Visual Text Analytics for Online Conversations

by

Md Enamul Hoque Prince

B.Sc., Chittagong University of Engineering & Technology, 2007

M.Sc., Memorial University of Newfoundland, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Computer Science)

The University Of British Columbia
(Vancouver)

May 2017

Abstract

With the proliferation of Web-based social media, asynchronous conversations have become very common for supporting online communication and collaboration. Yet the increasing volume and complexity of conversational data often make it very difficult to get insights about the discussions. This dissertation posits that by integrating natural language processing and information visualization techniques in a synergistic way, we can better support the user's task of exploring and analyzing conversations. Unlike most previous systems, which do not consider the specific characteristics of online conversations; we applied design study methodologies from the visualization literature to uncover the data and task abstractions that guided the development of a novel set of visual text analytics systems.

The first of such systems is ConVis, that supports users in exploring an asynchronous conversation, such as a blog. ConVis offers a visual overview of a conversation by presenting topics, authors, and the thread structure of a conversation, as well as various interaction techniques such as brushing and linked highlighting. Broadening from a single conversation to a collection of conversations, MultiConVis combines a novel hierarchical topic modeling with multi-scale exploration techniques. A series of user studies revealed the significant improvements in user performance and subjective measures when these two systems were compared to traditional blog interfaces.

Based on the lessons learned from these studies, this dissertation introduced an interactive topic modeling framework specifically for asynchronous conversations. The resulting systems empower the user in revising the underlying topic models through an intuitive set of interactive features when the current models are noisy and/or insufficient to support their information seeking tasks. Two summative studies suggested that these systems outperformed their counterparts that do not support interactive topic modeling along several subjective and objective measures.

Finally, to demonstrate the generality and applicability of our approach, we tailored our previous systems to support information seeking in community question answering forums. The prototype was evaluated through a large-scale Web-based study, which suggests that our approach can be adapted to a specific conversational genre among a diverse range of users.

The dissertation concludes with a critical reflection on our approach and considerations for future research.

Lay Summary

Since the rise of social-media, an ever-increasing amount of conversations are generated. Often many people contribute to the discussion, which become very long with hundreds of comments, making it difficult for users to get insights about the discussion. This dissertation integrates language processing and visualization techniques to support the user's task of exploring and analyzing conversations. Language processing mines topics and opinions from the conversations, while visualization techniques provide visual overviews of the mined data and support user exploration and analysis. User studies revealed significant improvements, when our systems were compared to traditional blog interfaces. This dissertation also introduces a new human-in-the-loop algorithm that helps the user to revise results of topic modeling. Two user studies show that these systems outperform their non-interactive counterparts. Finally, we tailored our previous systems to support information seeking in community question answering forums. The prototype was successfully evaluated through a large-scale user study.

Preface

Parts of this thesis are based on prior peer-reviewed publications by me (under the name Enamul Hoque) and with various coauthors:

Chapter 2 is based on the article *ConVis: A visual text analytic system for exploring blog conversations*, by Enamul Hoque and Giuseppe Carenini; in Journal of Computer Graphics Forum (Proceedings of EuroVis), 33(3):221230, 2014 [59]. My main contributions include: 1) user requirements analysis; 2) iterative design and development of prototypes; 3) designing and conducting the user study; 4) preparing the manuscript. Giuseppe Carenini advised me throughout the project and contributed to the editing process.

Portions of Chapter 2 also appeared (along with portions of Chapter 1) in a summarized form in *Interactive exploration of asynchronous conversations: Applying a user-centered approach to design a visual text analytic system*, by Enamul Hoque, Giuseppe Carenini, and Shafiq Joty; in Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces (ILLVI), 2014 [62].

A version of Chapter 3 has been published as *MultiConVis: A visual text analytics system for exploring a collection of online conversations*, by Enamul Hoque and Giuseppe Carenini; in Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI), pp. 96-107, 2016 [60]. My main contributions include: 1) user requirements analysis; 2) iterative design and development of prototypes; 3) designing and conducting the user study; 4) preparing the manuscript. My co-author Giuseppe Carenini played supervisory roles and contributed to the editing process.

Portions of Chapter 4 were published in *ConVisIT: Interactive topic modeling for exploring asynchronous online conversations*; Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI), pp. 169-180, 2016 [60]. An extended version of this paper has also appeared as a journal paper: *Interactive topic modeling for exploring asynchronous online conversations: Design and evaluation of ConVisIT*, by Enamul Hoque and Giuseppe Carenini; ACM Transactions on Interactive Intelligent Systems (TiiS), 6(1):7:17:24, Feb. 2016 [61]. My main contributions include: 1) proposing and implementing the interactive topic modeling approach; 2) iterative design and development of prototypes; 3) designing and conducting the user study; 4) preparing the manuscript. Giuseppe Carenini played supervisory roles during the design and evaluation of the system. He also contributed to the editing process.

Portions of Chapter 5 were published in *CQAVis: Visual text analytics for community question answering*, by Enamul Hoque, Shafiq Joty, Lluís Màrquez and Giuseppe Carenini; in Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI), 2017 [63]. Most of this research was done during my internship at the Qatar Computing Research Institute between January 2016 and May 2016. My main contributions include: 1) user requirements analysis; 2) iterative design and development of prototypes; 3) designing and conducting the user study. My co-authors provided feedback during the design and evaluation of the system. Shafiq Joty and Lluís Màrquez helped me in establishing the collaboration with Qatar Living administrators, advertising for the user study and deploying the tool online. I performed the majority of the writing for the paper, while all of my collaborators contributed to the editing process.

All the user studies described in Chapter 2, 3, and 4, were conducted with the approval of the UBC Behavioural Research Ethics Board (BREB): certificate H13-02132.

Table of Contents

List of Tables

List of Figures

Acknowledgments

First of all, I would like to express my deepest gratitude to my advisor, Giuseppe Carenini, for his enormous support throughout my Ph.D. study. As a great advisor, Giuseppe demonstrated to me how research should be done, provided insightful advice, and most importantly constantly challenged me to do better.

I would like to sincerely thank my other supervisory committee members Tamara Munzner and Raymond Ng, for their valuable support on my research and helpful comments and suggestions.

Many thanks go to my external examiner Sheelagh Carpendale for her constructive and insightful feedback on my thesis.

I thank my university examiners Ronald Garcia and Victoria L. Lemieux, as well as my thesis defense chair Cay Holbrook.

I thank all the current and past members of our Natural Language Processing and Intelligent User Interfaces research groups at UBC for the feedback and support they provided. I am especially indebted to Shafiq Joty, for his continuous support throughout my graduate student life. I also thank attendees of Tamara Munzner's InfoVis reading group, who provided valuable feedback and support on my research.

I thank my collaborators for providing valuable help during a design study, which I conducted at the Qatar Computing Research Institute (QCRI): Lluís Màrquez, Alberto Barrón-Cedeño, Giovanni Da San Martino, Alessandro Moschitti, Preslav Nakov, and Salvatore Romeo. I also thank Qatar Living, and especially Ash Bashir, for their help in deploying our research prototype in their website for running a web-based user study.

I would like to thank my M.Sc. supervisors Orland Hoeber and Minglun Gong, who inspired me to pursue my doctoral degree.

Last but not the least, I would like to thank my family for all their unconditional love and encouragement: my parents A K Fazlul Haque and Nargis Ashara Khanam, my brother Nazmul Huq, and my sister Fauzia Jahan. Many thanks to my beloved wife Farzana Afroze, whose support during the final stages of this Ph.D. is so appreciated.

Chapter 1

Introduction

Since the internet revolution and the subsequent rise of social media, an ever-increasing amount of human conversations are generated in many different modalities [17]. While email remains a fundamental way of communicating for most people, other conversational modalities such as blogs and microblogs have quickly become widely popular. These conversations are primarily asynchronous in nature, where participants communicate with each other at different times.

People engaged in asynchronous conversations to exchange ideas, ask questions, and comment on daily life events. Often many people contribute to the discussion, which can quickly become very long with hundreds of comments. The net result of this phenomena is that an enormous and growing volume of conversational data is generated everyday. Recent statistics from Alexa's Internet traffic rating service reveal that the top three blogging and microblogging sites Wordpress, Twitter, and Tumblr are among the top 50 most visited sites in the world [4]. In the Wordpress blogging platform alone, users produce about 80.7 million new posts and 44.5 million new comments each month and over 409 million people view more than 24.2 billion pages in the same period [7]. Conversations in social network platforms also continue to rise at an accelerating pace. A recent study from Pew Research Center shows that 79% of American internet users use Facebook, with roughly three-quarters (76%) of these Facebook users report that they visit the site daily [6].

These collections of online conversations can provide valuable insights in many domains including but not limited to marketing intelligence, business analytics, customer relationship management, journalism, and healthcare analytics. For instance, business analysts may want to analyze text conversations in social media to uncover consumer sentiment and insights about their products or company and to draw conclusions about commercial strategies [93]. In online news media, such as the New York Times, editors are interested to know which of their contributions generate most comments from their readers in order to make strategic decisions about how to balance the content of their online news. They may also want to assess the quality of comments to remove the low quality ones, and to identify high quality contributions to set community standards [32]. As another example, administrators in online health communities are interested in continuously monitoring the forum in order to foster lively discussions, while at the same time they need to prevent the propagation of misinformation and abusive comments [76]. Finally, a casual reader may want to skim through a blog to find out the community response to a particular topic and to decide whether and how she should contribute to the discussion.

While the abundance of conversational data opens up a great opportunity for important discoveries in a variety of domains, exploring and analyzing such large amounts of data has become a challenging problem in both personal and professional contexts. This problem is commonly known as *information overload*, where users feel overwhelmed by the vast amount of potentially relevant information [14]. To address this problem, this dissertation takes a visual text analytics approach, where we combine natural language processing methods for understating and summarizing discussions and information visualization techniques to present an overview of the conversational data to users.

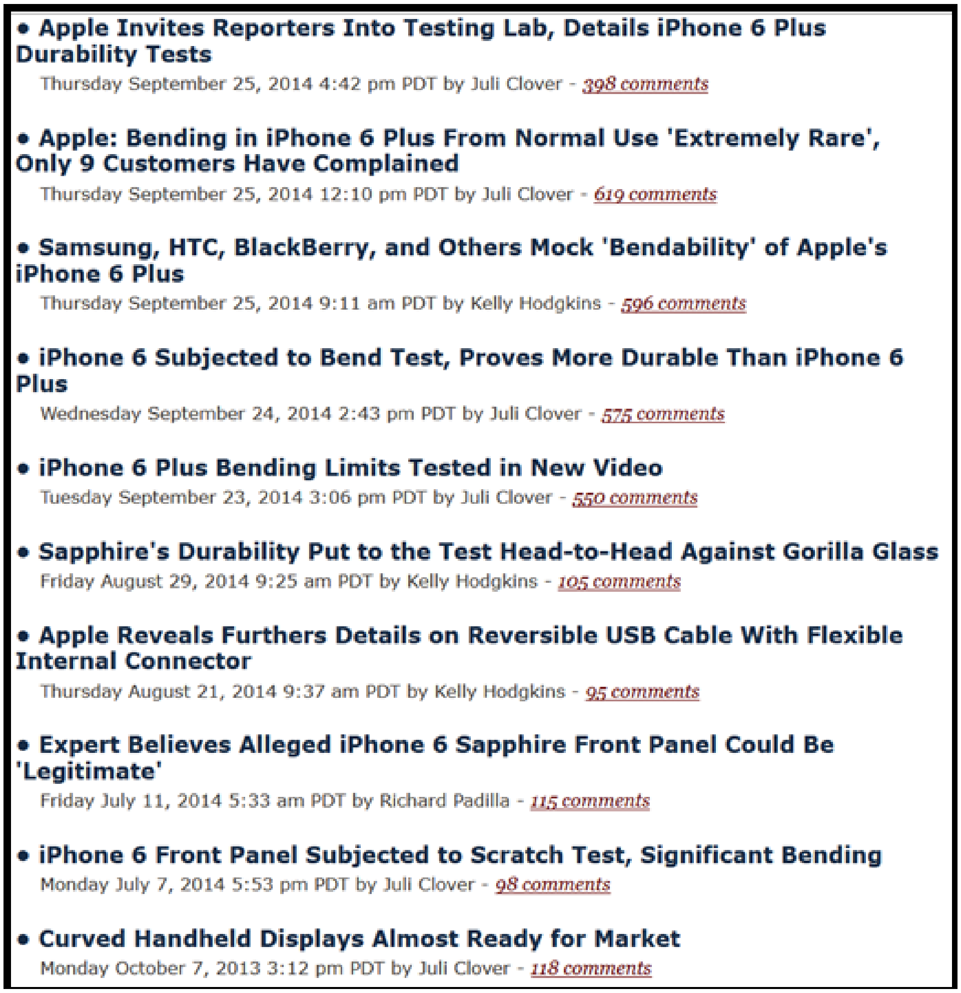
In the remainder of this introduction, we first discuss some key challenges arising from the volume and complexity of conversational data, and the short-comings of existing approaches in dealing with such challenges in Section 1.1. Then, we outline the research methodology used to tackle these challenges (Section 1.2), followed by defining the scope of this work (Section 1.3). Finally, we summarize our major contributions along with an outline of the dissertation in Section 1.4.

1.1 The Problem

An asynchronous conversation such as a blog may start with a news article or an editorial opinion and later may generate a long and complex thread as comments are added by the participants [17]. When a reader wants to explore such a large conversation, traditional social media sites provide very limited support. They simply present the original posts and subsequent replies as a paginated indented list. Thus the reader needs to go through a long list of comments sequentially, until her information needs are fulfilled. Going through such an overwhelming amount of textual data in this way often leads to information overload, i.e., the user finds it very difficult to get insights about the ongoing or past discussions [69]. The problem becomes even more serious when the user is interested in analyzing multiple conversations that are discussing similar issues.

To illustrate the problem, let us consider the issue of 'iPhone bending' that went viral on social media, when the iPhone 6 was launched in September 2014. Soon after the product was released, some people claimed that this new phone can easily bend in the pocket while sitting on it. This incident triggered a huge amount of discussions in Macrumors [1], a blog site that regularly publishes Apple related news and allows participants to make comments. Within a few days, more than a dozen conversations with thousands of comments were generated in Macrumors covering various related issues, such as 'what users reported about the bending issue', 'what Apple says to defend its new product', and 'what are the reactions from the rivals of Apple' etc. In this situation, we could imagine at least three different users who would like to explore this set of conversations. First, a potential customer, who intended to buy an iPhone may want to explore these conversations to verify whether the bending issue is really serious. Second, a journalist may want to publish a story about what people are saying about the 'bending issue'. Finally, an Apple marketing analyst may want to get a pulse from the online community to make an informed decision about how to react to the rumors and possibly redesign the products.

In all three cases, given the large number of conversations/comments, it would be extremely difficult and time-consuming for a user to explore and analyze all this information with the current blog interfaces. This is primarily due to the fact

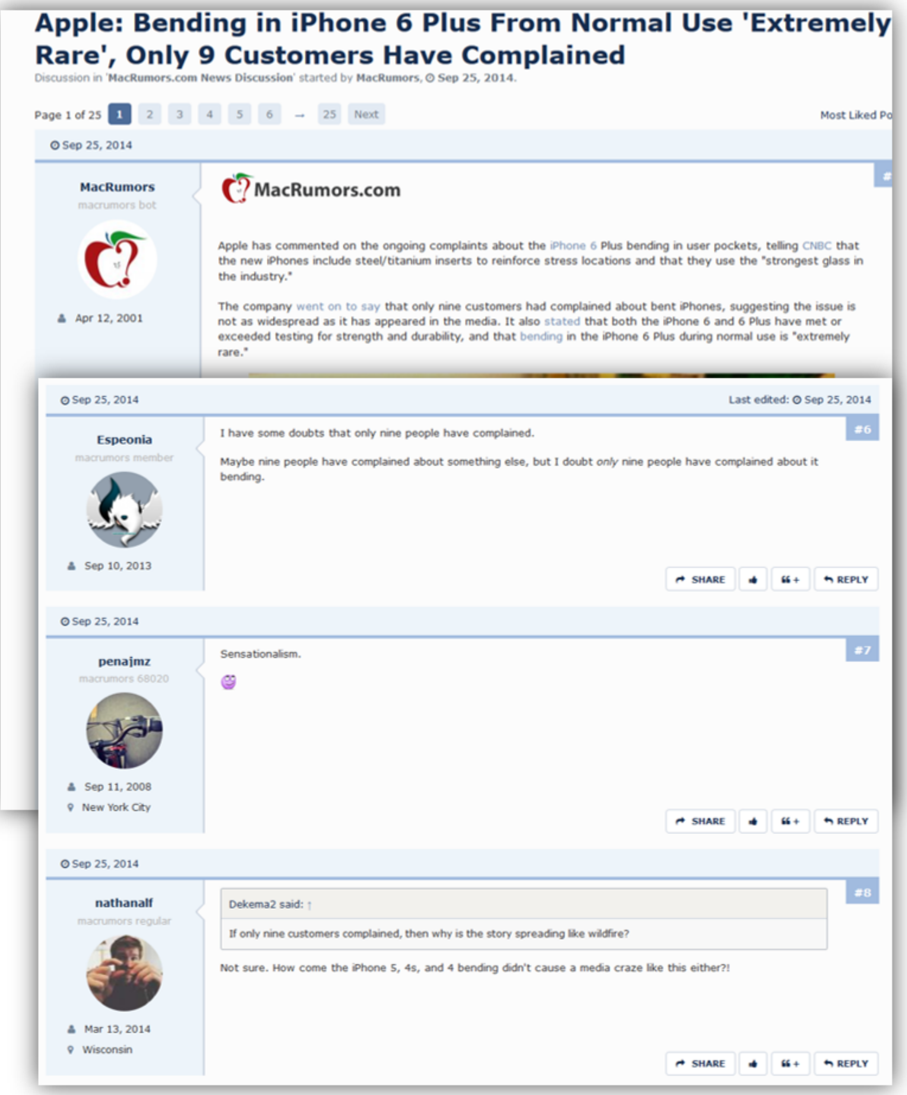


- **Apple Invites Reporters Into Testing Lab, Details iPhone 6 Plus Durability Tests**
  Thursday September 25, 2014 4:42 pm PDT by Juli Clover - 398 comments
- **Apple: Bending in iPhone 6 Plus From Normal Use 'Extremely Rare', Only 9 Customers Have Complained**
  Thursday September 25, 2014 12:10 pm PDT by Juli Clover - 619 comments
- **Samsung, HTC, BlackBerry, and Others Mock 'Bendability' of Apple's iPhone 6 Plus**
  Thursday September 25, 2014 9:11 am PDT by Kelly Hodgkins - 596 comments
- **iPhone 6 Subjected to Bend Test, Proves More Durable Than iPhone 6 Plus**
  Wednesday September 24, 2014 2:43 pm PDT by Juli Clover - 575 comments
- **iPhone 6 Plus Bending Limits Tested in New Video**
  Tuesday September 23, 2014 3:06 pm PDT by Juli Clover - 550 comments
- **Sapphire's Durability Put to the Test Head-to-Head Against Gorilla Glass**
  Friday August 29, 2014 9:25 am PDT by Kelly Hodgkins - 105 comments
- **Apple Reveals Furthers Details on Reversible USB Cable With Flexible Internal Connector**
  Thursday August 21, 2014 9:37 am PDT by Kelly Hodgkins - 95 comments
- **Expert Believes Alleged iPhone 6 Sapphire Front Panel Could Be 'Legitimate'**
  Friday July 11, 2014 5:33 am PDT by Richard Padilla - 115 comments
- **iPhone 6 Front Panel Subjected to Scratch Test, Significant Bending**
  Monday July 7, 2014 5:53 pm PDT by Juli Clover - 98 comments
- **Curved Handheld Displays Almost Ready for Market**
  Monday October 7, 2013 3:12 pm PDT by Juli Clover - 118 comments

Figure 1.1: A set of conversations returned for a query 'iPhone bending' are presented as a paginated list by Macrumors (accessed in March, 2015).

that a typical blog site presents both the list of conversations and the comments as paginated lists and only provides sequential access to conversations/comments (see Figure 1.1 and Figure 1.2). It neither provides any high-level overview of the conversations, nor provides sufficient navigational cues. As a result, users often become overwhelmed by the large amount of conversational data and leave the discussions without fulfilling their information needs [69].

