. However, we observed the first introduction information security during this era. Software and application development lost its focus to the lowest level with its prevalence rank 16th and crowd work lost its rank to 15th in the current era.

These observations go in line with our previous understanding. The huge technological advancements in mobile and internet technologies in 21st century led to a huge shift in focus in mobile technology and social network/media. Along with that, we saw that the topics became sufficiently matured to discuss major HCI research areas proportionately, but at the same time the published papers discussed comparatively more random topics, a reason for loss of topic centrality. The advancement of internet technology made data collections and surveys/case studies much easier, which is why we saw a growing popularity in social aspects of HCI research. At the same time, technological advancement explains the growing amount of performance considerations in the HCI research. The more prevalence of data and information on the internet made the work of information security more relevant as part of HCI research. Software and application development lost focus further because of the already said reason but having a lowest prevalence rank may be a matter of worry to the research community. The loss of focus in crowd and collaborative work in this era of mobile and internet has also become our reason for surprise. Finally, we have also shown (Figure 16) few notable works covered during this topic era.

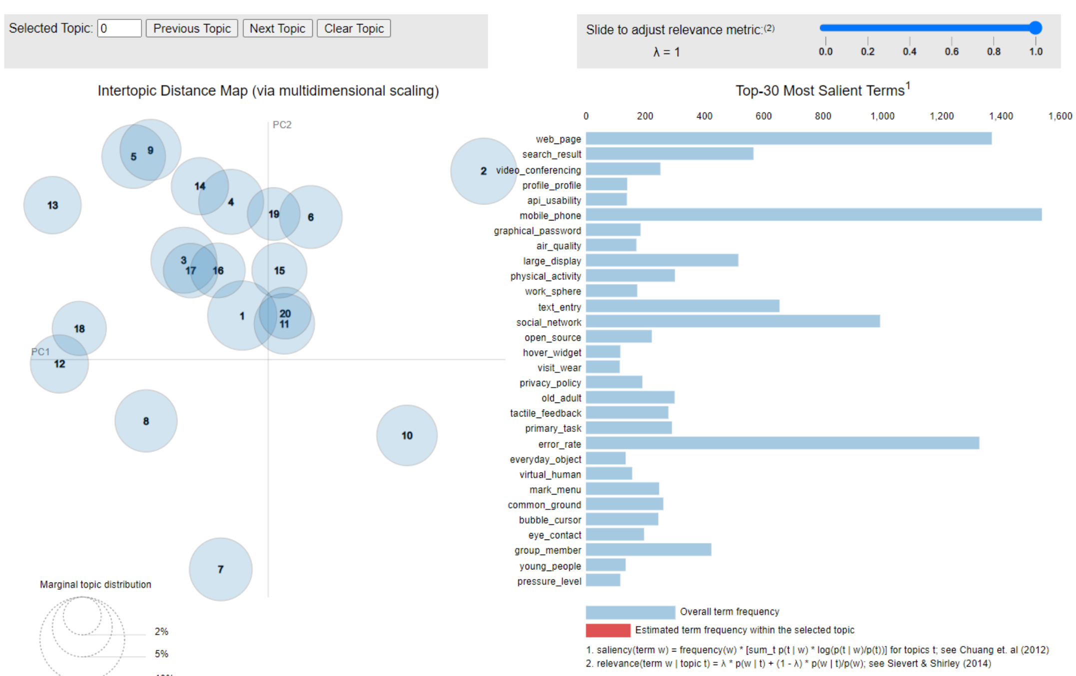

Fig 14: Bigram topic model for the topic era 2001 – 2010

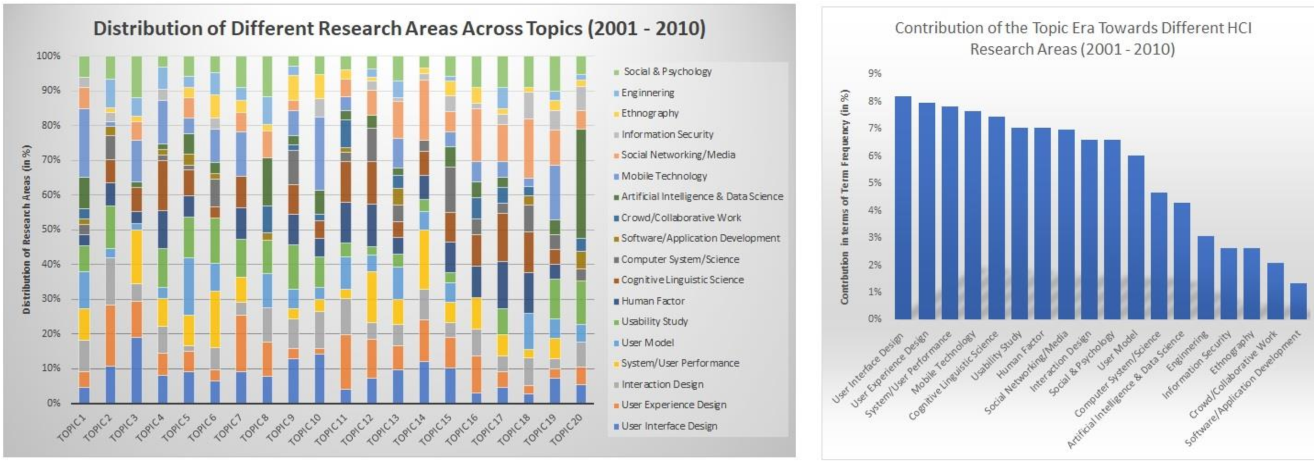

Fig 15: Prevalence of difference research areas towards different topics vs average contributions of 16 HCI research areas in topic era 2001 – 2010

Fig 16: Few notable works during topic era (2001 - 2010)

4.4.4 Analysis of Topic Model during 2011 – 2018

We have now shown the bigram topic model (Figure 17) for 4th decade (2011 – 2018). Analyzing the characteristics of the topic model during this decade, we have observed that topics got even more scattered as compared to the previous decade, leaving much fewer overlapping topics (only 8 of them) than those of previous decade. This scattered nature of topics broadened the knowledge