While both the Natural Language Processing (NLP) and the Information Vi-



Apple: Bending in iPhone 6 Plus From Normal Use 'Extremely Rare', Only 9 Customers Have Complained

Discussion in 'MacRumors.com News Discussion' started by MacRumors, Sep 25, 2014.

Sep 25, 2014

MacRumors
macrumors bot
Apr 12, 2001

MacRumors.com

Apple has commented on the ongoing complaints about the iPhone 6 Plus bending in user pockets, telling CNBC that the new iPhones include steel/titanium inserts to reinforce stress locations and that they use the "strongest glass in the industry."

The company went on to say that only nine customers had complained about bent iPhones, suggesting the issue is not as widespread as it has appeared in the media. It also stated that both the iPhone 6 and 6 Plus have met or exceeded testing for strength and durability, and that bending in the iPhone 6 Plus during normal use is "extremely rare."

Sep 25, 2014 Last edited: Sep 25, 2014

Espeonia
macrumors member
Sep 10, 2013

#6

I have some doubts that only nine people have complained.

Maybe nine people have complained about something else, but I doubt *only* nine people have complained about it bending.

Sep 25, 2014

penajmz
macrumors 68020
Sep 11, 2008
New York City

#7

Sensationalism.

Sep 25, 2014

nathanalf
macrumors regular
Mar 13, 2014
Wisconsin

#8

> Dekema2 said: ↑
>
> If only nine customers complained, then why is the story spreading like wildfire?

Not sure. How come the iPhone 5, 4s, and 4 bending didn't cause a media craze like this either?!

Figure 1.2: An example of an excerpt from a single conversation, which consists of an initial post followed by a set of comments (accessed in March, 2015).

sualization (InfoVis) community individually attempt to address this and similar problems, only little efforts have been devoted to integrating NLP and InfoVis techniques in a synergistic way. In general, previous work at the intersection of Infovis and NLP has developed relatively simple and generic approaches to visual text analytics.

A common visualization technique to help users in analyzing individual documents is to visually encode the frequency of keywords using font size, such as applying a tag cloud metaphor [113]. Another way to visually represent a document is to split it into blocks of text and then use color to indicate some key information within each block [54, 73]. For instance, TileBars [54] presents the documents retrieved from search queries using bars, where widths are relative to the length of the documents, and heights are relative to the number of query terms. The content of a document were divided into blocks and the color within each block represents the frequency of query terms. Similarly, Oelke and Keim extracted some features such as vocabulary richness or sentence length for each text block and represent them using color at different levels of granularity ranging from chapters to sentences to words [73]. However, since the above visualizations did not reveal semantic relationships among terms, another body of works attempted to capture such relationship using tree representations, such as Word Tree [126], Double Tree [29], and DocuBurst [25]. Additionally, they mapped term frequency to font size.

In the research presented in this thesis, we have used some of the common metaphors from the above works, such as colored bars to encode some features of the text (e.g., sentiment) and mapping frequency of discussion topics to font size. However, the above works are devised for generic documents, in contrast, we have considered additional data specific of conversations including the reply-relationships between comments and the relationships between comments and authors, which introduces more challenges from the visualization perspective that were not addressed in the above works.

Previous research that specifically focused on visualizing asynchronous conversations also have important limitations. Most of these works did not derive their visual encodings and interactive techniques from task and data abstractions based on a detailed analysis of specific user needs and requirements in the target domains. Instead, they either visualize only metadata such as the reply-relationship between comments that do not reveal any content information (e.g., [101, 123, 125]), or visualize the results of simple, often inaccurate text analysis techniques that are not adequate to support the user (e.g., [109, 124]). Furthermore, these text analysis methods are not designed to exploit the specific characteristics of asynchronous conversations, such as reply-relationships and use of quotation; despite recent evidence suggest that NLP methods such as topic modeling [70] are more accurate when these specific characteristics are taken into account.

In short, most previous works did not integrate text analysis and information visualization based on considering the specific characteristics of online conversations and of their users. This dissertation aims to address such short-coming of existing approaches.

1.2 Approach

The primary goal of this thesis is to develop a comprehensive understanding of how a combination of text analysis and interactive visualization can support users in exploring online conversations. The hypothesis is that *by tightly integrating NLP and InfoVis techniques, we can better support the user's task of exploring and analyzing conversations.* But how NLP and InfoVis techniques can be effectively integrated? More specifically, I pose the following research questions:

1. What tasks do users want to perform and what metadata and text analysis results are actually useful to support these tasks?

2. How can useful metadata and content be extracted from the conversation?

3. How should the extracted metadata and contents be visualized to the user?

4. How can we support the user when (she realizes that) the current text analysis results are not helping her anymore?

5. When we compare our proposed approach for exploring and analyzing conversations with traditional interfaces, is there any difference in user performance and subjective measures?

6. What specific aspects of the proposed approach are more/less beneficial for the potential users?

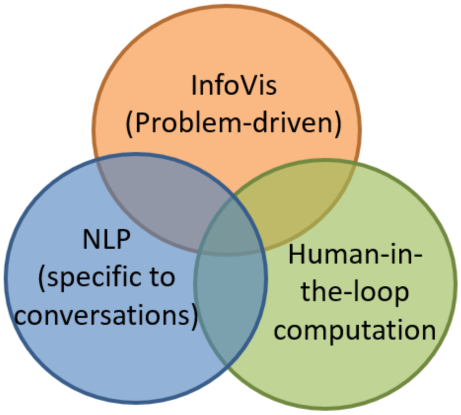


Figure 1.3: The research presented in this thesis falls into the intersection between information visualization, natural language processing, and human-in-the-loop computation.

My research falls into the cross section between three main research areas that are associated with my research questions: information visualization (Q1, Q3, Q5, Q6), natural language processing (Q2, Q4), and human-in-the-loop-computation (Q4). The overlap between these three areas defines the scope of my doctoral research, i.e., designing and testing visual analytics systems for asynchronous conversations (see Figure 1.3). The distinct role played by each area in my research is as follows:

-*Why InfoVis?* To address Q1 and Q3, I focus on applying human-centered design methodologies from the InfoVis literature [91, 111]. Starting from an analysis of user behaviours and needs in the target conversational domain, such methods help uncover useful task and data abstractions. On the one hand, task and data abstractions can characterize the type of information that needs to be extracted from the conversation (Q1); on the other hand, they can inform the design of the visual encodings and interaction techniques (Q3). More tellingly, as both the NLP and the InfoVis components of the resulting system are designed by referring to a common set of task and data abstractions, they are more likely to be consistent and synergistic. Finally, in order to answer Q5 and Q6, I focus on applying different techniques for user evaluation established in the InfoVis literature, such as informal evaluations, controlled studies, case studies and online user studies [78].

- *Why NLP for conversations?* To address Q2, I focus on devising and applying text mining and summarization methods specific to asynchronous conversations. Most of the existing visual text analytics systems use NLP methods that were originally devised for generic documents. These methods generally do not exploit the specific characteristics of asynchronous conversations (e.g., use of quotation, dialog acts), while it has been shown that text analysis results are more accurate when these specific characteristics are taken into account [70]. In order to address this limitation, I aim to adopt and extend text mining and summarization approaches that take advantage of the conversational features.

- *Why human-in-the-loop computation?* To address Q4, I focus on considering human feedback in the text analysis process. The motivation for such an approach is that the results of NLP systems can be either too noisy and/or may not match the user's mental model, and current tasks. In such situations, I aim to support the user in providing feedback to the underlying NLP system, so that the results can better match her information needs.

In essence, my approach to designing visual text analytics systems consists of apply design study methodology in InfoVis to uncover data and task abstractions; apply NLP methods for extracting the identified data to support the corresponding tasks; and incorporate human feedback in the text analysis process when the extracted data is noisy or may not match the user's mental model and current tasks.

1.3 Scope

In the initial, exploratory phase of my research, I focused on understanding and characterizing the broad range of domains, users, and data for asynchronous conversations with the aim of better defining the scope for the thesis. Here, I provide an overview of the different types of conversations and users, followed by the design scope of this thesis.

Types of online conversations: Online conversations can be characterized from at least three major perspectives, as shown in Table 1.1. First of all, the phenomenal adoption of novel Web-based social media has led to the rise of asynchronous conversations in many different *modalities*, ranging from blogs, to mi-

Dimension	**Examples**
Nature of initial post	Article, question, opinion, proposal, review
Genre: The subject of discussions	Politics, business, technology, education, art, lifestyle, entertainment, health, sports
Conversational modality: It refers to "a means or mode of communication, where a particular modality may be associated with both distinct communication technologies as well as distinct social conventions and language characteristics" [17].	Blogs, conversations in social networks, microblogs

Table 1.1: Characterizing online conversations from different dimensions.

croblogs, to discussions in social networks. Social news blog sites[1], such as Reddit, Slashdot, and Digg contain user-generated stories that are ranked based on popularity [2]. Users can comment on these posts and these comments may also be ranked. Online news sites such as New York Times [3] allow readers to contribute by commenting on articles on a broad range of topics. Moreover, for many users, microblogs such as Twitter and Tumblr and social networking sites such as Facebook and Google Plus have become part of their online life.

Second, an online conversation can be characterized based on the *content of its initial post*, which can be an article, question, opinion, or review. Some websites may focus on a specific type of initial post. For instance, Quora, a community question answering forum, allows people to start a conversation by asking a question.

Third, online conversations can be categorized based on their *genre*. Some websites may focus on a broad range of subjects, while others may focus on a particular genre. For example, Huffington Post and Daily Kos blogs are dedicated to the discussion of politics[4], while Slashdot and MacRumors focus on technology.

[1] Also known as social news aggregators

[2] reddit.com, slashdot.org, digg.com

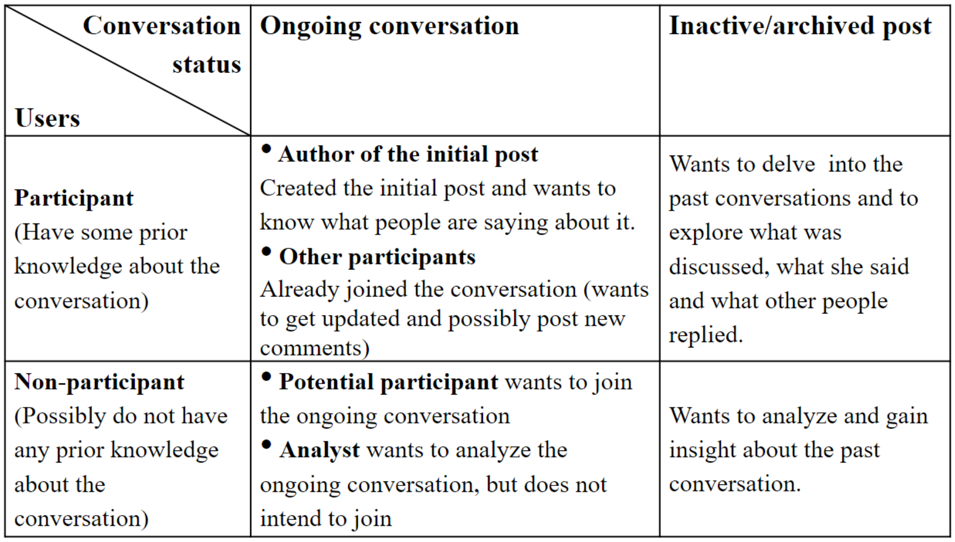
[3] www.nytimes.com

[4] www.huffingtonpost.com, www.dailykos.com

In this research, we developed a set of visual text analytics systems focusing on supporting a common set of tasks involved in exploring and analyzing conversations. However, for the purpose of design and evaluation of the approach, we mainly focused on blog conversations. According to the basic definition, a blog refers to “frequently updated website consisting of dated entries arranged in reverse chronological order so the most recent post appears first” [96]. Over the years, blogs have evolved in terms of style and content, enabling the production of diverse content [15, 79]. Based on the broad definition of blogs and its diverse nature, in this dissertation we refer to blogs as a variety of conversations ranging from personal blogs, to corporate blogs, to discussions on news articles, to online forums.

Blogs are appealing over other conversational modalities as an initial design target of this thesis for various reasons. First, blogging is a common way for people to freely publish their thoughts about almost any content published on Web [17]. Therefore, they are not limited to any specific type of initial contribution or any specific subject listed in Table 1.1. Second, blogs are mainly focused on high-quality content generation and information sharing as opposed to purely social interactions, which are more prevalent in social networking sites. Third, blogs are often archived and actively read over several years [42]. Finally, unlike microblog [108], they do not have fixed length comments; furthermore, they have finer conversational structure as participants often reply to a post and/or quote a fragment of other comments [70], making it a more challenging problem for users to explore and analyze such conversations.

Nevertheless, in the later part of this dissertation, we show how our solutions can be tailored and adapted to specific domain problems. Here, by domain problems we refer to problems faced by a user or a group of users in a specific conversational modality, possibly with a focus on a particular genre. For instance, in Chapter 5 we present a design study, where our visual text analytics systems were simplified and tailored to support information seeking tasks in a community question answering forum (i.e., a blog where the initial post is a question) for a user population possibly having low visualization expertise. Furthermore, in Chapter 6, we report how several other researchers have applied or partially adopted our data abstractions and visual encodings to address specific domains problems, such as



Conversation status **Users**	**Ongoing conversation**	**Inactive/archived post**
Participant (Have some prior knowledge about the conversation)	• **Author of the initial post** Created the initial post and wants to know what people are saying about it. • **Other participants** Already joined the conversation (wants to get updated and possibly post new comments)	Wants to delve into the past conversations and to explore what was discussed, what she said and what other people replied.
Non-participant (Possibly do not have any prior knowledge about the conversation)	• **Potential participant** wants to join the ongoing conversation • **Analyst** wants to analyze the ongoing conversation, but does not intend to join	Wants to analyze and gain insight about the past conversation.

Table 1.2: User categorization for asynchronous conversation.

problems faced by administrators of online health forums and instructors of educational forums.

Users: As shown in Table 2.2, users in a conversational domain can be categorized into two groups based on their activities: (a) *participants* who have already contributed to the conversations, and (b) *non-participants* who have not contributed to the conversations yet. Depending on different user groups the tasks might vary as well, something that needs to be taken into account in the design process.

For example, imagine a *participant* who has expressed her opinion about a major political issue. After some time, she may become interested to know what comments were made supporting or opposing her opinion, and whether those comments require a reply right away. In contrast, a *non-participant,* who is interested in joining the ongoing conversation on that particular political issue, may want to decide whether and how she should contribute by quickly skimming through a long thread of blog comments. Another group of users may include analysts who do not wish to join the conversation, but may want to analyze and gain insights from conversations. For instance, a journalist may want to summarize the major arguments that were used to support or oppose the political issue. Another example is an analyst who wants to discover important insights from conversations and present those to a policy maker for supporting her decision-making process.

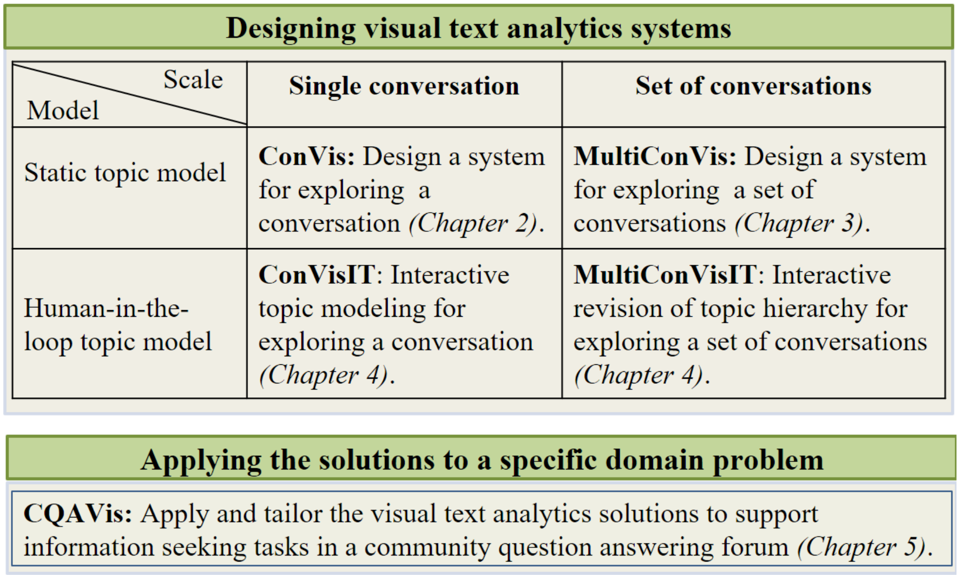
In this dissertation, we have mainly focused on supporting the non-participant's activity on archived conversations as opposed to ongoing ones. However, as we will discuss in Chapter 6, in future work our text analysis methods and visualization techniques could be extended to support other types of users and ongoing conversations.

1.4 Thesis Contributions

The fundamental contributions of this research have arisen from devising the approaches for tightly integrating natural language processing and information visualization techniques for interactive exploration of conversational data. In order to evaluate the effectiveness of these approaches, we explore a two-dimensional design space, as shown in Figure 1.4. These dimensions are: the scale of conversations (a single conversation vs. a set of conversations) and the topic modeling choice (static model vs. human-in-the-loop model). Here, a single conversation consists of an initial post followed by a set of comments, where these comments and the initial posts are connected by reply-relationships, as exemplified in Figure 1.2. A collection of conversations consists of two or more conversations that share some common themes. For instance, these conversations may be retrieved from a blog or forum site given a search query, as shown in Figure 1.1.

Along the two different dimensions of our design space, four different visual text analytics systems have been developed to explore and validate our fundamental approaches. After designing these systems, in Chapter 5 we have shown how our visual text analytics solutions can be applied and tailored to a specific domain problem. The resulting system, CQAVis, was designed for supporting information seeking tasks in a community question answering forum for a user population possibly having low visualization expertise. A summary of the available resources for our systems is provided in Table 1.3.

At different stages of designing the systems, we conducted user studies to validate our approach. Table 1.4 provides a complementary summary of these user studies organized by thesis chapters. As we can see, a variety of user studies were conducted ranging from informal evaluation, to more formal summative studies in



Designing visual text analytics systems

Scale Model	Single conversation	Set of conversations
Static topic model	**ConVis:** Design a system for exploring a conversation *(Chapter 2)*.	**MultiConVis:** Design a system for exploring a set of conversations *(Chapter 3)*.
Human-in-the-loop topic model	**ConVisIT**: Interactive topic modeling for exploring a conversation *(Chapter 4)*.	**MultiConVisIT**: Interactive revision of topic hierarchy for exploring a set of conversations *(Chapter 4)*.

Applying the solutions to a specific domain problem

CQAVis: Apply and tailor the visual text analytics solutions to support information seeking tasks in a community question answering forum *(Chapter 5)*.

Figure 1.4: The design space explored in this research.

lab settings, to a web-based study in the wild.

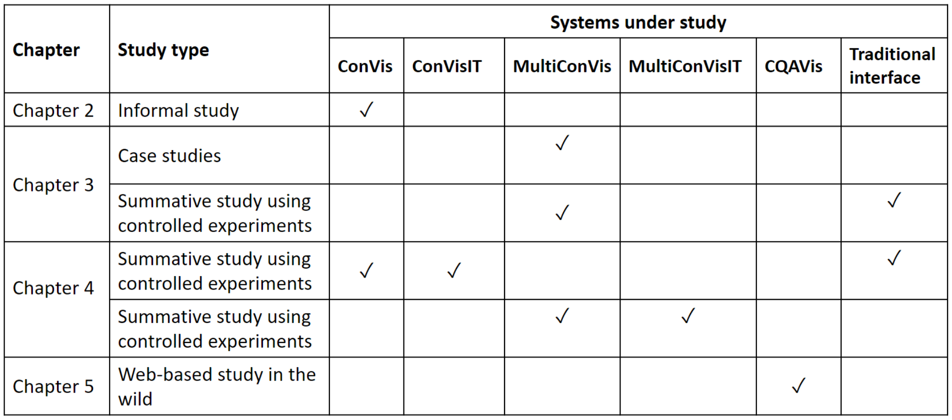
We now provide an overview of the visual text analytics systems we developed followed by the summary of contributions that have emerged from designing and evaluating these systems.

1.4.1 Exploring Conversations with Static Model

In the initial work, we proposed a visual text analytics system that supports users in exploring a single asynchronous conversation (Chapter 2). Following the design study methodology in InfoVis, we started with a user requirement analysis for the domain of blog conversations to derive a set of design principles. Based on these principles, we designed an overview+detail interface, named ConVis that provides a visual overview of a conversation by presenting topics, authors and the thread structure of a conversation (see Figure 1.5). The underlying topic modeling approach was specifically designed for asynchronous conversations that takes into account the unique features of conversations, namely reply relationships and use of quotation. By using this approach, we group the sentences of a conversation into a number of topical clusters using a graph-based clustering technique and label each

System	Resource	url
ConVis	Video demo Live demo	Short: https://goo.gl/gBQE3e Long: www.cs.ubc.ca/group/iui/convis.mp4 Long:www.cs.ubc.ca/ enamul/convis/
MultiConVis	Video demo	Short: https://goo.gl/ZmVYks Long: www.cs.ubc.ca/group/iui/multiconvis.mp4
ConVisIT	Video demo	Short: https://goo.gl/QALDvw Long: www.cs.ubc.ca/group/iui/convisit.mp4
MultiConVisIT	Video demo	Short: https://goo.gl/edT69x Long: www.cs.ubc.ca/group/iui/multiconvisit.mp4
CQAVis	Video demo Live demo	Short: https://goo.gl/IM3Gez Long: www.cs.ubc.ca/group/iui/cqavis.mp4 http://iyas.qcri.org

Table 1.3: A summary of the available resources for our systems.

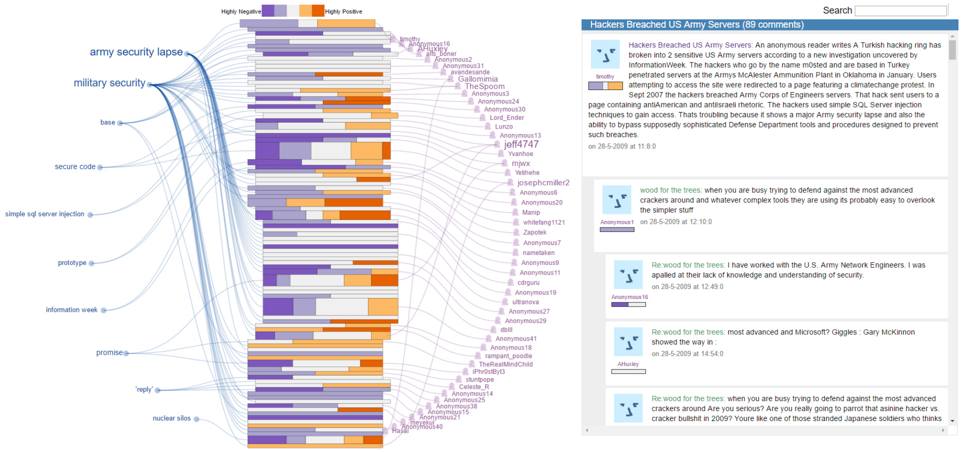


Chapter	Study type	Systems under study					
		ConVis	ConVisIT	MultiConVis	MultiConVisIT	CQAVis	Traditional interface
Chapter 2	Informal study	✓					
Chapter 3	Case studies			✓			
	Summative study using controlled experiments			✓			✓
Chapter 4	Summative study using controlled experiments	✓	✓				✓
	Summative study using controlled experiments			✓	✓		
Chapter 5	Web-based study in the wild					✓	

Table 1.4: Summary of user studies conducted in the dissertation.

cluster by generating semantically meaningful descriptors. The visual interface provides various interaction techniques such as brushing and highlighting based on multiple facets to support the user in exploring and navigating the conversation.

We performed an informal user evaluation, which provides anecdotal evidence about the effectiveness of ConVis as well as directions for further design. The participants' feedback from the evaluation suggests that ConVis can help the user to identify the topics and opinions expressed in the conversation; supporting the user



Hackers Breached US Army Servers (89 comments)

timothy

Hackers Breached US Army Servers: An anonymous reader writes A Turkish hacking ring has broken into 2 sensitive US Army servers according to a new investigation uncovered by InformationWeek. The hackers who go by the name m0sted and are based in Turkey penetrated servers at the Armys McAlester Ammunition Plant in Oklahoma in January. Users attempting to access the site were redirected to a page featuring a climatechange protest. In Sept 2007 the hackers breached Army Corps of Engineers servers. That hack sent users to a page containing antiAmerican and antiIsraeli rhetoric. The hackers used simple SQL Server injection techniques to gain access. Thats troubling because it shows a major Army security lapse and also the ability to bypass supposedly sophisticated Defense Department tools and procedures designed to prevent such breaches.

on 28-5-2009 at 11:8:0

Anonymous1

wood for the trees: when you are busy trying to defend against the most advanced crackers around and whatever complex tools they are using its probably easy to overlook the simpler stuff

on 28-5-2009 at 12:10:0

Anonymous16

Re:wood for the trees: I have worked with the U.S. Army Network Engineers. I was apalled at their lack of knowledge and understanding of security.

on 28-5-2009 at 12:49:0

AHuxley

Re:wood for the trees: most advanced and Microsoft? Giggles : Gary McKinnon showed the way in :

on 28-5-2009 at 14:54:0

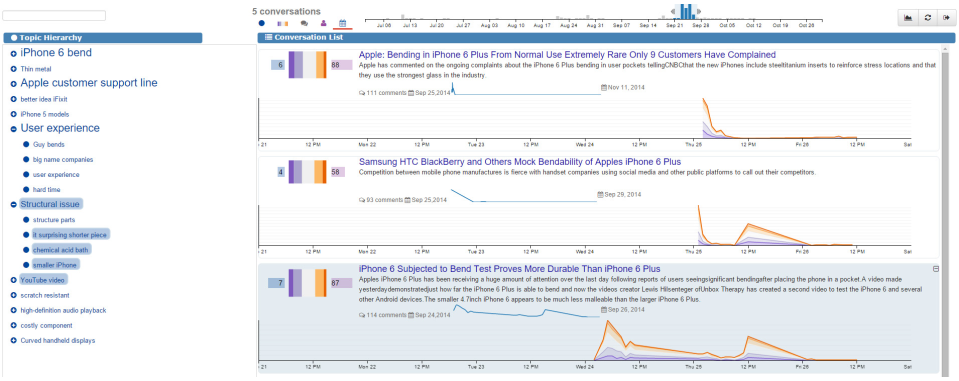
Re:wood for the trees: when you are busy trying to defend against the most advanced crackers around Are you serious? Are you really going to parrot that asinine hacker vs. cracker bullshit in 2009? Youre like one of those stranded Japanese soldiers who thinks

Figure 1.5: The ConVis interface (see Figure 2.2 for further description).

in finding comments of interest, even if they are buried near the end of the thread. The informal evaluation also reveals that in few cases the extracted topics and opinions are incorrect and/or they may not match the mental model and information needs of the user. To address this problem, we introduced a human-in-the-loop model as discussed in Section 1.4.2.

In subsequent work, we focused on supporting readers in exploring a collection of conversations related to a given query (Chapter 3). Exploring topics of interest that are potentially discussed over multiple conversations is a challenging problem, as the volume and complexity of the data increases. To address this challenge, we devised a novel hierarchical topic modeling technique that organizes the topics within a set of conversations into multiple levels, based on their semantic similarity. For this purpose, we extended the topic modeling approach for a single conversation to generate a topic hierarchy from multiple conversations by considering the specific features of conversations. We then designed a visual interface, named MultiConVis that presents the topic hierarchy along with other conversational data, as shown Figure 1.6. The user can explore the data, starting from a possibly large set of conversations, then narrowing it down to the subset of conversations, and eventually drilling-down to the set of comments relating to one conversation.

We evaluated MultiConVis through case studies with domain experts and a formal user study with regular blog readers. Our case studies demonstrate that the



Apple: Bending in iPhone 6 Plus From Normal Use Extremely Rare Only 9 Customers Have Complained

Apple has commented on the ongoing complaints about the iPhone 6 Plus bending in user pockets tellingCNBCthat the new iPhones include steeltitanium inserts to reinforce stress locations and that they use the strongest glass in the industry.

111 comments Sep 25,2014 Nov 11, 2014

Samsung HTC BlackBerry and Others Mock Bendability of Apples iPhone 6 Plus

Competition between mobile phone manufactures is fierce with handset companies using social media and other public platforms to call out their competitors.

93 comments Sep 25,2014 Sep 29, 2014

iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus

Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.

114 comments Sep 24,2014 Sep 26, 2014

Figure 1.6: The MultiConVis interface (see Figure 3.1 for further description).

system can be useful in a variety of contexts of use, while the formal user study provides evidence that the MultiConVis interface supports the user's tasks more effectively than traditional interfaces. In particular, all our participants, both in the case studies and in the user study, appear to benefit from the topic hierarchy and the high-level overview of the conversations. The user study also shows that the MultiConVis interface is significantly more useful than the traditional interface, enabling the user to find insightful comments from thousands of comments, even when they were scattered around multiple conversations, often buried down near the end of the threads. More importantly, MultiConVis was preferred by the majority of the participants over the traditional interface, suggesting the potential value of our approach for combining NLP and InfoVis.

Contributions:

1) We performed a user requirements analysis based on extensive literature review in the domain of blogs to inform our interface design for both a single conversation as well as a set of conversations. The analysis reveals the data and task abstractions for the problem domain and a set of design principles to support the user requirements.

2) We adopted a topic modeling method for effectively extracting topics from a single conversation. We also extended this method for creating a topic hierarchy for a whole collection of conversations, by aggregating the topics extracted from each conversation in the collection. The novelty of our approach is that in

both extracting the topics and organizing them into a hierarchy, our methods takes advantage of conversational features to enhance the quality of the topic model.

3) We presented the design and implementation of two novel visual interfaces: ConVis and MultiConVis. Unlike previous approaches which either visualize some metadata or only one type of content information from the conversations (e.g., the topics covered but not opinions), our interfaces visualize both *topic* and *opinion* mining results along with a set of metadata, such as authors and position of the comments. We also proposed a way to seamlessly integrate the two interfaces to allow users to switch from exploring a collection of conversations to a single conversation.

4) We designed and conducted a series of user studies, namely an informal evaluation, a formal lab-based study, and three case studies, which revealed the differences in user performance and subjective opinions when our systems were compared to traditional blog interfaces for exploring conversations. These studies also provide further directions for our subsequent research, including the need for a human-in-the-loop model.

1.4.2 Exploring Conversations with Human-in-the-loop Model

A preliminary evaluation of ConVis suggested that while the participants were generally positive about the interface, the results of the topic model were sometimes noisy and/or did not match their current information needs. This was particularly evident from the interviews, where users expressed a pressing need for enhancing their ability to revise the topic model according to their own information needs. This was also revealed by spontaneous users' comments, while they were performing the experimental tasks.

Motivated by this experience, we proposed a novel interactive topic modeling approach in Chapter 4 that revises the topics on the fly on the basis of users' feedback. We then designed a visual interface, named ConVisIT, by extending ConVis, where the user can explore long conversations, as well as revise the topic model when the current results are not adequate to fulfill her information needs (see Figure 1.7). By analyzing the tasks of exploring online conversations, we devised a set of topic revision operations that are critical to the user. For instance, the user

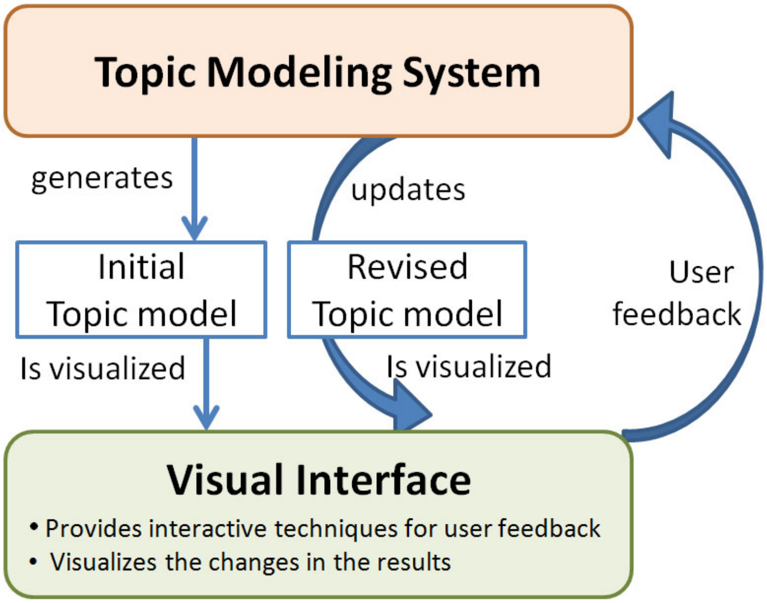


Figure 1.7: Our interactive topic modeling framework

could perform a *merge* operation on two topics if these topics are talking about similar issues. In other cases, if a topic is too generic the user could *split* this into further smaller sub-topics. For example, splitting 'ObamaCare' would create three subtopics namely 'health insurance', 'drugs' and 'healthcare cost'. By dynamically revising the topic model, the user could build a topic model that better matches her mental model and current information needs.

A similar human-in-the-loop model was investigated for exploring a set of conversations in Chapter 4. The motivation is that while a topic hierarchy is useful to organize the discussion themes within a set of conversations into a coherent structure, such a hierarchy may be too noisy and/or may not match the user's current tasks or her mental model. To support the user in this situation, we devised an approach for revising the topic hierarchy based on users' information needs. According to this approach, the user can provide the feedback to the system through the visual interface MultiConVisIT, which incorporates a set of features for revising the topic hierarchy. The system then updates the topic hierarchy, which is visualized in the interface for further exploration.

We ran two summative user studies in lab-based settings to compare ConVisIT and MultiConVisIT with interfaces that do not support human-in-the-loop topic modeling. In essence, both studies suggest that most users benefit from getting more control over the topic modeling process while exploring conversations. The first study, as described in Section 4.6.1, reveals that ConVisIT outperformed both a traditional interface as well as ConVis along several subjective metrics. Similar results were found in the second study reported in Section 4.6.2, where MultiConVisIT was found to be more useful and was also preferred over its counterpart that does not provide interactive topic revision operations.

Contributions:

1) We proposed a novel interactive topic modeling approach specifically devised for asynchronous conversations. Existing systems for interactive topic modeling (e.g., [19, 66, 81]) were mainly devised for generic documents without considering the unique features of conversations.

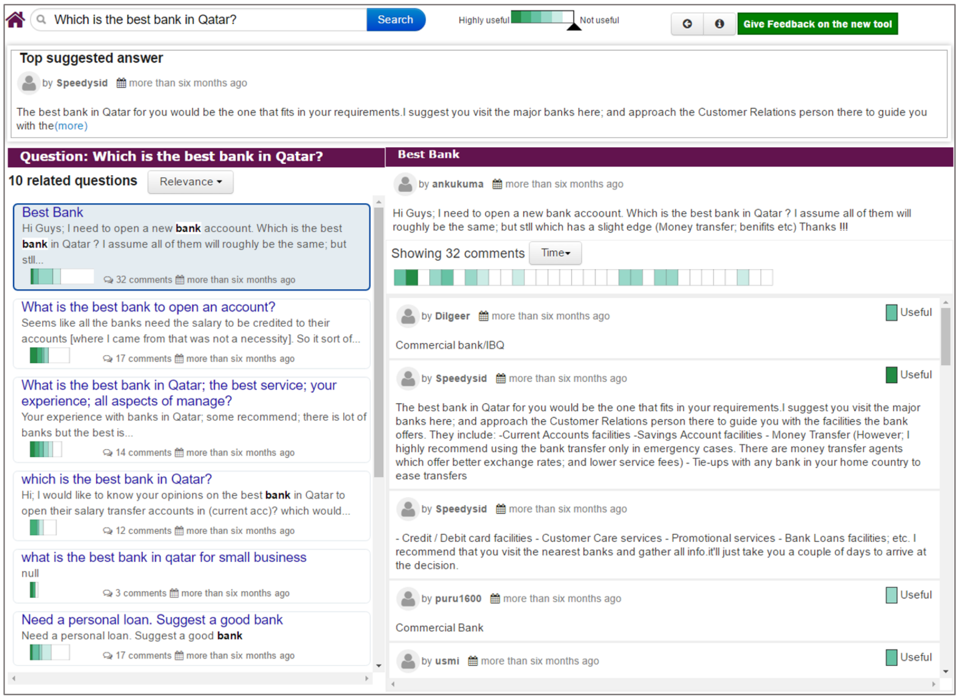
2) We designed a set of interactive features that allow the user to revise the current topic model. In response, the interface updates and re-organizes the modified topics by means of intuitive animations, so that the user can better fulfill her information needs.

3) We conducted two lab-based summative studies, which revealed the potential utility of our human-in-the-loop topic modeling approaches.

1.4.3 Applying and Tailoring the Solutions to Specific Domains

After designing the visual text analytics systems, we have analyzed how our solutions for generic blog conversations can be applied to specific domain problems in Chapter 5 and Chapter 6. To answer this question, we conducted a design study in the domain of community question answering (CQA) forum, in which the initial post is a question. Here, our generic visual text analytics solutions were applied and tailored to support information seeking tasks for a user population possibly having low visualization expertise. Figure 1.8 shows a screenshot of our interface, presenting the results for a user-provided query.

Our system was evaluated by deploying it in an online study, in which it was tested with hundreds of real users. This large-scale Web study underlines the potential for tightly integrating NLP and InfoVis in practice, offering the users a new


Which is the best bank in Qatar?
Search
Highly useful
Not useful
Give Feedback on the new tool
Relevance
Time
Useful

Top suggested answer

by **Speedysid** more than six months ago

The best bank in Qatar for you would be the one that fits in your requirements.I suggest you visit the major banks here; and approach the Customer Relations person there to guide you with the(more)

Question: Which is the best bank in Qatar?