gaps between the topics. We saw a small cloud of 6 overlapping topics trying to accumulate the knowledge, but the relative positions of other scattered topics indicate a significant failure to balance the similarity and gaps between topics. We could conclude that the topics in this era are quite random in nature with the underlying research papers discussing random subjects. However, we could also observe that most of these topics (more than 90% of them) became sufficiently matured with their average distribution of 5% in the corpus. So, even though the topics in this era have lost their centrality, they became sufficiently matured to stand independently.

Plotting Figure 18, we observed a quite similar trend of topics in this era as compared to the previous decade because we could see proportionate distributions (varying between 6% - 9%) of different important research areas within all topics. Engineering and ethnography still remained the quite unexplored areas of CHI proceedings. Software/application developments again remained another deprived section in this era. We have, however, seen that performance factors ranked quite low this time even though receiving enough attention. Computer systems lost significant rank here, while crowd/collaborative work gained importance within this era.

These observations go hand in hand with our understanding so far. This topic era continued with the trend of the previous one, letting the knowledge more scattered across the topic space, making the research topics even more random in nature and broadening the knowledge gaps, while at the same time, developing each topic sufficiently large, matured and independent. The continued technological advancements claimed the predominant focus in mobile and internet technology, but at the same time the research papers became more and more comprehensive, taking the holistic approach to discuss most of the HCI factors under specific research context. We have seen that engineering and ethnography have been the most unexplored areas in CHI proceedings year after year and, at the same time, computer system/science and software/application developments systematically have lost their research focus in the later decades. Finally, we have also shown (Figure 19) few notable works covered during this topic era.