10 related questions

Best Bank
Hi Guys; I need to open a new **bank** accoount. Which is the best **bank** in Qatar ? I assume all of them will roughly be the same; but stll...
32 comments more than six months ago

What is the best bank to open an account?
Seems like all the banks need the salary to be credited to their accounts [where I came from that was not a necessity]. So it sort of...
17 comments more than six months ago

What is the best bank in Qatar; the best service; your experience; all aspects of manage?
Your experience with banks in Qatar; some recommend; there is lot of banks but the best is...
14 comments more than six months ago

which is the best bank in Qatar?
Hi; I would like to know your opinions on the best **bank** in Qatar to open their salary transfer accounts in (current acc)? which would...
12 comments more than six months ago

what is the best bank in qatar for small business
null
3 comments more than six months ago

Need a personal loan. Suggest a good bank
Need a personal loan. Suggest a good **bank**
17 comments more than six months ago

Best Bank

by **ankukuma** more than six months ago

Hi Guys; I need to open a new bank accoount. Which is the best bank in Qatar ? I assume all of them will roughly be the same; but stll which has a slight edge (Money transfer; benifits etc) Thanks !!!

Showing 32 comments

by **Dilgeer** more than six months ago — Useful

Commercial bank/IBQ

by **Speedysid** more than six months ago — Useful

The best bank in Qatar for you would be the one that fits in your requirements.I suggest you visit the major banks here; and approach the Customer Relations person there to guide you with the facilities the bank offers. They include: -Current Accounts facilities -Savings Account facilities - Money Transfer (However; I highly recommend using the bank transfer only in emergency cases. There are money transfer agents which offer better exchange rates; and lower service fees) - Tie-ups with any bank in your home country to ease transfers

by **Speedysid** more than six months ago

- Credit / Debit card facilities - Customer Care services - Promotional services - Bank Loans facilities; etc. I recommend that you visit the nearest banks and gather all info.it'll just take you a couple of days to arrive at the decision.

by **puru1600** more than six months ago — Useful

Commercial Bank

by **usmi** more than six months ago — Useful

Figure 1.8: CQAVis is a visual interface to support information seeking tasks in a community question answering forum (See Figure 5.3 for further description).

way of seeking information in CQA forums. It also reveals important lessons for designing and studying such systems for real users with varying levels of expertise, which can arguably be generalizable for the design and evaluation of visual analytics systems for other conversational domains.

In addition to our own work, we also conducted a survey focusing on how other researchers have recently applied or partially adopted the data abstractions and visual encodings of MultiConVis and ConVis in a variety of domains, such as to online health forums and educational forums. In Section 6.2, we analyze these research works to understand the potential applicability of our systems to different domain problems.

Contributions:

1) We characterized the CQA forums by identifying user tasks and some key

design needs.

2) We designed and implemented a visualization tool that demonstrates how our generic solutions for integrating NLP and InfoVis techniques presented in Chapter 2 and 3 can be applied and tailored to the information seeking tasks in CQA.

3) We evaluated the new CQA forum tool in the wild in an ecologically valid testing by deploying the system among real forum readers.

4) We identified and summarized generalizable lessons that can be useful to design visual interfaces for online conversations in other domains, as well as to design for user population possibly having low visualization literacy.

Chapter 2

Supporting Users in Exploring a Single Conversation

In this chapter, we present a visual text analytic system that tightly integrates interactive visualization with novel text mining and summarization techniques to fulfill information needs of users in exploring a single conversation. At first, we perform a user requirement analysis for the domain of blog conversations to derive a set of design principles. Following these principles, we present an interface that visualizes a combination of various metadata and textual analysis results, supporting the user to interactively explore the blog conversations. Finally, we conducted an informal user evaluation, which provides anecdotal evidence about the effectiveness of our system and directions for further design[1]. A further evaluation of our system, which was conducted in the form of a summative user study in a controlled setting, is described in Chapter 4.

2.1 Introduction

A single asynchronous conversation such as a blog conversation consists of an initial post such as an article or a question, followed by a set of subsequent replies. Often many people contribute to the discussion, which can quickly become very

[1] This chapter is a slightly modified version of our paper *ConVis: A visual text analytic system for exploring blog conversations*, by Enamul Hoque and Giuseppe Carenini; in Journal of Computer Graphics Forum (Proceedings of EuroVis), 33(3):221230, 2014 [59].

long with hundreds of comments. Traditional social media sites present the original posts and subsequent replies as a paginated indented list (see Figure 1.2). Thus the reader needs to go through a long list of comments sequentially, until her information needs are fulfilled. Going through such an overwhelming amount of textual data in this way often leads to information overload, i.e., the user finds it very difficult to get insights about the ongoing (or past) discussion. The end result is that the readers start to skip comments, generate simpler responses and leave the conversation without satisfying their intent [69].

To illustrate the problem, consider a scenario where Sarah is interested in technology-related blogs. She opens a blog discussion about a news article of hacking in US army servers. She is curious to know what are the different opinions about the US cyber security lapses. She finds that the top few posts blame the ‘shoddy work’ done by the contractor companies, while others believe that the incident was merely ‘a honeypot for hacker’. Sarah wants to know more about what other people are saying about the hacking issue, but soon realizes that the topic of discussion is shifted to ‘US involvement in the Vietnam war’, which she is not interested in. So Sarah keeps on skimming comments and notices that some others are discussing the technical details of hacking. At this point, Sarah is quite exhausted; she does not know whether the long list of remaining comments discuss the reasons for cyber security lapses; but she decides to end the reading without fulfilling her information needs.

To support readers in dealing with similar situations, we have developed ConVis: a visual exploratory text analytic system for blogs that tightly integrates interactive visualization with text mining techniques that are especially devised to deal with conversational data. Motivated by the nested design model [91], we started by characterizing the domain. While asynchronous conversations comprise emails, blogs, microblogs (e.g., Twitter), and messaging; in this chapter we focus on the domain of blogs. In fact, blog conversations often have finer conversational structure as participants often reply to a post and/or quote a fragment of other comments [70], making it a more challenging problem for users to explore and analyze such conversations. Once we have characterized our domain, we derive a set of design principles, which then guide the visual encoding and interaction techniques of ConVis. The primary contributions of this work are as follows:

1) We performed a user requirements analysis based on extensive literature review in the domain of blogs, as described in Section 2.2. The analysis includes data and task abstractions for the problem domain and a set of design principles to support the user requirements.

2) To the best of our knowledge, ConVis is the first visual text analytic system for blog conversations that visualizes both *topic* and *opinion* mining results along with a set of metadata such as authors and position of the comments, which were identified as primary means for browsing and navigation from the user requirement analysis. Existing systems either visualize some metadata or only one type of content information from the conversations (e.g., the topics covered but not opinions), thus limiting the ability of the user to explore and analyze the conversation.

3) We present the design, implementation, and evaluation of ConVis. ConVis visually represents the overview of a blog and then allows the user to explore this conversation based on multiple facets (e.g., topics and authors). This is a major shift from traditional blog reading interfaces which provide a long list of paginated comments, thus only supporting linear navigation.

2.2 From User Requirements to Design Principles

Blog reading has been extensively studied in the fields of computer mediated communications (CMC) [35, 72, 135], social media [48, 57], human computer interaction (HCI) [13, 30, 90], and information retrieval [69, 74, 83, 87, 119]. This literature provides a detailed analysis of the motivations and goals for reading blogs, along with the unique behaviours of blog reading. Based on this analysis, we characterize the data and tasks in the domain of blogs and then identify the user requirements (*UR*), which are finally translated into a set of design principles.

2.2.1 Why and How People Read Blogs?

Over the years, several studies have been conducted to identify the motivations and goals for reading blog conversations [9, 30, 72, 74, 87]. Kaye performed a web survey among active bloggers to find the reasons why they access blogs [72]. These reasons were grouped into 10 general motivational blocks, including information seeking, fact checking, guidance/opinion seeking, and political surveillance. In particular, the users reported that often they read blogs to seek information about their area of interests such as education, technology, and politics [30, 72]. Blogs also help users to quickly verify and compare accounts of news and information and check the accuracy of traditional media (fact checking) [72, 74]. Frequently, users read blogs to seek a wide variety of opinions and to help them make up their minds about important issues [30, 72, 74, 87]. Mishne noted that the information in blogs is often subjective or opinionated [87]. In fact, it has been found that readers consider blogs with a mixture of positive and negative posts more credible [9]. Overall, this suggests that the interface should facilitate a visual overview of the diverse range of opinions covering positive and negative sentiments about important topics, allowing the user to understand various viewpoints (*UR-1*).

The people-centric nature of the domain of blogs was reported in various studies [30, 87]. Dave et al. reported that Blog readers are looking to find ideas or information, take the pulse of a community and meet people [30]. In other words, blogging can promote a sense of belonging in the blogosphere among others who try to publicly express their opinions and to affiliate with like-minded individuals ("find people who think like I do") [72, 74]. This indicates the importance of identifying the key participants and their opinions (*UR-2*).

In reality, users do not always look for important information or opinions, they may read blogs simply for enjoyment or personal fulfillment [13, 72]. An ethnographic study reveals that "the participants visit blogs for information, inspiration, entertainment, and to a certain extent because it is just what they have always done" [13]. Kaye suggests that blogs bring more novelty and thus users find blogs to be more fun and interesting than formal media content [72]. This aspiration for novelty and fun should be encouraged by the interface by promoting exploration and serendipitous discoveries (*UR-3*).

Previous studies suggest that many blog readers are inherently *variety-seekers* [90, 119], i.e., they are often looking for a variety of opinions and discussion themes. Singh et al. found the individual's tendency to switch from one set of topics to another [119]. Even in the case when a reader may read only content on the same topic, she essentially reads distinct posts leading to some variety within a topic. Thus, being able to browse the conversations based on different possible topics and sub-topics can effectively support this variety seeking behaviour

(*UR-4*). Many users also exhibit *skimming* tendency [95, 135], i.e., they seek to quickly scan through a set of posts to understand what the authors are saying. This behaviour might be explained by the exploratory nature of blog reading. It has been found that readers remain in an *exploratory state* (intermediate state) before entering into a *focused state* from another *focused state* [119]. The reading in this exploratory state provides clues of what the reader may expect to find if she focused on the comments she is currently skimming. In other words, the reader needs to quickly skim through (i.e., explore) a few posts about a topic before delving deeper into its details (i.e., entering into a focused state). Therefore, the interface should facilitate open-ended exploration within the conversation space, by providing navigational cues that help the user to seek interesting comments and to quickly decide whether they are worthwhile to read (*UR-5*).

2.2.2 Data and Tasks Abstraction

From the analysis of primary goals of blog reading, we compile a list of tasks and the associated data variables that one would wish to visualize for these tasks. In addition, we analyze the Blog track in the Text REtrieval Conference (TREC), which defines a set of tasks on opinion finding (e.g., *What do people think about X?*) and blog distillation (e.g., *Find me a blog with a principle interest in X*) [83]. Based on these analyses, we create a set of tasks (phrased as questions) that the blog reader might ask, along with the possible associated variables as listed in Table 2.1. Most of these questions involve *topics* and the *sentiment* expressed in the conversation, which are relevant to some of the key goals of the users, including information seeking, fact checking, and guidance seeking. Q1 and Q2 are related to finding topics, while Q3 through Q6 can involve both topic and sentiment information. Q7 through Q9 may additionally require to know the people-centric information and relate such information with other data such as topics and sentiment (extending *UR-2*). The last question (Q10) reflects the motivations for personal fulfillment/enjoyment. Finally, to reflect the exploratory behaviour associated with most of the tasks listed here, both thread (to support *exploratory state*) and comments (to support *focused state*) are included as data variables.

Upon identifying the data involved in the list of tasks, we abstract them in

No	Questions (Q)	Topic	Author	Opinion	Thread	Comment
1	What this conversation is about?	X				X
2	Which topics are generating more discussions?	X				
3	What do people say about topic X?	X		X	X	X
4	How controversial was the conversation? Were there substantial differences in opinion?	X	X	X	X	X
5	How do other people's viewpoints differ from my current viewpoint on topic X?	X		X	X	X
6	Why are people supporting/opposing an opinion?			X	X	X
7	Who was the most dominant participant in the conversation?		X		X	X
8	Who are the sources of most negative/positive comments on a topic?	X	X	X	X	X
9	Who has similar opinions to mine?		X	X		X
10	What are some interesting/funny comments to read?	X	X	X		X

Table 2.1: A set of tasks (phrased as questions) that a user may likely have to perform/answer while exploring a blog conversation to satisfy her information needs.

terms of scale and type. Table 2.2 lists a comprehensive set of conversational data to be visualized and their abstract types. We also compute average and maximum counts for different types of data to better understand what scale the visualization needs to deal with. These values are computed based on a set of 20 Slashdot blogs which comes with human generated topic annotations [70]. Here, the *topics* and the *sentiment* are added since they can be useful for performing almost all of the tasks in Table 2.1. The position of the comment in the discussion space and comment length are added since they have been found to be useful cues for

Attributes from data	**Abstract type**	**Counts**	
		Avg.	**Max.**
Thread structure	Tree	Depth: 4.3 nodes: 60.3	Depth: 5 nodes: 101
Topic	Categorical	10.77	23
Author	Categorical	57.71	92

Derived Attributes	**Abstract type**	**Range**
Topic length	Quantitative	[0.0,1.0] (normalized)
Comment length	Quantitative	[0.0,1.0] (normalized)
Position of the comment	Ordinal	[1,101]
Sentiment	Ordinal	[-2,-1,0,+1,+2]

Table 2.2: Set of conversational data to be visualized and their abstract types. The avg. and max. counts for different types of data are provided based on the Slashdot dataset.

navigation [13, 95] (*UR-6*).

Another study suggests that the exact timestamp of a comment is much less important to users than its chronological position with respect to the other comments [13] (*UR-7*). Therefore, we wanted to encode the *position of the comments* (ordinal) as opposed to their timestamps (quantitative).

2.2.3 Design Principles

Based on the user and tasks analysis, we have identified the following key design principles (*DP*) that form the basis of our visualization system. Each design principle is derived from one or more of the User Requirements, as follows:

1. **Show comprehensive set of relevant data:** The visual interface should display a comprehensive set of user/system generated metadata namely comment length, position of the comment, and moderation score *(UR-6, UR-7)*, as well as the results of text analysis *(UR-1)* as listed in Table 2.2.

2. **Provide faceted exploration:** Considering the exploratory nature of blog

reading, the interface should provide various facets (e.g., topics and authors) as a means for navigation and browsing. Once these primary facets are effectively presented, users will arguably take a more active role in exploring conversations in a non-linear fashion, by quickly navigating through comments of a particular facet (addressing *UR-3, UR-4, UR-5*).

3. **See relationship between multiple facets:** Many of the common tasks for browsing conversation require the user to perceive the relations between multiple facets and comments. For example, to perform the task in Q8, the user needs to know how the author, opinion, and topic facets are related to each other. Thus, we aim to effectively reveal the relation between multi-facets to the user, to better support the critical tasks identified in Table 2.1 (*UR-2*).

4. **Provide overview at multiple granularity levels:** We aim to integrate the high level summarized view of the conversation (e.g., topics), the visual overview of the thread (showing sentiment information of all the comments), and the actual comments (detailed content) in a seamless way, so that the user can easily switch between the different levels of overview and the actual conversation (*UR-1, UR-5*).

5. **Lightweight interactions:** To enhance learnability, the interface should facilitate the open-ended exploration of conversations through a set of low-cost interactions [77], that can be easily triggered and reversed without requiring much cognitive overload (*UR-5*). Low cost interactions, along with interface metaphors that are easily understood, can make the exploration process more enjoyable (*UR-3*).

2.3 Related Work

Previous work on visualizing asynchronous conversations can be classified into two categories: metadata-based and content-based visualization; depending on whether the focus of the research was more on visualizing the system and user generated metadata (e.g., thread structure), vs. the results of some text analysis (e.g., finding topical clusters).

2.3.1 Metadata-based Visualization

Earlier works for visualizing asynchronous conversations primarily focused on revealing the structural and temporal patterns of a conversation [34, 101, 125]. Typically, the goal was to effectively represent the thread structure of a conversation using tree visualization techniques, such as thumbnail metaphor (a sequence of rectangles) [125] and radial tree layout [101]. Various interaction techniques, such as highlighting user-specified search terms [125] and zooming into an area of the thread overview [101] were proposed to deal with space constraints for larger threads. Other works visualize various system and user generated metadata such as timestamp [34]; comment length and moderation score [95]. Metadata-based visualization has also been applied to blog archives [68], as opposed to a single blog conversation, which shows the history of social interactions to help users identify potentially useful blog entries.

Even though metadata-based visualizations help to understand the social interaction patterns or the quality of the comments in a conversation, they may be inadequate to support users in most of the tasks shown in Table 2.1. For example, if the user is reading a political blog to know *"what do people think about Obama's recent healthcare policy?"*, knowing how nested the thread structure is or how many replies are made to a particular post would be insufficient. Also, the type of metadata can vary among different forums or blog sites, hence it is hard to generalize the utility of some metadata in supporting the browsing and exploration processes. Therefore, in this work, we are interested in complementing useful metadata by analyzing the textual content and conveying the results to the user. The aim is to provide insights that are based on a more comprehensive view of the conversation.

2.3.2 Content-based Visualization

Some early works aimed to identify and visualize the primary themes or topical clusters within conversations [30, 109]. In contrast, [131] focused more on the organization of the discussion by creating a tree layout, where the parent comment is placed on top as a text block, while the space below the parent node is divided between supporting and opposing statements. In general, the main limitation of these approaches is that they rely on simple, generic text analysis methods, which do not consider the structure of the conversation. More recently, the TIARA system proposes an enhanced Latent Dirichlet Allocation (LDA)-based topic modeling technique, which automatically derives a set of topics to summarize a collection of documents and their content evolution over time [127]. Each layer in the graphical representation represents a topic, where the keywords of each topic are distributed along time. From the height of each topic and its content distributed over time, the user can see the topic evolution. In contrast to visualizing topics, Opinion Space visualizes the differences in opinions in an online conversation [44] by projecting users on a two-dimensional map based on Principal Component Analysis(PCA), where the participants with similar opinions are positioned near to each other. The expectation is that by exploring the map, users can better understand a broad range of viewpoints.

While there has been a clear trend of moving beyond using only metadata to an increasing use of text analysis within the interactive visualization process, current systems generally suffer from two fundamental limitations. First, they use generic text analysis techniques. Secondly, current systems only convey one type of mined information (e.g., either topic or opinion), thus limiting the user's ability to perform most of the tasks in Table 2.1. In this work, we aim to address both limitations.

2.3.3 Faceted Exploration

Faceted browsing has been widely used in general text and multimedia search [55]. According to this approach, various metadata and content information can be used as facets for exploring and filtering content. Various techniques have been developed to interactively explore the faceted datasets [16, 38, 80, 132]. SolarMap arranges entities of the topic facet as cluster nodes and interactively highlights the relations with other facets located in the surrounding circular ring to this cluster region [16]. FacetLens introduces linear facets (e.g., year) and integrates richer faceted navigation techniques to expose trends and relationships between attribute values within a facet [80]. PivotSlice allows the user to construct a series of dynamic queries using facet values to divide the entire dataset into different subsets in a tabular layout, while directed edges are drawn between related items upon selection [132].

In general, the above methods require the user to apply some interactive techniques (e.g., dynamic queries [132], context switching [16]) in order to explore the relationships between facets. In contrast, our work is more similar to [38], where all relationships between facets are permanently displayed and are directly accessible to the user.

2.4 Mining and Summarizing Conversations

We now discuss the two computational approaches that were applied for mining and summarizing conversations: topic modeling and sentiment analysis.

2.4.1 Topic Modeling

In topic modeling, the sentences of a blog conversation are first grouped into a set of topical clusters/segments (segmentation). Then, representative key phrases are assigned to each of these segments (labeling). We adopt a novel topic modeling approach that captures finer level conversation structure in the form of a graph called Fragment Quotation Graph (FQG) [70]. All the distinct fragments (both new and quoted) within a conversation are extracted as the nodes of the FQG. Then the edges are created to represent the replying relationship between fragments. If a comment does not contain any quotation, then its fragments are linked to the fragments of the comment to which it replies, capturing the original 'reply-to' relation. Here, we briefly describe how topic segmentation and labeling can take advantage of the FQG, interested readers are directed to [70] for a more detailed description.

Topic Segmentation:

First, a Lexical Cohesion-based Segmenter (LCSeg) [49] is applied to find the segmentation boundary within each path (from roots to the leaves) of a FQG (see Figure 3.4). Then an undirected weighted graph $G(V,E)$ is constructed, where each node in V represents a sentence within the conversation, and each edge $w(x,y)$ in E represents the number of segments on different paths in which the two sentences appear together. If x and y do not appear together in any segment, their cosine similarity (always between 0 and 1) is used as edge weight. By construction, any

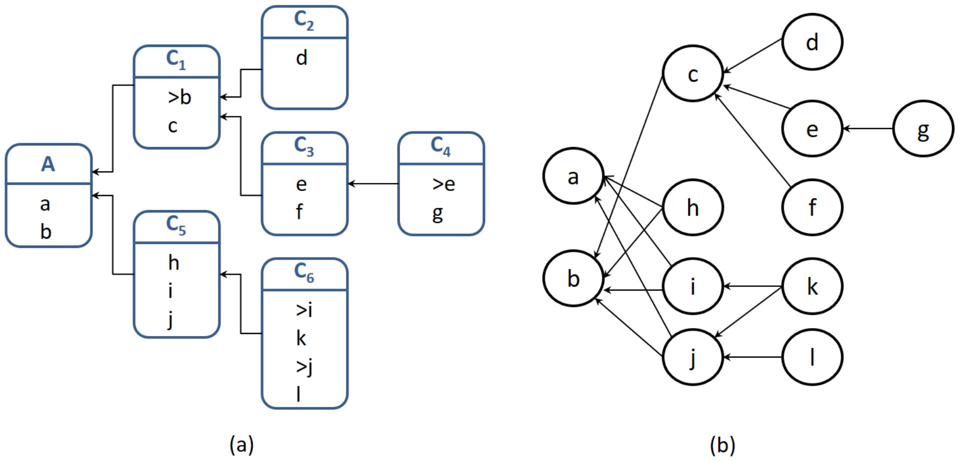


Figure 2.1: a) Reply-to relationships between the initial post A and the comments $C_1, C_2, ..., C_6$ (left). Here, '>' represents the quotation mark and each lowecase letter corresponds to a text fragment that may comprise one or more sentences. b) the corresponding FQG (right) where each node represents a text fragment and the edges represent replying relationships between fragments.

subgraph of G whose nodes are strongly connected represent a set of sentences that should belong to the same topical segment.

To identify subgraphs whose nodes are strongly connected, a k-way min-cut graph partitioning algorithm is applied on the graph $G(V,E)$ with the normalized cut (Ncut) criteria. Since Ncut is an NP-complete problem, an approximate solution is found following an efficient method proposed by Shi and Malik [118]. At the end of this process, each sentence of the conversation is assigned to one of the topical segments.

Topic Labeling

Topic labeling takes the segmented conversation as input and generates keyphrases to describe each topic in the conversation. The conversation is first tokenized and a syntactic filter is applied to select only nouns and adjectives from the text. Then a novel graph-based ranking model is applied that exploits two conversational features: information from the leading sentences of a topical segment and the FQG.

For this purpose, a heterogeneous network is constructed that consists of three subgraphs: the FQG; the word co-occurrence graph (G_W) that captures the co-occurrence of each word in the topic cluster with respect to the words in the leading sentence of that cluster; and a bipartite graph that ties these two graphs together. A co-ranking method [134] is then applied to this heterogeneous network to generate the ranked list of words for each topic. The top-M selected keywords from the ranked list are then marked in the text, and the sequences of adjacent keywords are collapsed into keyphrases. Finally, to achieve broad coverage of the topic, the Maximum Marginal Relevance (MMR) criterion is used to select the labels that are most relevant, but not redundant.

2.4.2 Sentiment Analysis

For sentiment analysis, we applied the Semantic Orientation CALculator (SO-CAL) [121], which is a lexicon-based approach for determining whether a text expresses a positive vs. negative opinion. SO-CAL computes polarity as numeric values. Its performance is consistently good across various domains and on completely unseen data, thus making a suitable tool for our purpose. At first, we apply SO-CAL to generate the polarity for each sentence of the conversation. We define 5 different polarity intervals, and for each comment in the conversation we count how many sentences fall in any of these polarity intervals. Then, we normalize the value in each polarity interval by the total number of sentences in the comment to compute the polarity distribution for that comment.

2.4.3 Corpora and Preprocessing

While designing and implementing ConVis, we have been mainly working with two quite different blog sources: Slashdot [2] (a technology related blog site) and Daily Kos [3] (a political analysis blog site). The Slashdot corpus, which was collected from [70], consists of 20 conversations annotated with topics by three human annotators. The other corpus was created by crawling blog conversations from the Daily Kos site.

After obtaining the conversations, we converted them into a common format (representing various metadata and the actual conversation) that our text mining methods can process. Then, we applied topic modeling and sentiment analysis to each conversation. For the Slashdot corpus, we automatically generate a topic model comprising of x topics, where x represents the average number of topics produced by the annotators for that conversation. Since the Daily Kos corpus was not annotated by any human rater, we simply used the average number of topics (i.e., 11) among all the blog conversations annotated in [70]. Finally, the results of topic modeling and sentiment analysis along with different metadata are mapped to the abstract data type as shown in Table 2.2.

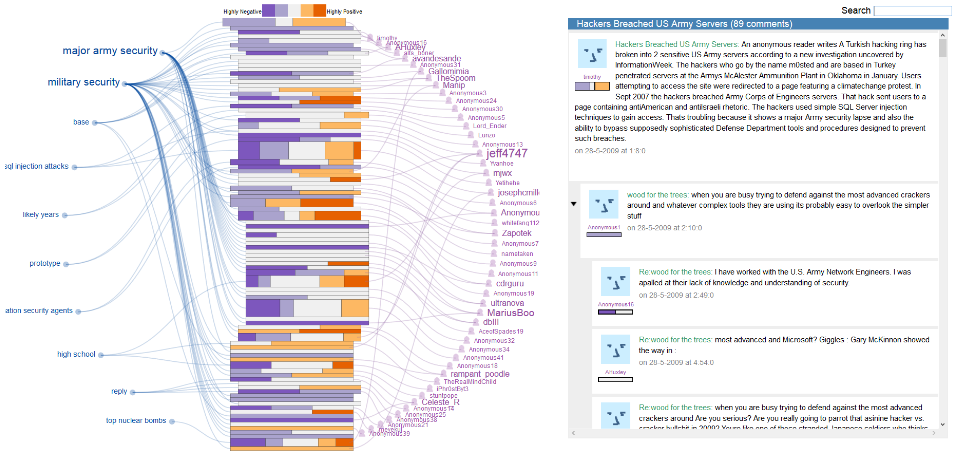
2.5 ConVis Design and Implementation

2.5.1 Visual Encoding

ConVis is designed to support multi-faceted exploration of blog conversations[2]. The visual encoding was guided by the design principles presented in Section 2.2, and the information to be presented is generated by the text mining techniques described in Section 2.4. A high-level design decision for the interface was to follow an overview+detail approach to deal with the space constraints. The rationale is that several studies have found the overview+detail approach to be more effective for text comprehension tasks than other approaches such as zooming and focus+context [24]. Overview+detail also allows us to provide information at multiple granularities *(DP-4)* by displaying a high-level overview of what was discussed by whom (i.e., topics and authors), a visual summary of the whole conversation (in the Thread Overview) and the most detailed view representing the actual conversation (see Figure 2.2). The interactions between these views are performed in a coordinated way. Below, we describe the design of each component along with careful justification of crucial design decisions.

The **Thread Overview** hierarchically represents a visual summary of the whole conversation, and allows the user to navigate through the comments (see Figure 2.2, middle). It displays each comment as a horizontal stacked bar. Each stacked bar encodes three different metadata (comment length, position in the thread, and depth of the comment within the thread) and the text analysis results (i.e., sentiment) for

[2] A video demonstration of ConVis is available here https://goo.gl/gBQE3e.



Hackers Breached US Army Servers (89 comments)

Hackers Breached US Army Servers: An anonymous reader writes A Turkish hacking ring has broken into 2 sensitive US Army servers according to a new investigation uncovered by InformationWeek. The hackers who go by the name m0sted and are based in Turkey penetrated servers at the Armys McAlester Ammunition Plant in Oklahoma in January. Users attempting to access the site were redirected to a page featuring a climatechange protest. In Sept 2007 the hackers breached Army Corps of Engineers servers. That hack sent users to a page containing antiAmerican and antiIsraeli rhetoric. The hackers used simple SQL Server injection techniques to gain access. Thats troubling because it shows a major Army security lapse and also the ability to bypass supposedly sophisticated Defense Department tools and procedures designed to prevent such breaches.

on 28-5-2009 at 1:8:0

wood for the trees: when you are busy trying to defend against the most advanced crackers around and whatever complex tools they are using its probably easy to overlook the simpler stuff

on 28-5-2009 at 2:10:0

Re:wood for the trees: I have worked with the U.S. Army Network Engineers. I was apalled at their lack of knowledge and understanding of security.

on 28-5-2009 at 2:49:0

Re:wood for the trees: most advanced and Microsoft? Giggles : Gary McKinnon showed the way in :

on 28-5-2009 at 4:54:0

Re:wood for the trees: when you are busy trying to defend against the most advanced crackers around Are you serious? Are you really going to parrot that asinine hacker vs. cracker bullshit in 2009? Youre like one of those stranded Japanese soldiers who thinks

Figure 2.2: A snapshot of ConVis for exploring blog conversation: The Thread Overview visually represents the whole conversation encoding the thread structure and how the sentiment is expressed for each comment(middle); The Facet Overview presents topics and authors circularly around the Thread Overview; and the Conversation View presents the actual conversation in a scrollable list (right). Here, topics and authors are connected to their related comments via curved links.

a comment, which are identified to be potentially useful navigational cues *(DP-1)*. The stacked bars are vertically ordered according to their positions in the thread starting from the top with indentation indicating thread depth, allowing the user to see the whole thread structure at a glance. The height of each bar encodes the normalized comment length, while the width of all the bars remain equal. Thus one could easily notice the differences in length among comments. The current implementation can reasonably show up to 200 comments when the visualization is used on a 1920 × 1080 screen. This scale was sufficient for all the conversations we have examined (see Table 2.2) and is plausibly adequate for the vast majority of blog conversations.

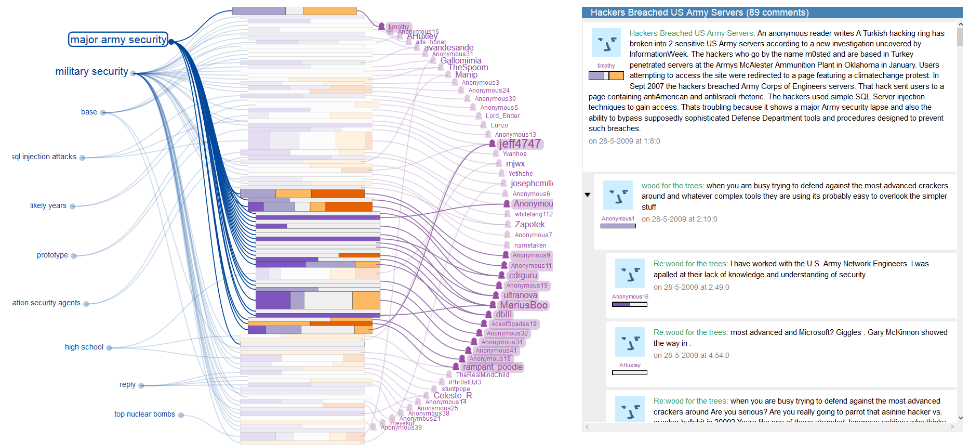
The distribution of sentiment orientation of a comment is encoded using color within each stacked bar, where width of each cell of a stacked bar indicates the number of sentences that belongs to a particular sentiment orientation. A set of five diverging colors was chosen from ColorBrewer [5] to visualize this distribution in a perceptually meaningful order, ranging from purple (highly negative, -2) to orange (highly positive, $+2$). Thus, the distribution of colors in the Thread Overview can help the user to perceive whether this conversation is mainly neutral /positive /negative, or very controversial. For example, if the Thread Overview is mostly in strong purple color, then the conversation has many negative comments.

Facet Overview: To support multifaceted exploration of the conversation *(DP-2)*, the primary facets, namely topics and authors are presented in a circular layout around the Thread Overview (see Figure 2.2). Topics and authors are presented to the left and right side of the Thread Overview respectively, creating a symmetric view. Both topics and authors are positioned according to their chronological order in the conversation starting from top, allowing the user to understand how the conversation evolves as the discussion progresses. Two distinctive qualitative colors are used to encode the facet links and the facet elements. The font size of a *topic* encodes how much it has been discussed when compared to the other topics within the whole conversation. Likewise, the font size of an *author* encodes how many times a participant has posted in a conversation. Thus, the font size of both facets helps the user to quickly identify what are the most discussed themes and who are the most dominant participants within a conversation.

To convey how facets and comments of the conversations are inter-related *(DP-3)*, the facet elements are connected to their corresponding comments in the Thread Overview via subtle curved links indicating topic-comment-author relationships (the relation between topic and comments can be many-to-many). While a common way to relate various elements in multiple views is synchronized visual highlighting, we choose visual links because it has been found that users can locate visually linked elements in complex visualizations faster and with greater satisfaction than plain highlighting [120]. By default, these visual links are drawn in the de-saturated tone of the corresponding facet's color.

The design decision of arranging facet elements in a circular layout is motivated by two primary reasons. First, more elements can be accommodated in this way than in a linear fashion. In fact, the current implementation can reasonably show up to 100 topics /authors when the visualization is used on a 1920×1080 screen. Second, a circular layout helps to encode the curved links between facets and comments without much visual clutter.

The **Conversation View** displays the actual text of the comments as a scrol-



Hackers Breached US Army Servers (89 comments)

timothy

Hackers Breached US Army Servers: An anonymous reader writes A Turkish hacking ring has broken into 2 sensitive US Army servers according to a new investigation uncovered by InformationWeek. The hackers who go by the name m0sted and are based in Turkey penetrated servers at the Armys McAlester Ammunition Plant in Oklahoma in January. Users attempting to access the site were redirected to a page featuring a climatechange protest. In Sept 2007 the hackers breached Army Corps of Engineers servers. That hack sent users to a page containing antiAmerican and antiIsraeli rhetoric. The hackers used simple SQL Server injection techniques to gain access. Thats troubling because it shows a major Army security lapse and also the ability to bypass supposedly sophisticated Defense Department tools and procedures designed to prevent such breaches.

on 28-5-2009 at 1:8:0

Anonymous1

wood for the trees: when you are busy trying to defend against the most advanced crackers around and whatever complex tools they are using its probably easy to overlook the simpler stuff

on 28-5-2009 at 2:10:0

Anonymous16

Re:wood for the trees: I have worked with the U.S. Army Network Engineers. I was apalled at their lack of knowledge and understanding of security.

on 28-5-2009 at 2:49:0

AHuxley

Re:wood for the trees: most advanced and Microsoft? Giggles : Gary McKinnon showed the way in :

on 28-5-2009 at 4:54:0

Re:wood for the trees: when you are busy trying to defend against the most advanced crackers around Are you serious? Are you really going to parrot that asinine hacker vs.

Figure 2.3: Hovering the mouse over a topic element ('major army security') causes highlighting the connecting visual links, brushing the related *authors*, and providing visual prominence to the related comments in the Thread Overview.

lable list (see Figure 2.2, right). Like in the Thread Overview, comments are indented according to their depth in the thread hierarchy, thus revealing the reply-to relationships. At the left side of each comment, the following metadata are presented: title, author name, photo, and a stacked bar representing the sentiment distribution (mirrored from Thread Overview). Overall, the Conversation View provides a familiar web discussion interface to the user, thus potentially enhancing the learnability for those who are accustomed to the current blog interfaces *(DP-5)*.

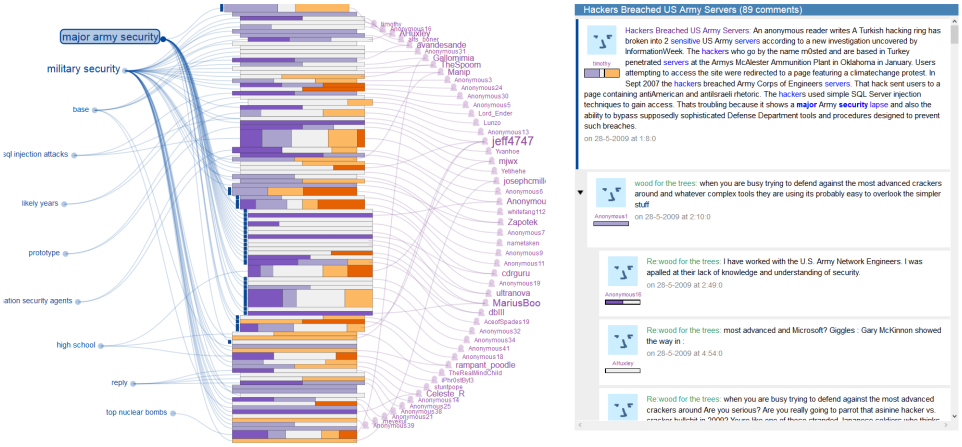
2.5.2 User Interactions

ConVis provides a set of lightweight interactions [77]. These interactions are designed so that they can be easily triggered without causing drastic modifications to the visual encoding, thus allowing the user to comprehend their effect without much cognitive overload *(DP-5)*.

Both overviews and the Conversation View interact in a coordinated way. Hovering the mouse over a facet element causes a rectangular border to be drawn around that element and subsequently highlights the connecting curved links by changing their color to a darker tone. This also causes brushing the elements in the other facet, and provides visual prominence to the related comments in the Thread Overview by de-saturating the rest of the stacked bars (see Figure 2.3). As such, the user can perceive relations between multiple facets *(DP-3)*. If the user becomes further interested in a facet element (e.g., a specific topic), she can subsequently select that item by clicking on it, resulting in drawing a thick vertical outline next to the corresponding stacked bars in the Thread Overview (see Figure 2.4). As a result, the comments of a particular topic/author remain persistently selected. The color of the vertical outlines is the same color as its facet, thus distinguishing between the selections of different types of facets. This encoding is also mirrored in the Conversation View (see Figure 2.4, right). Moreover, the user can select multiple facet items so that the comments shared among them become more apparent.

Highlighting and selection is also possible for each individual comment both from the Thread Overview and the Conversation View. Hovering the mouse over the stacked bar representing a comment causes it to be highlighted by drawing horizontal outlines on the top and bottom of the bar. It also causes the related topic(s) and author to be brushed along with the visual links connecting the comment to be highlighted. This highlighting is also mirrored in the Conversation View. Conversely, hovering the mouse over a comment in the Conversation View highlights the corresponding stacked bar in the Thread Overview. The user can subsequently select a comment either in the Thread Overview (see Figure 3.5) or in the Conversation View, so that this highlighting remains persistent unless the user toggle the state by clicking on it again.