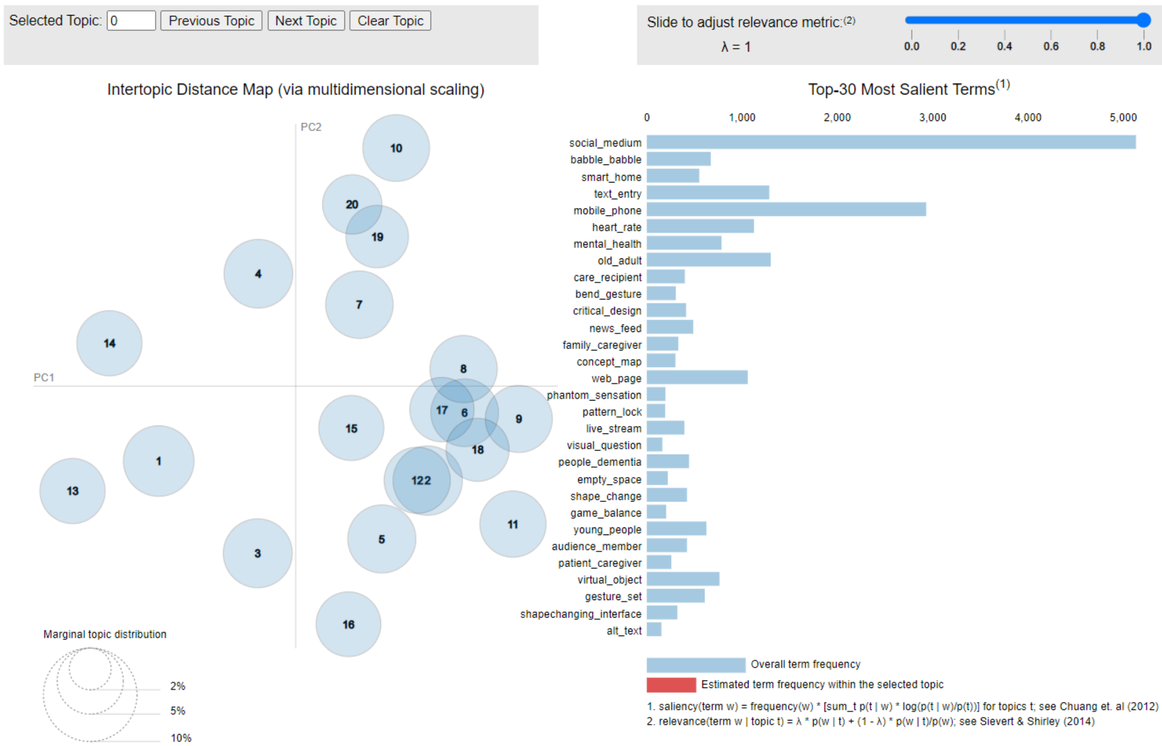

Fig 17: Bigram topic model for the topic era 2011 – 2018

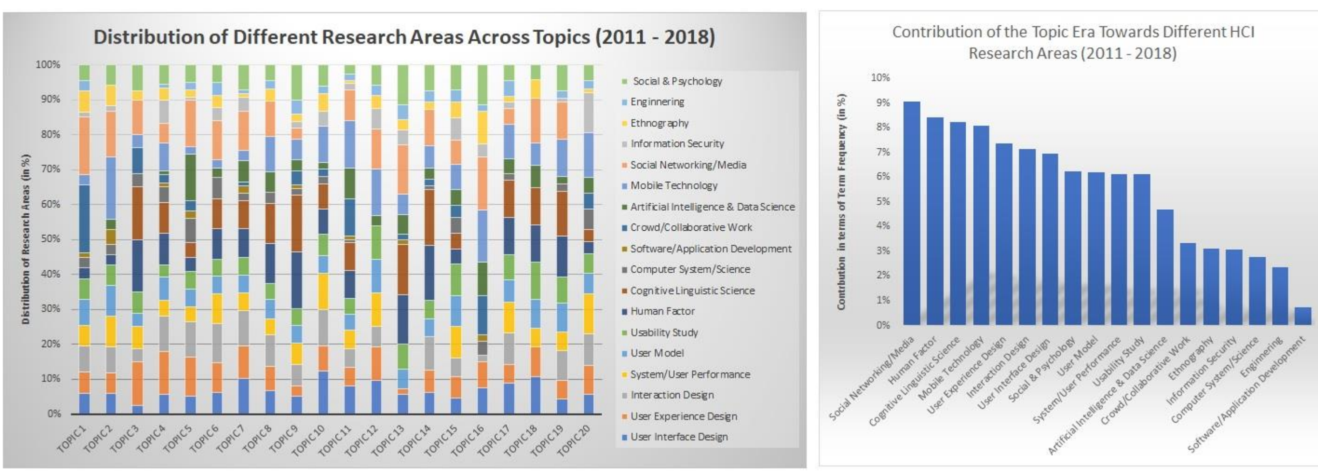

Fig 18: Prevalence of difference research areas towards different topics vs average contributions of 16 HCI research areas in topic era 2011 – 2018

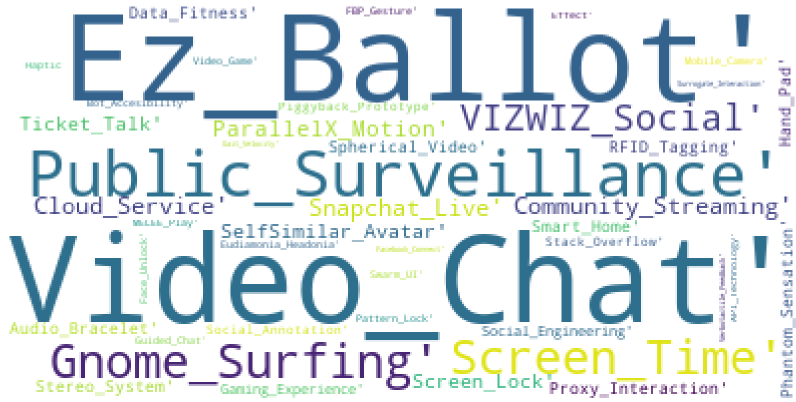

Fig 19: Few notable works during topic era (2011 – 2018)

5. CHAPTER 5: RESEARCH OUTCOME AND CONCLUSION

5.1 Introduction

In this chapter we are aimed to discuss our research outcomes based on our methods and observations. We would like to discuss this by summarizing our observations in answering our research questions. We will then proceed to discuss few of our research limitations and scope of future works and finally we will conclude. So, the discussion points of this chapter are mentioned below.