A selection of a comment in the Thread Overview or of a facet in the Facet Overview causes scrolling to the relevant comment in the Conversation View via a smooth animation. In this way, the user can easily locate the comments that belong to a particular topic and/or author. Moreover, the keyphrases of the relevant topic and sentiments are highlighted in the Conversation View upon selection, providing more details on demand about what makes a particular comment positive/negative or how it is related to a particular topic. The user can also scroll through the comments with traditional interactions using the mouse wheel, or standard arrow and page keys. Finally, any branch of the conversation can be expanded/collapsed by clicking the up/down arrow to the left side of parent posts.



Hackers Breached US Army Servers (89 comments)

Hackers Breached US Army Servers: An anonymous reader writes A Turkish hacking ring has broken into 2 sensitive US Army servers according to a new investigation uncovered by InformationWeek. The hackers who go by the name m0sted and are based in Turkey penetrated servers at the Armys McAlester Ammunition Plant in Oklahoma in January. Users attempting to access the site were redirected to a page featuring a climatechange protest. In Sept 2007 the hackers breached Army Corps of Engineers servers. That hack sent users to a page containing antiAmerican and antiIsraeli rhetoric. The hackers used simple SQL Server injection techniques to gain access. Thats troubling because it shows a **major** Army **security** lapse and also the ability to bypass supposedly sophisticated Defense Department tools and procedures designed to prevent such breaches.
on 28-5-2009 at 1:8:0

wood for the trees: when you are busy trying to defend against the most advanced crackers around and whatever complex tools they are using its probably easy to overlook the simpler stuff
on 28-5-2009 at 2:10:0

Re:wood for the trees: I have worked with the U.S. Army Network Engineers. I was apalled at their lack of knowledge and understanding of security.
on 28-5-2009 at 2:49:0

Re:wood for the trees: most advanced and Microsoft? Giggles : Gary McKinnon showed the way in :
on 28-5-2009 at 4:54:0

Re:wood for the trees: when you are busy trying to defend against the most advanced crackers around Are you serious? Are you really going to parrot that asinine hacker vs.

Figure 2.4: Clicking on a topic results in drawing a thick vertical outline next to each of the related comments.

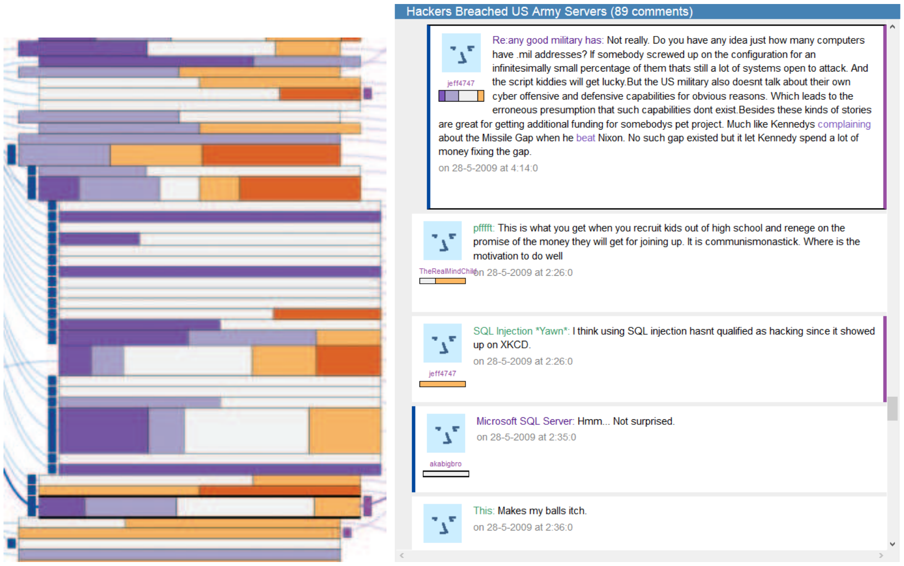
2.5.3 Implementation

A server-side component (in PHP) retrieves conversations annotated with topics and sentiment information. The visualization component, on the other hand, is implemented in JavaScript (using the D3 and JQuery library), which is sufficiently fast to respond in real time to the user actions[3].

2.6 Informal Evaluation

During the design and implementation of ConVis, we conducted formative evaluations to identify potential usability issues and to iteratively refine the prototype. Once the prototype was completed, we ran an informal evaluation [78] with a different set of target users to evaluate the higher levels of the nested model [91]. In this evaluation, we aimed to: 1) understand to what extent the overall visualization and its specific components are perceived to be useful by the potential users; 2) identify differences among users in how they performed the tasks; and 3) solicit ideas for improvements and enhancements.

[3] A live demo of ConVis is available here www.cs.ubc.ca/enamul/convis



Hackers Breached US Army Servers (89 comments)

jeff4747

Re:any good military has: Not really. Do you have any idea just how many computers have .mil addresses? If somebody screwed up on the configuration for an infinitesimally small percentage of them thats still a lot of systems open to attack. And the script kiddies will get lucky.But the US military also doesnt talk about their own cyber offensive and defensive capabilities for obvious reasons. Which leads to the erroneous presumption that such capabilities dont exist.Besides these kinds of stories are great for getting additional funding for somebodys pet project. Much like Kennedys complaining about the Missile Gap when he beat Nixon. No such gap existed but it let Kennedy spend a lot of money fixing the gap.

on 28-5-2009 at 4:14:0

TheRealMindChild

pfffft: This is what you get when you recruit kids out of high school and renege on the promise of the money they will get for joining up. It is communismonastick. Where is the motivation to do well

on 28-5-2009 at 2:26:0

jeff4747

SQL Injection *Yawn*: I think using SQL injection hasnt qualified as hacking since it showed up on XKCD.

on 28-5-2009 at 2:26:0

akabigbro

Microsoft SQL Server: Hmm... Not surprised.

on 28-5-2009 at 2:35:0

This: Makes my balls itch.

on 28-5-2009 at 2:36:0

(a) **(b)**

Figure 2.5: An example showing: (a) The user clicked on a comment (the one with horizontal outlines) in the Thread Overview. (b) As a result, the system automatic scrolled to the actual comment in the Conversation View.

2.6.1 Procedure and Participants

A pre-study questionnaire was administered to capture demographic information and prior experience of participants with blog reading. Then the ConVis interface was demonstrated to the participants. After that, they were allowed to choose three conversations of their interest from a set of six blogs from Slashdot, all of them having similar length (avg. number of comments is 91.33). Instead of asking some abstract questions (such as the ones in Table 2.1), we provided an open-ended task to reflect the exploratory nature of blog reading. We asked the participant to explore the conversations according to her own interests and write down a summary of the key insights (if any) gained while exploring each conversation. During the study, we primarily focused on gathering qualitative data such as observations, user-generated summaries, and semi-structured interviews.

We conducted the study with five participants (age range 18 to 24, 2 female),

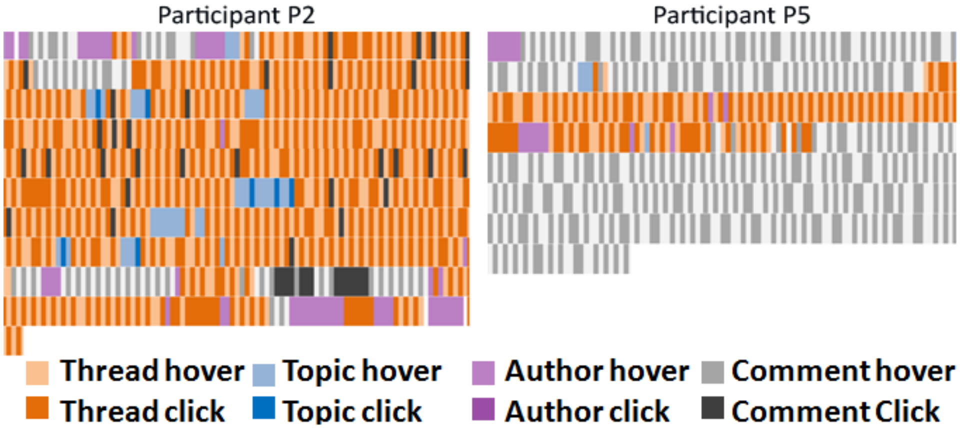


Figure 2.6: Comparison of uses patterns between two participants using the two different strategies on the conversation titled “Music Streaming to Overtake Downloads”.

who are frequent blog readers (four of them reported to read blogs at least several times a day and one reported several times a week). The three most common reasons for them to read blogs are information seeking, guidance/opinion seeking, and enjoyment. They are primarily interested in blogs about technology, politics, and education.

2.6.2 Results and Analysis

Browsing strategies: From the interaction log data and semi-structured interviews, we identified two main strategies for reading comments: exploring by topic facets, and skimming through detailed comments. Figure 2.6 shows the sequence of interface actions made by participant P2, who followed the former strategy, and P5 who followed the latter, on the same conversation. Overall, of the five participants, two followed the exploration by topic strategy, while the other three followed the skimming comments one. The two participants who followed the former strategy, reported that they would begin by quickly scanning the topics and selecting either the most discussed topic first or the ones that were interesting to them, and then reading the comments linked to that topic. We also observed that to find the comments of interest in the selected topic, they often relied on the sentiment and comment length encoded in the Thread Overview. After going through the comments on a specific topic, they either went on reading the next topic that appeared in the conversation, or went back to scan the topic list to find the next topic of interest. This navigational behaviour can be observed from the sequence of actions made by Participant P2 (see Figure 2.6(left)). The other three participants followed the traditional way of blog reading, primarily skimming through the comments in the Conversation View. This is illustrated in Figure 2.6 (right), where P5 mainly hovered over different comments. However, at the same time these participants acknowledged that they tended to coordinate with the topics and Thread Overview where the related items were highlighted so that they could get a sense of what part of the conversation they were reading and when the topic was about to change. Supporting evidence came from interaction log data, where those who followed the first strategy, on average clicked on different topics 13 times and hovered 68 times for each conversation. On the contrary, those who followed the second strategy hovered only 11 times on average per conversation and never clicked on a topic.

Interface features: In general, all participants, independently of their preferred browsing strategy, agreed that showing the set of topics and then visually linking them to the comments in the Thread Overview helped them to quickly understand what a conversation is about and to focus on its most interesting parts. P2 said: *"I just try to find topics that are interesting to me which is really useful. I could look into a comment of that topic and then look at other comments replying to that comment, so this navigation feature was really good."* Another useful feature according to the participants is the Thread overview displaying the comments and sentiment. P1 said: *"In the visualization, it is very clear to see what kind of article I am going to dealing with... the last conversation has lot of purple, indicating its something going to have many negative comments"*; however, P3 reported that the sentiment classification was incorrect in some cases, making it less reliable. Encoding the comment length was found to be useful to P4: *"The height of the bar was really useful, cz the thicker comments were generally more interesting and insightful than the shorter ones."*.

Users were also interested in seeing how much an author contribute to a specific topic. According to one participant, *"My primary interest with the author would be to see how much they have participated back into the topic and that happens* *in various occasions, so I found the linking between topics and authors quite useful'.* P2 also found some utility of the author facet: '*"If I find someone's comment interesting, then I wanna know what other comments she made, and how people reacted to that."* In such scenario, linking the comments to the corresponding author is valuable. But participants also emphasized that if they would have been part of the community, the author facet would have been much more useful: *"If I would know some people, I would be really interested in what they are saying. But since these are random people, I don't know if I would incline to care"* (P1). The participants also acknowledged that if they had been participants in the conversation, they would have been interested to know who is replying to their posts.

Preference: When the participants were asked to compare their experience using ConVis with their regular blog reading interface, the answers were generally in favour of ConVis, due to its ability to show a visual overview of the whole conversation and allowing the user to explore through facets. Moreover, the visualization tool was found to be easy to learn by the participants. According to P1: *"Seeing the sort of pagination in current interfaces, you don't get the overall. I have to read through all of them."* On the contrary, *"Using ConVis I would read more important parts of the conversation as opposed to just people talking. I can navigate through the comments without actually reading them, which is really helpful."* P5 who followed the strategy of skimming through the conversation mentioned: *"I am so much used to scroll up and down in the list of comments, but using this additional visual overview, I had a sense of where I am reading right now and what topic I am currently reading".* P2 said that ConVis provides a quicker way to explore comments: *"It allows me to navigate through the most insightful stuff out of five minutes which could take say 15 minutes otherwise. Actually I found many comments to be interesting towards the end of conversations, which I probably wouldn't notice if I would use my blog interface."*

2.6.3 Revisiting Task Abstraction

Analyzing the user-generated summaries from the evaluation helps us to reflect on the task abstraction in Section 2.2. After mapping each sentence of the summaries to one or more possible tasks in Table 2.1, we find that some of the tasks were performed more frequently than others. All of the participants answered Q1 and Q2 in their summaries, suggesting that understanding the *topics* is a fundamental task. A substantial portion of each summary answers questions Q3 through Q6, which are related to the *opinion* variable. We also realize that the exploratory behaviour can be largely influenced by participant's own viewpoints (Q5) and what they perceive as interesting/funny (Q10). However, the summaries reveal very little interest of the participant in looking for questions specific to *authors* (Q7 through Q9), suggesting that being a part of the community might be highly relevant for these tasks as mentioned by a participant. Thus, it is important to consider the characteristics of the target blog community into the design process.

2.7 Discussion

Based on the results and analysis of the informal evaluation, we discuss more generally various visualization design issues and directions for future improvements.

Improve faceted exploration: An important aspect of our visualization was to explicitly depict the relations between multiple facets of the conversation with the related comments. However, depending on the tasks additional facets can become more useful to the participants (e.g., moderation scores, named entities), while an existing facet being less useful (e.g., author). In the future, we plan to devise an interactive visualization technique that allows the user to dynamically change the facets of interest and reveal relations between them.

Enhance scalability: On scalability, while ConVis can deal with conversations with hundreds of comments, additional techniques are needed for longer conversations. In some cases when the discussion topic is very popular, the conversation can become very large with thousands of comments. To deal with such situations, we suggest integrating additional computational methods such as detecting high quality comments [45] to guide the way of filtering and aggregating comments, as well as to apply focus+context techniques to the Thread Overview.

Need for human-in-the-loop model: In general, the utility of visual text analytic systems can be substantially improved if more accurate natural language processing techniques are adopted. Even though the text analytic methods used in this chapter achieve significantly higher accuracy than traditional methods [70], the informal evaluation reveals that still in few cases the extracted topics and opinions are incorrect. In particular, during the interviews users expressed a pressing need for enhancing their ability to revise the topic model according to their own information needs.

In such cases, a promising approach could be to incorporate users feedback in the text mining loop, so that the underlying models can be iteratively refined. Motivated by this experience, we have designed ConVisIT (an extended version of ConVis), by incorporating an interactive topic modeling approach. This extended interface supports the user in revising the topic model, while she is exploring the conversation. We discuss this interactive topic modeling approach in details in chapter 4.

Further user evaluation: While the informal evaluation provided some preliminary feedback from users about ConVis, further evaluations were necessary to compare this interface with regular blog reading interface. For this purpose, later we conducted a summative evaluation [78] using a lab-based study to understand the effectiveness of ConVis compared to traditional blog reading interfaces as well as an interface that supports interactive topic modeling (i.e., ConVisIT). We discuss the results of this study in Chapter 4.

2.8 Summary

We have presented ConVis, a visual text analytic system designed to support the exploration and analysis of blog conversations. Our approach incorporates novel mining methods that take advantage of conversational features, with interactive visualization that supports multifaceted exploration. The participants' feedback from our informal evaluation suggests that ConVis can help the user to simultaneously explore the topics and opinions expressed in the conversation; supporting the user in finding comments of interest, even if they are buried near the end of the thread. Interestingly, ConVis is beneficial also to users who follow the traditional strategy of scrolling through the Conversation View, because the other views provide situational awareness (e.g., what topic is expected next).

Exploring a large set of conversations is arguably an even more challenging task than exploring only one conversation, because the volume and complexity of the textual data may drastically increase and the information overload problem could be even more prevalent and serious among users [69]. Therefore, in our subsequent work, we have extended our approach to handle a large collection of asynchronous conversations, where the user is able to explore topics that are discussed over many different threads. In the next chapter, we discuss this approach for exploring a set of conversations in detail.

Chapter 3

Supporting Users in Exploring a Set of Conversations

In Chapter 2, we presented ConVis, a visual text analytics system for exploring a single conversation. We now describe how we have extended the ConVis system, that we called MultiConVis, to support users in exploring and analyzing a collection of conversations. The resulting system supports the user exploration, starting from a possibly large set of conversations, then narrowing it down to a subset of conversations, and eventually drilling-down to comments of one conversation. Similarly to what we did for ConVis, the development of MultiConVis is based on the integration of NLP techniques for topic modeling and sentiment analysis with information visualizations, by considering the unique characteristics of online conversations. Later in this chapter, we present a set of case studies with domain experts and a formal user study with regular blog readers, which illustrate the potential benefits of our approach, when compared to a traditional blog reading interface[1].

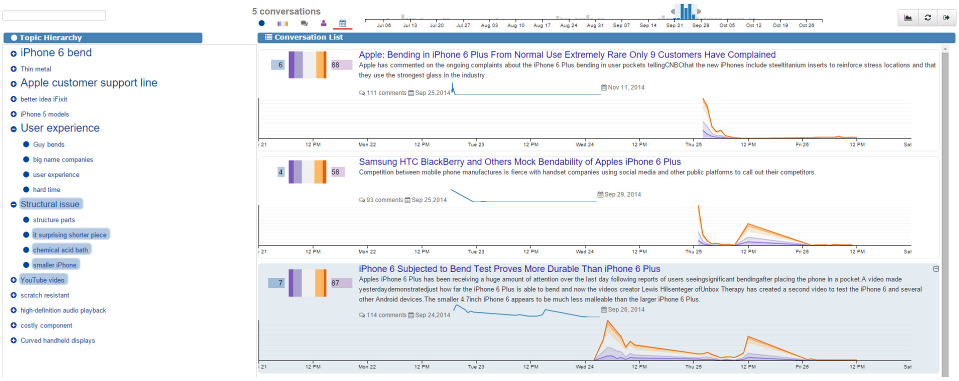
[1] This chapter is a modified version of our paper *MultiConVis: A visual text analytics system for exploring a collection of online conversations*, by Enamul Hoque and Giuseppe Carenini; in Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI), pp. 96-107, 2016 [60].

3.1 Introduction

With the proliferation of web-based social media, there has been an exponential growth of asynchronous online conversations discussing a large variety of popular issues like ‘ObamaCare’, ‘US immigration reform’, and ‘Apple iWatch release’. Given a query, traditional blog sites only present the set of relevant blogs as a paginated list ordered by their recency, without providing any high-level summary of the conversations. This navigational support is often inadequate to explore a set of blogs that may be of great interest to readers [68].

To understand the problem, let us recall the ‘iPhone bending’ query example introduced in Chapter 1. After the iPhone 6 was launched, some people claimed that this new phone can easily bend in the pocket. This incident triggered a lot of discussions in Macrumors [1], a blog site for Apple-related news. Within just a few days, more than a dozen conversations with thousands of comments were generated in Macrumors covering various related topics. In this context, we could imagine three different users who would like to explore this set of conversations. First, a *potential customer*, who intended to buy an iPhone may want to explore these conversations to verify whether the bending issue is really serious or not. Second, a *journalist* may want to publish a story about what people are saying about this issue by analyzing this set of conversations. Finally, an *Apple marketing analyst* may want to know how the online community is responding to this issue to make an informed decision about how to react to the rumors and possibly redesign the products. In all these cases, given the large number of conversations/comments, it would be difficult and time-consuming for a user to explore and analyze all this information with traditional blog interfaces, which only provide sequential access to conversations/comments.

In this work, we tightly couple NLP techniques for topic modeling and sentiment analysis with interactive visualizations to support the exploration and analysis of a large set of conversations by considering the specific characteristics of blog conversations. As we have pointed out in Chapter 1, blog conversations exhibit several unique characteristics: unlike microblog or messaging [108], they do not have fixed length comments; furthermore, they have finer conversational structure as participants often reply to a post and/or quote a fragment of other comments [70].



Apple: Bending in iPhone 6 Plus From Normal Use Extremely Rare Only 9 Customers Have Complained

Apple has commented on the ongoing complaints about the iPhone 6 Plus bending in user pockets tellingCNBCthat the new iPhones include steeltitanium inserts to reinforce stress locations and that they use the strongest glass in the industry.

111 comments Sep 25,2014 Nov 11, 2014

Samsung HTC BlackBerry and Others Mock Bendability of Apples iPhone 6 Plus

Competition between mobile phone manufactures is fierce with handset companies using social media and other public platforms to call out their competitors.

93 comments Sep 25,2014 Sep 29, 2014

iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus

Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.

114 comments Sep 24,2014 Sep 26, 2014

Figure 3.1: The MultiConVis interface, showing a subset of blog conversations returned by the query 'iPhone bending' from Macrumors in November 2014. Here, the user filtered out some conversations from the list using the Timeline located at the top, and then hovered on a conversation item (highlighted row in the right). As a consequence, the related topics from the Topic Hierarchy were highlighted (left).

In this chapter, we consider these unique characteristics in devising our novel NLP and InfoVis techniques.

We built the MultiConVis system on top of ConVis (described in Chapter 2). As we move from a single conversation to a collection of conversations, critical challenges emerge from the fact that users need to deal with a much larger amount of data, with different levels of granularity. For instance, the number of topics increases drastically for a set of conversations, therefore understanding and exploring these topics can be much more time consuming and cumbersome. Since some of these topics are similar in their semantic meaning, grouping them into a hierarchical topic organization may support the understanding and navigation of topics more effectively.

To address this challenge, we devise a hierarchical topic modeling technique that organizes the topics within a set of conversations into multiple levels, based on their semantic similarity. The resulting topic hierarchy is intended to better support user's understanding and navigation of the topics. We then design a visual interface that presents the hierarchical topic structure along with other conversational data as shown in Figure 3.1. The main contributions of this work are:

1) A **hierarchical topic modeling method** over a collection of conversations. While Chapter 2 describes how to effectively extract topics from a single conversation, here we propose a method which creates a topic hierarchy for a whole collection of conversations, by aggregating the topics extracted from each conversation in the collection.

2) The design and implementation of the **MultiConVis** interface, which supports exploration of a collection of blog conversations based on the topic hierarchy and sentiment. In essence, MultiConVis can be seen as an interface built on top of ConVis to allow the user to seemingly switch from exploring a collection of conversations to a single conversation. In particular, MultiConVis initially visualizes all the conversations in the whole collection, next supports the user in filtering out conversations that are irrelevant to her information needs, and then allows the user to drill down to a specific conversation, which is visualized with the ConVis interface.

3) The evaluation of MultiConVis through a set of **case studies**, and a **user study** to investigate how the system influences user performance and subjective opinions when compared to a sample, traditional blog reading interface similar to existing interfaces, like Slashdot [2] and Macrumors [1].

3.2 Related Work

In Chapter 2, we have already provided an overview of related work which primarily focused on visualizing a single online conversation. Here, we discuss research prototypes that aim to support the exploration of a large collection of conversations. These prototypes can be categorized based on the information they extract and visualize: (a) metadata of the conversations, such as timestamps, tags, and authors, (b) the results of text analysis, such as topic model and opinion.

3.2.1 Metadata Visualization

Some earlier works have focused on how to support the exploration of a blog archive using only metadata, for example, by visualizing tags and comments arranged along a time-axis [68], or by providing faceted visualization widgets for visual query formulation according to time, place, and tags [36]. While these works may assist users to find the blogs they are looking for, they are not designed to support users in understanding the actual content (i.e., the text) of these conversations. However, many tasks for blog readers, that we have identified in Chapter 2, require the user to get overviews of the actual content of a collection of conversations, such as "Find out what are people feeling about *X* over time." Therefore, our goal is to visualize a combination of various metadata and textual analysis results that are identified as important in our user requirements analysis.

3.2.2 Topic Modeling and Visualization

In contrast to simply showing the metadata of the conversations, recently there have been some attempts to visualize the topics discussed within a collection of conversations [37, 124, 127]. A common approach is to use probabilistic topic models such as Latent Dirichlet Allocation (LDA), where topics are defined as distributions of words and documents are represented as a mixture of topics. Many of these works also consider the temporal aspects of topics by showing the evolution of topics over time. For example, *Themail* visualizes how topics in a collection of email conversations develop over time by arranging keywords selected based on term-frequency inverse document-frequency (TF-IDF) along a horizontal time axis [124]. *TIARA* [127] represents the temporal evolution of topics from an email collection by applying the ThemeRiver visualization [53], where each layer in the stacked graph represents a topic and the keywords of each topic are distributed over time. From the height of each topic and its content distributed over time, the user can see the topic evolution.

More recent works have tried to move beyond visualizing topics as a flat list, by organizing them into a hierarchy [28, 40, 82]. For example, *HierarchicalTopics* organizes a large number of topics into a tree structure by considering the distance between the probability distributions of topics [40]; and then utilizes a hierarchical ThemeRiver view to explore temporal trends of topics. Using the same algorithm, *TopicPanorama* builds topic hierarchies from multiple corpora (i.e., news, blogs, and microblogs), followed by matching these hierarchies using a graph-matching technique, so that the common and distinctive topics from different corpora can be visualized [82]. It combines a radially stacked tree visualization with a density-based graph visualization to facilitate the examination of the matched topic graph from multiple perspectives. Compared to these approaches that generate static topic hierarchies, *RoseRiver* focused on exploring the evolutionary patterns of hierarchical topics generated at different timeframes by conveying topic merging and splitting relationships over time using Sankey diagrams [28].

Organizing topics into a hierarchy can be very useful to our work as well, because the number of distinct topics in a collection of conversations may be quite high, compared to a single conversation. However, existing hierarchical topic modeling approaches are not designed specifically for conversational data. In contrast, MultiConVis creates a topic hierarchy for a collection of conversations by aggregating the topics of each conversation. And such topics are generated by taking specific characteristics of asynchronous conversations such as reply-relationship into account [59].

3.2.3 Opinion Visualization

There is a growing interest in visualizing the opinions expressed in conversations, mostly focusing on microblogs [33, 85, 129]. Diakopoulos et al. presented *Vox Civitas* [33] that displayed sentiment and tweets volume over time for events discussed in microblogs to support the tasks of journalistic inquiry. *TwitInfo* [85] was also designed for visualizing microblogs with a focus on providing more accurate aggregation of sentiment information over a collection of tweets. Unlike these works, *OpinionFlow* focused more on visualizing the spreading of opinions about a particular topic (e.g., ‘US government shutdown’) among participants with a combination of a density map and a Sankey diagram [129]. Often the opinion information is summarized with other important aspects of information spreading such as temporal information, and the connections among conversation threads and authors [133].

A critical issue when abstracting data for sentiment analysis is how to aggregate sentiment information across sentences, comments, and conversations. While all the works described above dealt with twitter data, in which tweets are only organized as a list, here we focus on a set of much more structured blog conversations, where each conversation consists of a set of comments organized in multiple threads with reply-relationships. We exploit this additional structure when we visually represent sentiment over multiple, different levels.

3.3 User Requirements Analysis

In Chapter 2, we analyzed why and how people read blogs and used this analysis to derive the data and task abstractions. Here, we are going to identify useful data abstractions for a set of conversations and compare them with data abstractions for a single conversation.

In essence, the primary goals of reading blogs include information seeking, fact checking, and opinion seeking [30, 72], which require the reader to understand what *topics* are discussed in the conversations and what *opinions* are expressed on those topics. Furthermore, users often exhibit a *variety seeking behaviour*, i.e., they tend to switch frequently from a topic to its sub-topics or to a completely different topic [119].

Blog readers also care about temporal aspects of the conversations [31, 57], for instance, the start and end time of a conversation, the chronological position of a comment with respect to the other comments within a conversation [13], and the volume of comments over time when exploring multiple conversations. Information about authors of the comments is also considered to be valuable [57], especially for blogs in which the same users participate frequently.

Table 3.1 summarizes our design choices for what information our interface should display, in light of the current literature on blog readers. The row in the table corresponds to data facets and the columns to whether the facet is for multiple conversations vs a single one.

Since the number of topics for a collection of conversations is potentially much larger than for a single conversation, all the *topics* within a collection are organized into a hierarchy, while the topics of each single conversation are organized as a flat list and are explicitly connected to the comments of that conversation. To support the goals related to the *time* facet, the volume of comments over time is computed for each conversation in the collection of conversations, whereas within each conversation the chronological position of the comments is used. For the *sentiment* facet the distribution of sentiments across five polarity intervals, ranging from -2

Levels Facets	**Collection of Conversations**	**One Conversation**
Topics	Hierarchy with all topics from all conversations	Conversation-level topics with explicit links to the topic hierarchy and to comments
Time	- Start day/time - Volume of comments over time	Ordinal time representations: comments are ordered chronologically
Sentiment	- Sentiment distribution for each conversation - Sentiment evolution over time for each conversation	Sentiment distribution for each comment
Authors	Number of authors for each conversation	Conversation-level authors with explicit links to comments

Table 3.1: A summary of how facet elements are abstracted for a collection of conversations vs. one conversation.

to +2, is computed by counting how many sentences fall in each of these intervals. Here, for a collection of conversations, we compute the sentiment distribution for *each conversation*, whereas for one conversation, we compute this distribution at a finer level, i.e., for *each comment*. Finally, for the *authors* facet, while for a set of conversations only counts of authors are computed without providing the detailed list of authors, for one conversation the list of authors for that conversation is shown.

Current literature on blog reading not only inspired our data choices, but also guided the development of MultiConVis interactive visualization techniques. Considering the exploratory nature of blog reading, MultiConVis supports the user in browsing the set of conversations and comments by means of all the key facets, namely topics, sentiment, and authors. Furthermore, the interface facilitates the exploration through the facets at different levels of granularity: from all conversations, to a subset of conversations, to one conversation. For consistency, elements of the same facet across different levels of granularity have similar visual mappings

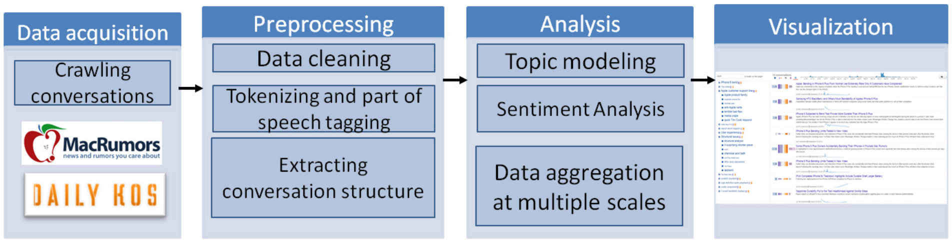
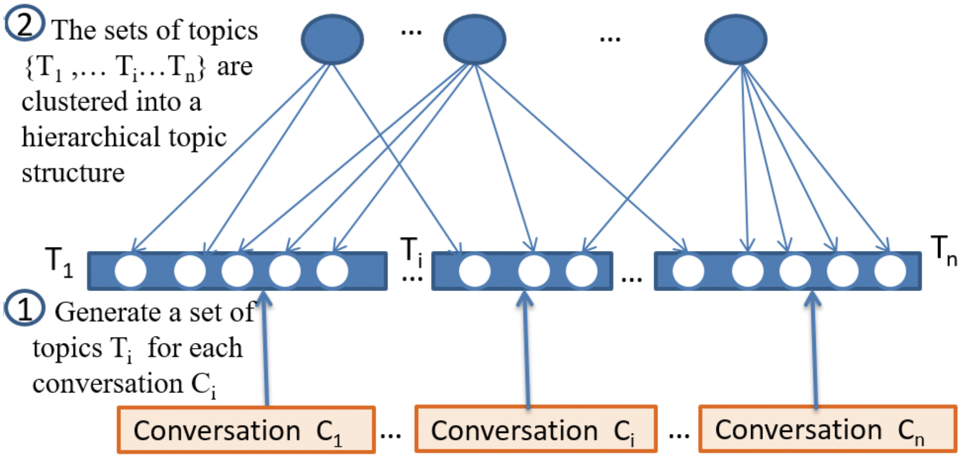

Data acquisition
Crawling conversations
MacRumors
news and rumors you care about
DAILY KOS
Preprocessing
Data cleaning
Tokenizing and part of speech tagging
Extracting conversation structure
Analysis
Topic modeling
Sentiment Analysis
Data aggregation at multiple scales
Visualization

Figure 3.2: Overview of the MultiConVis system.

in terms of color, shape, and other visual encoding channels. Finally, to facilitate the exploration and filtering of conversations, important attributes of each conversation, namely the number of topics/authors, the number of comments, and the overall sentiment distribution are encoded as information scent [128].

3.4 System Overview

The MultiConVis system consists of four major components as shown in Figure 3.2. Given a specified query (e.g., ‘iPhone bending’), the *data acquisition module* invokes a blog site such as Macrumors to crawl the set of conversations obtained from the first page of the search results returned by that site. Next, the *preprocessing module* performs data cleaning to retain only the conversational data in the crawled pages, followed by extracting the conversational structure, i.e., reply-relationships and quotation. We also use a state-of-the-art tagger [8] to tokenize text and annotate the tokens with their part-of-speech tags. After that, the *analysis module* performs topic modeling and sentiment analysis over the whole set of conversations. It then aggregates both metadata and results of text analysis at different granularity levels as described in the user requirements analysis. Finally, the *visualization module* displays the results obtained from the analysis module, and supports the user to interactively explore the conversations.



2 The sets of topics {T1 ,… Ti…Tn} are clustered into a hierarchical topic structure
…
…
T1
Ti
…
…
Tn
1 Generate a set of topics Ti for each conversation Ci
Conversation C1
…
Conversation Ci
…
Conversation Cn

Figure 3.3: Hierarchical topic model generation.

3.5 Text Analysis

3.5.1 Topic Hierarchy Generation

Our topic modeling approach takes a collection of n blog conversations $C = \{c_1, c_2, ..., c_n\}$ that satisfies a user query and generates a topic hierarchy following a bottom-up approach. In the resulting hierarchy, each node represents the cluster of sentences in the conversations that discuss the topic described by the label of the node. One could think of a top-down approach to be more suitable for generating the topic hierarchy, as it considers the whole set of conversations while generating the initial set of clusters (the roots of the hierarchy); however, we choose a bottom-up approach because in this way we are able to take into account the conversational structure extracted from each conversation. In other words, we first generate a set of topic clusters for each conversation by taking advantage of its conversational structure, and then we organize these topic clusters from all the conversations into a hierarchy. More specifically, our topic hierarchy generation involves two primary steps as shown in Figure 3.3: 1) generate a set of topics T_i for each conversation $c_i \in C$; 2) aggregate all the T_i into a hierarchical topic structure for the whole collection.

3.5.2 Topic Modeling Over Each Conversation

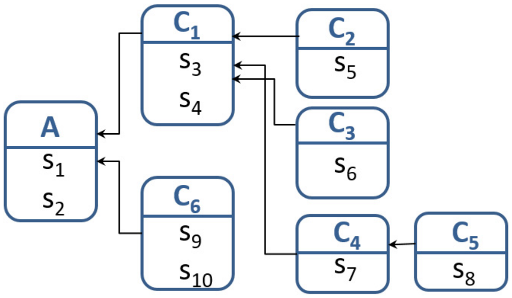
In order to generate a topic model over each conversation, we adopt the method described in Chapter 2. We briefly summarize it here, because our topic modeling method for a collection of conversations exploits similar data structures and techniques. Topic modeling of a single conversation starts by grouping the sentences of the conversation into a number of topic clusters (*segmentation*). Then, representative key phrases are assigned to each of these clusters (*labeling*).

In essence, topic segmentation applies a Lexical Cohesion-based Segmenter (LCSeg) [49] to each thread in the conversation as shown in Figure 3.4, where each thread represents a path from the initial message to a leaf message. Notice that after running the LCSeg algorithm, two sentences (e.g., s_1 and s_4) may appear together in the same segment in one thread ($A, C1, C2$), while falling into different segments in another thread ($A, C1, C5$). To consolidate all the (possibly conflicting) segmentation decisions made on each thread, we apply an efficient min-cut graph partitioning algorithm [118]. The optimal number of topics for each conversation is automatically determined by maximizing a clustering objective function proposed in [97].

Topic labeling takes the segmented conversation as input and generates a set of keyphrases to describe each topic cluster in the conversation. This is done by adapting the co-ranking method proposed in [134], in which a list of the top keyphrases is extracted from a graph of words that captures the co-occurrence of each word in the topic cluster with respect to the words in the leading sentence of that cluster, as well as the position of each word with respect to the thread structure of the conversation.

Creating the Topic Hierarchy Over the Collection

This is the key computational contribution of this chapter. Once the sets of topics T_i for each conversation c_i are generated, we organize all of them into a single topic hierarchy to create a structured overview of the whole collection of conversations. To achieve this, we have devised a graph-based method similar to the one that we apply to single conversations. The main difference here is that the nodes of the graph we create are not sentences anymore, but topics.



(a)

Thread	Sentences
A,C1,C2	$s_1, s_2, s_3, s_4 \mid s_5$
A,C1,C3	$s_1, s_2, s_3, s_4 \mid s_6$
A,C1,C4,C5	$s_1, s_2 \mid s_3, s_4 \mid, s_7, s_8$
A,C6	s_1, s_2, s_9, s_{10}

(b)

Figure 3.4: a) Reply-to relationships between the initial post A and the comments $C_1, C_2, ..., C_6$ of a conversation (left). Each post may comprise of one or more sentences as denoted by $s_1, s_2, s_3, ..., s_{10}$. b) the corresponding list of threads along with segmentation results after running the LCSeg algorithm on each of these threads. Here, the segmentation boundary is denoted by '|' (right).

In particular, we create a weighted undirected graph $G(V_C, E_C)$, where the nodes V_C represent the union of all the topics T_i from the set of conversations $C = \{c_1, c_2, ..., c_n\}$ and the edge weight $w(x;y)$ in E_C, between any two given topic nodes x and y, are generated by computing the average similarity between all pairs of sentences, in which one sentence belongs to topic x and the other one belongs to topic y. More formally, consider S_x is a set of l sentences and S_y is a set of m sentences for topics x and y respectively. Then we compute the edge weight $w(x;y)$ as follows:

$$w(x;y) = \frac{1}{l \times m} \sum_{s_i \in S_x, s_j \in S_y} sim(s_i, s_j) \tag{3.1}$$

Here, $sim(s_i, s_j)$ is the measure of similarity between a pair of sentences s_i and s_j. This measure is based on cosine similarity between s_i and s_j, if topic x and topic y belong to two different conversations c_x and c_y. Also, the same cosine similarity measure is used when s_i and s_j are from the same conversation, but never appear in the same segment in the segmentation results of the LCSseg algorithm. However, if s_i and s_j are both from the same conversation and they appear together in the same segment at least once, then the similarity is determined by k, where k is the number of times ($k >= 1$) in which s_i and s_j appeared in the same segment. This is based on the intuition that two topics that are from the same conversation and have stronger cohesion in the threads of that conversation should be more likely to be clustered together than those that do not. More formally,

$$sim(s_i, s_j) = \begin{cases} \begin{cases} CosineSim(s_i, s_j) \end{cases} & \text{if } c_x \neq c_y \\ \begin{cases} k & \text{if } k >= 1 \\ CosineSim(s_i, s_j) & \text{if } k = 0 \end{cases} & \text{else} \end{cases} \tag{3.2}$$

$$CosineSim(s_i, s_j) = \frac{\sum_{w \in s_i, s_j} tf_{w,s_i} \cdot tf_{w,s_j}}{\sqrt{\sum_{p \in s_i} tf^2_{p,s_i}} \cdot \sqrt{\sum_{q \in s_j} tf^2_{q,s_j}}} \tag{3.3}$$

$$0 \leq CosineSim(s_i, s_j) \leq 1 \tag{3.4}$$

Here, $tf_{a,b}$ denotes the term frequency of term a in the sentence b [2].