1. Answering the research questions
2. Research limitations
3. Scope of future works
4. Conclusion

5.2 Answering the Research Questions

In our research introduction (Chapter 1), we outlined few research questions that we were committed to answer. These questions are in line with our research objective, which is finding the trends and gaps in CHI proceedings for1982 – 2018. The answers are as follows.

5.2.1 What did CHI conferences offer between 1982 and 2018?

Because the scope of this question was to understand the trends in the topics discussed in CHI proceedings and their contributions towards 16 major HCI research areas, this question should be answered by answering the below two sub-questions.

5.2.1.1 What were the trends and techniques in CHI conferences during 1982 - 2018?

In the previous chapter, we have discussed the trends in CHI proceedings for four consecutive decades mainly from three different perspectives, which are nothing but finding trends in CHI publications, trends at the overall corpus level and trends at each topic level. As part of answering this question, we will summarize and combine these three points of views. The 1st decade (1982 - 1991) started with only 447 publications and hence the scope of CHI proceedings was very limited. It was the era of primitive computing and the technologies were under their initial development stages, so the focus of HCI was mainly to develop computer systems or interface/interactions technologies itself rather than building impressive interfaces. During this time, the research topics were mostly at the budding stages and most of them were very similar to each other. So accordingly, topics were narrowly focusing on very few major CHI research areas like UI/UX designs, usability studies and computer/system designs, while the other important HCI research factors were either discussed little or never discussed. The less prevalence of interactive computers became a hinderance to interface designs, however, surprisingly, some advanced computing topics such as AI, data science were under scope of discussion as part of future scope of research. Few notable works during this time included Command-based interface design, Manual based usability study, DBMS Developments, Windows OS, Heuristics Evaluation etc.

The next era (1992 - 2000) grabbed the pace of HCI research with 681 publications. The scope of CHI proceedings started to get expanded because of technological advancements. During this era, the computer technology started to advance, and the scope of CHI proceedings also increased. This era, because of the introduction of graphical interfaces, had excessively huge focus mainly into UI/UX designs and their usability studies. The topics were somewhat dissimilar to each other as compared to the previous decade, but we could also see many similar topics that are essentially on UI/UX designs. We furthermore observed that the individual topics became more matured on an average than the previous era. We saw that the research papers started talking in a bit broader HCI perspective, somewhat considering the areas such as interaction design, social & psychology, human factors, cognitive science, system performance etc. Few notable works in this era included Traversable Interface design, On Demand interface design, Gom's Model, Fitt's Law, Graphic design etc.

The inception of 21^{st} century made a huge change in focus of HCI research. The introduction of mobile, internet and advanced computing technologies shifted the complete horizon of CHI conferences. We saw that the 3^{rd} decade welcomed 1521 CHI publications, around 123% more papers than the previous decade. This era was highly focused to research mobile, web and internet technologies. Social networking also received huge attention during this time. The CHI topics during this decade started talking much more different things, but at the same time, most of the topics became sufficiently matured to become themes or theories. The topics started taking about diversified HCI research areas such as social science, psychology, human factors, cognitive sciences, AI and Data science etc. Advancement in mobile and web technology brought huge attention to system or user performances and user/case studies became much easier through data collections; so, we found an increasing popularity in social aspect of HCI. This prevalence of data gave much consideration to information and data security. Few notable works in this decade included Smart Phone technology, Storytelling Alice, Touch sensor, Search Engine, Virtual Reality etc.

The final era has been observed to follow the similar trend of the previous decade. The number of publications further increased to 3929. We could see that similar research areas such as mobile, internet, social networking etc. were trending in this decade; however, they were discussed at much more advanced level and from the perspective of finest user experience. In this era, most of the topics became even more dissimilar to each other, but at the same time almost all the topics became sufficiently matured to become interdisciplinary. The published papers continued to discuss the topics from holistic perspective, covering most of the HCI factors with proportionate amount of importance. However, they were going into much deeper sides of factors. For example, case studies were not limited to different types of users but were considering their individual behaviors. Tools or technologies were being designed with the considerations of much advanced level of cognitive or human factors. AI and data science became trending topics in this era with the increasing popularity of computer gaming. Few notable works included Video chatting, Gnome surfing, Public surveillance systems, Game development, Community streaming etc.

5.2.1.2 What were the notable contributions and addition of knowledge in CHI conferences happened during 1982 - 2018??

Throughout the duration of 37 years, CHI conferences made their footprints towards many notable contributions. Firstly, throughout the four decades, HCI has grown from a very tiny and conservative discipline to a very vast and interdisciplinary field. When we observed the topic trends, we saw that the topics were immature in nature during the initial period, while as the time progressed, CHI topics became more matured and expanded, talking about certain subject matters from different perspectives. For example, we saw that during the initial days of computing, topics were discussing limited areas of HCI, while with the progress of time, they were viewed from holistic perspectives. On the similar line, an interface design was not only limited to the use of fancy graphics, but also it was considered from different uses' perspectives, different human factors (such as psychological and cognitive/intellectual sides), and different performance and usability perspectives etc. We found that HCI has huge contributions towards UI/UX designs and usability studies. From year after year, this side of HCI research has become continuously stronger and systematically developed. We observed that, from very initial level of command prompts, HCI has progressed towards augmented reality, smart home and advanced gaming technology. Furthermore, since the beginning of 21^{st} century, CHI conferences made much contributions towards mobile and internet technologies. From the very initial level of mobile technologies such as calling technology, ringtones, display etc., we progressed towards much advanced level of smart phone technologies such as touch screen, face recognition, pattern lock, touch sensors etc. We also advanced from the websites and webpage designs to much higher standard internet and social networking technologies such as search engine, live streaming, online tutorials, news feeds etc. We also made huge progress towards the usage of various terms in HCI. Before, 21^{st} century the term "user" resembled to user interfaces, user interactions or user experiences, while after 21^{st} century, the term "user" was understood further to user case studies, user behavior and other social and intellectual perspectives. So, in conclusion we can firmly say that HCI research not only made huge contributions to itself to broaden it as a discipline, but also enriched other disciplines related to computer, social, psychological and cognitive sciences.

5.2.1.2.1 Can we offer a quantitative metric to calculate the addition or loss of knowledge within a corpus?

At this stage, with our current approach, it is difficult to answer a yes or a no to this question. Because of the nature of the topic modeling we have used (LDA in this case) here, we cannot quantify the impact of the addition and subtraction of a topic within a system and henceforth, the amount of knowledge being added to or subtracted from a topic model is unquantifiable at this stage. However, in machine learning, we have "Entropy" or "Information Gain" theory (Gray, 2011), that can effectively quantify the gain of information with the addition of a feature within the training set. Here, in our case, columns of the document vectors used for LDA are effectively the features. So, introducing Entropy theory may give us an opportunity to quantify the gain of information with the addition of a document within a text corpus. Furthermore, we also have an existing research work (Ximing Li, 2018) that have used this Entropy theory within LDA topic modeling to provide weights to certain words that are relatively more important for a topic, extending our horizon from the document level to a topic level. So, this provides us a ray of hope that we may be able to answer this question as part of our future work.

5.2.2 What were the holes or gaps in knowledge produced in CHI conferences during 1981 - 2018?

There is a say that success comes with criticism, as is the case for CHI proceedings without an exception. Kostakos's (2015) study determined that CHI proceedings systematically produced knowledge gaps using co-word analysis. Our study validated the same result to some extent. Using topic modeling, we saw that CHI proceedings systematically produced gaps of knowledge within the topics. According to our observations in the four bigram topic models shown above, during the four consecutive decades, the topics became more separated from each other, continuously increasing the gaps between themselves. During the last two decades, the topics became more random in nature, indicating that the CHI publications mostly picked up random subjects as research topics. Furthermore, while analyzing the unigram models, we observed continuous disappearance of most popular terms in consecutive decades. We observed that second, third and final decades have lost 63%, 50% and 43% of the most popular words that appeared during their respective previous decades. This indicates abrupt shifts in research focus over consecutive topic eras. We observed that the focus during the 1st era was to develop basic computer technologies, while the 2nd era was biased with its focus heavily into UI/UX designs. The inception of 21st century snatched the complete CHI research attention into developing gadgets and internet technologies. Even though last two decades maintained similar trends, the random nature of publications resulted in disappearance of 43% popular words. So, all these results validate one fact that CHI proceedings have systematically lost their research centrality and thus, producing systematic knowledge gaps over the years.

5,2.2.1 What were the problems in CHI conferences left unsolved over time and the possible reasons behind?

While analyzing the knowledge gaps during four consecutive decades we also aimed to find problems that remained persistent over time. In doing so, firstly we observed that engineering and ethnography were the two research areas systematically deprived in CHI conferences. Even though, we saw that, in later two decades, the topics discussed a little bit from these two research areas, they never became important areas of considerations, but just a tiny area within some specific research contexts. An explanation against this could be that these two fields have never been the core CHI research areas and would have always been a matter of introduction under specific contexts. We further saw that another two areas such as computer system/science and software application/developments have systematically lost their focus. During the initial days, the computer technology was scarce, which is why CHI publications during their 1st decade felt the need the research focuses on developing computer system. However, during the later decades, when the computer systems became sufficiently developed, CHI publication did not have the need to continue the research in this aspect, which is why this area of research lost its focus. Similarly, HCI lost its focus towards software and application developments because its excessive focus towards user centric interface/experience designs over the decades, thus the core area of software engineering remained an unexplored area within this discipline.

5.2.2.2 Was there any attempt made to address any persistent gaps?

While we saw the persistent gaps in HCI research being unsolved over time, we also saw few nice attempts to bride any gaps appeared. We observed that interaction design, mobile technology and social networking remained underdeveloped during the first two decades but started getting heavy attention during the last two decades. It is much evident that interactive computers and applications were not prevalent before the 21st century and mobile technology & social networking were moderately introduced only during that time, which suffice the reason behind these fields to not have introduced during first two decades. But, more interestingly, CHI conferences took this nice opportunity during 2000 – 2018 to research those areas, an example of addressing the research gaps. Furthermore, with similar reason, system/user performance was a deprived area of HCI research before 21st century because of the limitations in computer technology. However, with the introduction of fast computers during 21st century, this area became much explored in order to improve the user experience and their interactions with systems/applications. We found that AI and Data Science was a scope of discussion during the 1st era but lost its focus during the second decade because of the biased attempt of CHI conferences to design better interfaces. However, with the beginning of information era during 21st century, it again regained its lost focus. In similar line, information security and social & psychology were two barely explored areas during the 1st two decades. They gained their priorities during the final two decades because of the increasing advancement of internet technologies to exercise user or case studies and abundance or data over web explains the increasing need to explore information security during 21st century.

5.3 Research Limitations

After discussing our research outcomes, it's time to discuss about two of our research limitations. Firstly, after building the topic models we had resorted to a manual approach to correctly classify the relevant terms describing the topics under one or more of the above discussed 16 research areas in order to understand the topic contexts. This manual classification technique took around 50% of our research effort and might be error prone for a person who is not an HCI field expert. Secondly, we have built the cost function for prevalence metric entirely based on the term frequencies of different terms. We have used this derived metric to calculate the relative prevalence of topics or terms towards one or more major HCI research areas in order to measure the contribution of that topic towards those research areas. However, considering LDA and LDAvis both as probabilistic models, we have somewhat discouraged the possibility of our prevalence metric to be probabilistic in nature.

5.4 Scope of Future works

Our work has a handful of scopes of future developments. Firstly, we would like to see, if we can come up with an automatic approach for classifying any term (unigram/bigram) under one or more of 16 CHI research areas. For that we would like to use our currently classified data set as the base training set. We would then plan to implement few resampling techniques such as Bootstrap or Cross Validation (Kohavi, 1995) in combination with a supervised machine learning technique to train on the data and validate the accuracy of the classification method. We will then look forward to implementing this data model to make prediction on some unseen data to judge the test accuracy, which will be an automated approach. Secondly, we would see, if we can come up with a probabilistic cost function for the current prevalence metric that we have used to quantify the research contribution of a topic towards various HCI research areas. Finally, continuing to answer the question 5.2.1.2.1, we would like to come up with a quantitative metric that can help us to estimate the information gain with the addition of a topic to and subtraction of a topic from a topic system in order to understand how much new knowledge is contributed by a new topic within the topic system. This will help us understand the literary contribution of a topic in a much deeper level.

5.5 Conclusion

As part of this study, we have analyzed the trends of CHI conferences for the past 37 years. We have also investigated the gaps in the CHI proceedings that appeared during the same timeframe. But, with all this information, we cannot conclude whether HCI has been a healthy scientific discipline over these years, which Kostakos threw a light upon in his research. There are two reasons behind this fact. Firstly, Kostakos used some approved metrics and techniques to do this health check that are already proven as standard techniques in other fields. Secondly, engaging into such a strong discussion on whether a discipline is healthy or not needs sufficient field knowledge (in this case it is the HCI knowledge) and years of research experience in the same discipline. And unfortunately, I don't possess any of them. However, our research takes itself way beyond the scope of "The Big Hole is HCI Research" (2015). It offers a systematic and comprehensive method of trends and gaps analysis for CHI proceedings using a completely new approach. We see our contribution lying in that, the rest is based on our readers' perspective.

BIBLIOGRAPHY

A. Alahmadi, A. J. (2013). A new text representation scheme combining Bag-of-Words and Bag-of-Concepts approaches for automatic text classification. *7th IEEE GCC Conference and Exhibition (GCC)* (pp. 108-113). Doha: doi: 10.1109/IEEEGCC.2013.6705759.

An, X. a. (2011). Q.Q. Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics 88, 1*, 133–144.

An, X. W. (2011). Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics 88, 1 (2011)*, 133–144.

Barkhuus, L. &. (2007). From Mice to Men – 24 years of Evaluation in CHI. *CHI '07 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* San Jose, California, USA: alt.chi.

Biglu, M.-H. a. (2016). Subject-based information retrieval system in digital libraries. *Journal of Software Engineering and Applications, Vol. 09 No. 7*, 346-352.

Biplab Ketan Chakraborty, D. S. (2017). A Review of Constraints on Vision-based Gesture Recognition for Human-Computer Interaction. *IET Research Journals.*

Blei, D. (2012). Probabilistic topic models. *Communications of the ACM, Vol. 55 No. 4*, 77-84.

Blei., A. J. (2012). Visualizing topic models. *ICWSM.* ICWSM.

Bornholdt, J. R. (2006). Statistical mechanics of community detection. *Physical Review E 74.*

Callon, M. C.-P. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information 22, 2* , 191–235.

Canny, J. (2006). The future of human-computer interaction. *Queue 4, 6*, 24–32.

Carson Sievert, K. E. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63–70). Baltimore, Maryland, USA: Association for Computational Linguistics.

Clement Lee, A. G. (2019). Weaving the Topics of CHI: Using Citation Network Analysis to Explore Emerging Trends. *CHI EA '19 Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA.

Coulter, N. M. (1998). Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science 49, 13*, 1206–1223.

Coulter, N. S. (1998). Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science 49, 13* , 1206–1223.

Daniel Ramage, E. R. (2009). Topic Modeling for the Social Sciences. *NIPS Workshop on Applications for Topic Models: Text and Beyond.*

David M. Blei, A. Y. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, 993-1022.

Dix, A. (2009). Human-Computer Interaction. *In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA*, 41-46.

Finetti., d. (1990). *Theory of probability. Vol. 1-2*. Chichester: John Wiley & Sons Ltd. Reprint of the 1975 translation.

Gabrilovich E. & Markovitch, S. (2007). Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, (pp. 6 - 12).

Gollapalli, S. D. (2014). Document analysis and retrieval tasks in scientific digital libraries. *Information Retrieval: 8th Russian Summer School, RuSSIR 2014* (pp. 3-20). Nizhniy, Novgorod, Russia: Revised Selected Papers, Springer, London.

Gray, R. M. (2011). *Entropy and Information Theory*. Springer.

Grudin, J. (1990). THE COMPUTER REACHES OUT: THE HISTORICAL CONTINUITY OF INTERFACE DESIGN. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)* (pp. 261–268). New York, NY, USA: ACM: http://dx.doi.org/10.1145/97243.97284.

Grudin, J. (2005). Three faces of human-computer interaction. *IEEE Annals of the History of Computing 27, 4*, (pp. 46–62).

Hofmann. (1999). Probabilistic latent semantic indexing. *Proceedings of the Twenty-Second Annual.*

Hu, C. B. (2009). Scientometric Analysis of the CHI Proceedings. *SIGCHI Conference on Human Factors in Computing Systems (CHI '09)* (pp. 699–708). New York, NY, USA: ACM, DOI:http://dx.doi.org/10.1145/1518701.1518810.

Jason Chuang, C. D. (2012). Termite: Visualization Techniques for Assessing Textual Topic Models. *AVI*. AVI.

Justin Snyder, R. K. (2013). Topic Models and Metadata for Visualizing Text Corpora. *NAACL HLT Demonstration Session*. Proceedings of the 2013.

K. Nigam, A. M. (2000). Text classification from labeled and unlabeled documents using EM. *Machine Learning, 39(2/3)*, (pp. 103–134).

KASPER HORNBÆK, A. M. (2019). What Do We Mean by "Interaction"? An Analysis of 35 Years of CHI. *ACM Trans. Comput.-Hum. Interact. 26, 4, Article 27*. https://doi-org.ezproxy.lib.monash.edu.au/10.1145/3325285.

Kaye, J. '. (2009). Some Statistical Analyses of CHI. *Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)* (pp. 2585–2594). New York, NY, USA: ACM.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *The International Joint Conference on Artificial Intelligence*.

Kong, X. M. (2018). VOPRec: Vector Representation Learning of Papers with Text Information and Structural Identity for Recommendation. *IEEE Transactions on Emerging Topics in Computing*. IEEE.

Kostakos, V. (2014). CHI 1994–2013:Mapping two decades of intellectual progress through co-word analysis. *Proc. CHI 2014* (pp. 3553–3562). New York: ACM.

Kostakos, V. (2015). The Big Hole in HCI Research. *Interactions Homepage archive Volume 22 Issue 2*, 48-51.

Landauer, T. D. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104.

Lund, K. B. (1995). Semantic and associative priming in a high-dimensional semantic space. *Cognitive Science Proceedings (LEA)*, (pp. 660-665).

Lund, K. B. (1996). Producing highdimensional semantic spaces from lexical cooccurrence. *Behavior Research Methods, Instruments & Computers, 28(2)*, (pp. 203-208).

Marc, L. (1991). The dynamics of research in the psychology of work from 1973 to 1987: From the study of companies to the study of professions. *Scientometrics 21, 1*, 69–86.

Matthew J. Gardner, J. L. (2010). The topic browser: An interactive tool for browsing topic models. *NIPS Workshop on Challenges of Data Visualization.*

Matveeva, I. L. (2005). Generalized latent semantic analysis for term representation. *In Proc. of RANLP*.

McGill, S. a. (1983). *Introduction to Modern Information Retrieval.* McGraw-Hill.

Michael Nokel, N. L. (2015). Topic Models: Accounting Component Structure of Bigrams. *Proceedings of the 20th Nordic Conference of Computational Linguistics*. NODALIDA 2015.

Michael, C. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information 22, 2* , 191–235.

Myers, B. A. (1998). A Brief History of Human-computer Interaction Technology. *interactions 5, 2 (March 1998), DOI:http://dx.doi.org/10.1145/274430.274436*, 44–54.

Nicola Dell, N. K. (2016). The Ins and Outs of HCI for Development. *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2220–2232). https://doi.org/10.1145/2858036.2858081.

Okoli, C. a. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems. *Sprouts: Working Papers on Information Systems 10 (26)*, 10-26.

Quinn, A. J. (2011). Human computation: a survey and taxonomy of a growing field. *In Proc, of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM (2011)*, (pp. 1403-1412).

Ravikumar, S. A. (2015). Mapping the intellectual structure of scientometrics: a co-word analysis of the journal scientometrics (2005-2010). *Scientometrics, Vol. 102 No. 1.*

Reeves, S. (2015). Locating the 'big hole' in HCI research. *Interactions*, 25 June 2015, Vol.22(4), pp.53-56.

Retrieval, I. t. (2009). *Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze*. Cambridge, England: Cambridge University Press.

Robert Tibshirani, T. H. (2013). An Introduction to Statistical Learning. In T. H. Robert Tibshirani, *Chapterb10: Unsupervised Learning* (pp. 380 - 385). New York Heidelberg Dordrecht London: Springer DOI 10.1007/978-1-4614-7138-7.

Rowley, J. &. (2004). Conducting a literature review. *Management Research News 27(6)*, 31-39.

S. Deerwester, S. D. (1990). Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6):391–407.

Sedighi, M. (2016). Application of word co-occurrence analysis method in mapping of the scientific fields (case study: the field of informetrics). *Library Review, Vol. 65 Nos 1/2*, 52-64.

Stackoverflow. (2020, 07 01). *nlp - Expanding English language contractions in Python - Stack Overflow*. Retrieved from Stack Overflow: https://stackoverflow.com/questions/19790188/expanding-english-language-contractions-in-python

Wael H. Gomaa, A. A. (2013). A Survey of Text Similarity Approaches. *International Journal of Computer Applications (0975 – 8887)*.

Wang, C. a. (2011). Collaborative Topic Modeling for Recommending Scientific Articles. *ACM*, (pp. 448-456). New York, NY.

Whittaker, S. T. (2000). Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI. *Journal Human-Computer Interaction Volume 15 Issue 2*, 75-106.

Ximing Li, A. Z. (2018). Exploring coherent topics by topic modeling with term weighting. *Elsevier Ltd.*