Once we have built the graph $G(V_C, E_C)$, we apply the same graph partitioning algorithm used in topic segmentation for single conversation, i.e., approximate solution to n-Cut [118] on $G(V_C, E_C)$. As a result, topic nodes that are mostly similar i.e., strongly connected in $G(V_C, E_C)$ will form n different clusters. Each of these clusters can be interpreted as a parent topic (in the topic hierarchy) of all the topic nodes that form that cluster. Here, the number of clusters n is automatically determined by maximizing a clustering objective function proposed in 4.1 [97].

For the final step of topic labeling, we assign a set of keyphrases to each parent topic by taking all the sentences from all the children topic nodes under it, and by then extracting and ranking keyphrases from all those sentences. This process is similar to the topic labeling method described for a single conversation, except that given the absence of a thread structure between multiple conversations, we modify the ranking process by creating a graph that only captures word co-occurrence relationships.

[2] For the sake of simplicity, we measured the cosine similarity between two sentences based on word frequency. Nevertheless, one could replace this simple representation with more recent neural embeddings for sentences like [75] to obtain better performance.

3.5.3 Sentiment

For sentiment analysis, we apply the Semantic Orientation CALculator (SO-CAL) [121], which has been shown to work well on user-generated content. SO-CAL computes sentiment polarity as numeric values. At first, we generate the polarity for each sentence of the conversation using SO-CAL. We defined five different polarity intervals (-2 to +2), and aggregate the results at various levels. For instance, at the level of a single conversation for each *comment*, we count how many sentences fall in any of these polarity intervals to compute the polarity distribution for that comment. Similarly, when dealing with a set of conversations, for each *conversation* we count how many sentences fall in any of these five polarity intervals to compute the polarity distribution for that conversation.

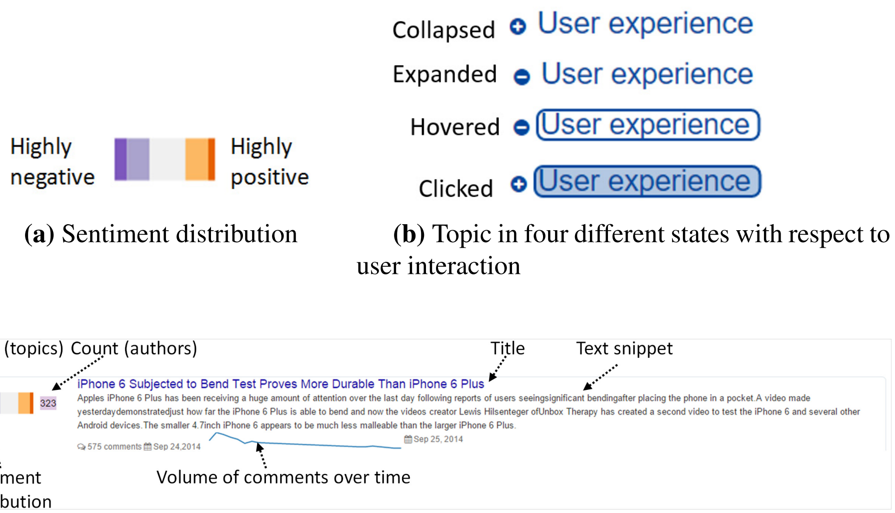
3.6 MultiConVis

In order to explore various design choices, we carried out an iterative design process, starting from early mockups and prototypes, to a fully functional system. Throughout this process, we performed formative evaluations [78] to identify potential usability issues and to iteratively refine the prototype. We now present the final design of the MultiConVis interface[3], along with justifications for the key design decisions based on our user requirements analysis and the InfoVis literature.

3.6.1 Visual Encoding

Facets: As mentioned earlier, a key design goal of MultiConVis is to facilitate the exploration of a set of conversations at multiple levels of granularity, while maintaining consistent visual mapping across different levels. We maintained consistency in the visual encodings across different levels as follows: 1) Sentiment distributions are represented in the same way (as a stacked bar) for a conversation, for a topic as well as for a comment (see Figure 3.5a). A set of five diverging colors was used in a perceptually meaningful order purple (highly negative) to orange (highly positive) to visualize the distribution of sentiment orientations

[3] A video demonstration of MultiConVis is available here https://goo.gl/ZmVYks.



iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus

Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.

575 comments Sep 24,2014 Sep 25, 2014

(a) Sentiment distribution

(b) Topic in four different states with respect to user interaction

(c) Conversation

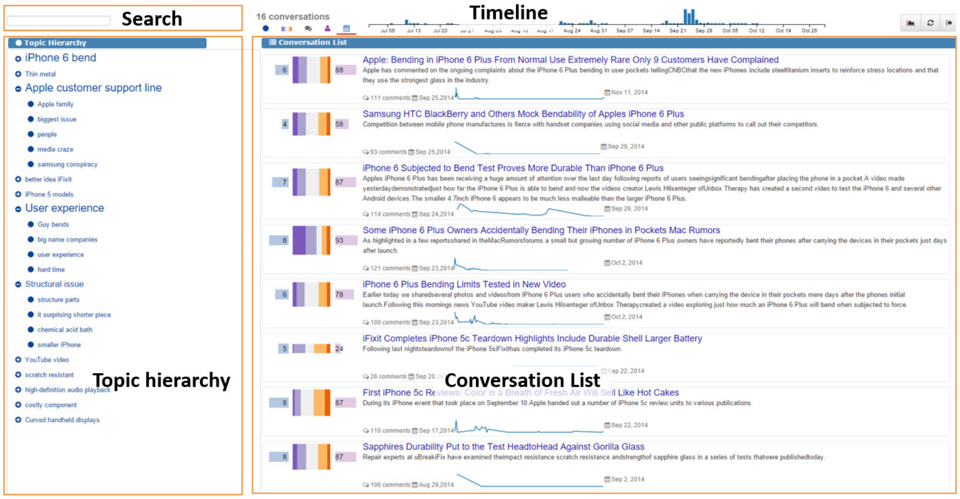
Figure 3.5: The main visual encodings in MultiConVis: a) Sentiment distribution is shown as stacked bar; b) Visual encoding of topics changes according to different user interactions; c) Visual encoding of a set of aggregated metadata and text analysis results for a conversation.

at all the three different levels of granularity[4]. 2) For all the attributes related to topics/authors facet, the same color coding was used across different levels (see Figure 3.5b).

All conversations: Initially, when the user starts exploring the whole collection of conversations MultiConVis displays three components as shown in Figure 3.6: 1) a Topic Hierarchy; 2) an overview of the set of conversations (Conversation List); and 3) a Timeline View showing the volume of comments of the whole collection over time. These three components are interactively coordinated, so that any operation in one view is reflected in the other views.

The **Conversation List** shows the current set of conversations, where each item in the list represents a set of aggregated metadata and the results of text analysis for the corresponding conversation (See Figure 3.5c). In particular, we encode the following attributes of each conversation: 1) the overall sentiment distribution us-

[4] The orange and purple colors were selected instead of the standard green and red to avoid the color blindness effects.



6 88 Apple: Bending in iPhone 6 Plus From Normal Use Extremely Rare Only 9 Customers Have Complained
Apple has commented on the ongoing complaints about the iPhone 6 Plus bending in user pockets tellingCNBCthat the new iPhones include steeltitanium inserts to reinforce stress locations and that they use the strongest glass in the industry.
111 comments Sep 25,2014 Nov 11, 2014

4 58 Samsung HTC BlackBerry and Others Mock Bendability of Apples iPhone 6 Plus
Competition between mobile phone manufactures is fierce with handset companies using social media and other public platforms to call out their competitors.
93 comments Sep 25,2014 Sep 29, 2014

7 87 iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus
Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.
114 comments Sep 24,2014 Sep 26, 2014

8 93 Some iPhone 6 Plus Owners Accidentally Bending Their iPhones in Pockets Mac Rumors
As highlighted in a few reportsshared in theMacRumorsforums a small but growing number of iPhone 6 Plus owners have reportedly bent their phones after carrying the devices in their pockets just days after launch.
121 comments Sep 23,2014 Oct 2, 2014

6 78 iPhone 6 Plus Bending Limits Tested in New Video
Earlier today we sharedseveral photos and videosfrom iPhone 6 Plus users who accidentally bent their iPhones when carrying the device in their pockets mere days after the phones initial launch.Following this mornings news YouTube video maker Lewis Hilsenteger ofUnbox Therapycreated a video exploring just how much an iPhone 6 Plus will bend when subjected to force.
100 comments Sep 23,2014 Oct 2, 2014

5 24 iFixit Completes iPhone 5c Teardown Highlights Include Durable Shell Larger Battery
Following last nightsteardownof the iPhone 5siFixithas completed its iPhone 5c teardown.
26 comments Sep 20,2014 Sep 22, 2014

8 87 First iPhone 5c Reviews: Color is a Breath of Fresh Air Will Sell Like Hot Cakes
During its iPhone event that took place on September 10 Apple handed out a number of iPhone 5c review units to various publications.
110 comments Sep 17,2014 Sep 22, 2014

8 87 Sapphires Durability Put to the Test HeadtoHead Against Gorilla Glass
Repair experts at uBreakiFix have examined theimpact resistance scratch resistance andstrengthof sapphire glass in a series of tests thatwere publishedtoday.
106 comments Aug 29,2014 Sep 2, 2014

Figure 3.6: A snapshot of MultiConVis for the ‘iPhone bending’ dataset: the Topic Hierarchy represents the set of topics and their sub-topics as an indented tree (left); the Conversation List shows a set of aggregated metadata and text analysis results for each conversation (row); the Timeline at the top shows the volume of comments over time for all conversations.

ing a stacked bar, 2) the number of comments, which is encoded as the height of this stacked bar, 3) the count of topics and authors as horizontal bars, and 4) a sparkline that represents the volume of comments over time in a more space efficient way [50]. In addition, the title and a text snippet of the conversation are shown to the right side of its visual summary. Overall, these attributes summarize the set of conversations, facilitating the discovery of interesting subsets of conversations that are of interest to the user.

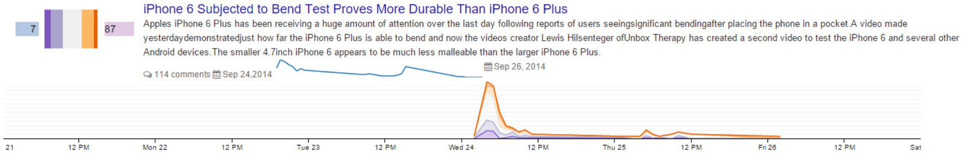
The **Topic Hierarchy** visually conveys all the topics in the whole collection of conversations using an indented tree representation. Here, topics are sorted chronologically within each level of the hierarchy. Each topic node is represented by its top keyphrase label returned by the topic modeling method, however, when the user hovers on a topic additional keyphrases are also shown to provide more context about that topic. The font size of a topic node represents how much it has been discussed compared to other topics. We present the Topic Hierarchy as an indented tree, where the parent-child relationship is represented by relative vertical position along with the horizontal position. We made this choice because an indented tree representation is much more compact than explicitly showing hierarchical links between topic nodes.

3.6.2 Multi-level Exploration

From the whole collection to subsets of conversations: While the user initially gets an overview of all the conversations in the collection, her subsequent goal is to find the subset of conversations that are more interesting or relevant, given her current information needs. We support this goal by providing a set of interactive features: linked highlighting, selection, filtering, and reordering. The Timeline View, shown in Figure 3.6, allows the user to quickly filter out conversations that do not fall within the time range in which the discussions were more active or relevant. In addition to filtering, the user can reorder the set of conversations based on the following attributes: number of topics/authors/comments, sentiment distribution, and date of the first post of a conversation.

To promote exploration based on the topic facet, we provide coordinated highlighting and selection of conversations by topic. For example, hovering on a topic highlights all the conversations where this topic was discussed, and conversely hovering on a conversation temporarily highlights topics in the Topic Hierarchy. Moreover, when the user selects a topic by clicking on it, a vertical outline is drawn alongside the related conversations, allowing the user to see the conversations in which this topic was discussed, even when she is exploring different conversations/topics. Throughout the filtering and selection processes, the representation of various attributes from both topics and conversations serve as information scent, thus enhancing the ability of the user to navigate and filter data more effectively [128].

Often, as the user finds a subset of conversations that are relevant to her information needs, she may become interested to know more detailed information about them, for instance, to see the temporal evolution of sentiment over time for each conversation. We provide such feature based on user interactions, i.e., as the user clicks on the ‘Show timeline’ button, the sentiment distribution of comments over time is represented as a stacked area chart, within each conversation item in the list (See Figure 3.7). This helps the user to understand temporal patterns of sentiment



iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus

Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.

114 comments Sep 24,2014

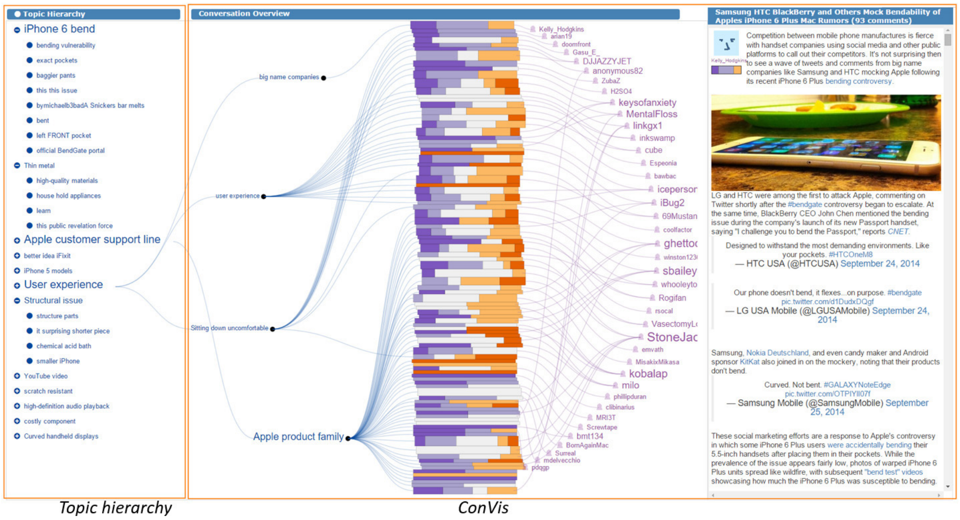
Sep 26, 2014

Figure 3.7: A conversation from the ‘iPhone bending’ dataset, showing stacked area chart to represent how sentiment distribution evolves over time.

in different conversations, supporting her to fulfill information needs related to the time facet.

Drill down to one conversation: As the user continues her exploration, she may become particularly interested in a specific conversation. In this case, she can drill down into that conversation with the ConVis interface, which was designed to explore a single conversation (described in Chapter 2) [59]. Here, an important design question arises: once the exploration has reached a single conversation, should we show ConVis along with both the Conversation List and the Topic Hierarchy so that the user can simultaneously glance at all of them? Notice that showing all the levels would be extremely challenging because of horizontal space limitations. However, we found this not even to be necessary. Our initial formative evaluations and case studies indicate that users do not need to jump back and forth to the Conversation List while exploring a single conversation. On the contrary, users tend to spend most of the time reading specific comments of the conversation they have decided to focus on before going back to the Conversation List. In light of this, when the user drills down into one conversation the Conversation List is replaced with the ConVis interface, as shown in Figure 3.8.

Now, we briefly describe how the visualization components of the ConVis interface interact with other views of MultiConVis (a more detailed description of ConVis is provided in Chapter 2). Recall that ConVis consists of an overview (Thread Overview) of the conversation along with two primary facets, topics and authors, which are presented circularly around this overview. Once ConVis is displayed within MultiConVis, the Topic Hierarchy over the whole collection is still shown to provide helpful context to the user in understanding the relationship between the topics of the selected conversation and the topics of the other conversa-



Samsung HTC BlackBerry and Others Mock Bendability of Apples iPhone 6 Plus Mac Rumors (93 comments)

Competition between mobile phone manufactures is fierce with handset companies using social media and other public platforms to call out their competitors. It's not surprising then to see a wave of tweets and comments from big name companies like Samsung and HTC mocking Apple following its recent iPhone 6 Plus bending controversy.

LG and HTC were among the first to attack Apple, commenting on Twitter shortly after the #bendgate controversy began to escalate. At the same time, BlackBerry CEO John Chen mentioned the bending issue during the company's launch of its new Passport handset, saying "I challenge you to bend the Passport," reports *CNET*.

> Designed to withstand the most demanding environments. Like your pockets. #HTCOneM8
> — HTC USA (@HTCUSA) September 24, 2014

> Our phone doesn't bend, it flexes...on purpose. #bendgate pic.twitter.com/d1DudxDQgf
> — LG USA Mobile (@LGUSAMobile) September 24, 2014

Samsung, Nokia Deutschland, and even candy maker and Android sponsor KitKat also joined in on the mockery, noting that their products don't bend.

> Curved. Not bent. #GALAXYNoteEdge pic.twitter.com/OTPIYIl07f
> — Samsung Mobile (@SamsungMobile) September 25, 2014

These social marketing efforts are a response to Apple's controversy in which some iPhone 6 Plus users were accidentally bending their 5.5-inch handsets after placing them in their pockets. While the prevalence of the issue appears fairly low, photos of warped iPhone 6 Plus units spread like wildfire, with subsequent "bend test" videos showcasing how much the iPhone 6 Plus was susceptible to bending.

Figure 3.8: As the user selects a particular conversation, the Conversation List is replaced by the ConVis interface, where the Thread Overview visually represents the whole conversation encoding the thread structure and how the sentiment is expressed for each comment(middle); The Facet Overview presents topics and authors circularly around the Thread Overview; and the Detail View presents the actual conversation in a scrollable list (right). Here, topics a are connected to their related comments as well as to their parents in the Topic Hierarchy via curved links.

tions. As shown in Figure 3.8, the topics of the selected conversation displayed with ConVis are explicitly linked to the ones in the Topic Hierarchy.

The user can explore a conversation using the interactive features of ConVis, such as hovering and selecting a topic of interest. While exploring a topic, she might become interested to know whether similar topics are discussed in other conversations. At any point, the user can look at the Topic Hierarchy to see what are the other similar topics to her current topic of interest, but not discussed in this conversation. For instance, when the user is exploring the topic ‘Thin metal’ in the current conversation, she may select a related topic labeled ‘Structural issue’ in the Topic Hierarchy, which results in abandoning the ConVis interface and switching back to the Conversation List, where the conversations related to ‘Structural issue’ would be highlighted. Finally, at any time the user can return to the Conversation

List by clicking on the ‘Back’ button.

3.7 Implementation

The data acquisition, preprocessing, and analysis components were developed using python and a server-side component (in PHP) which feeds the data to the visualization pipeline. The visual interface was implemented using a combination of HTML and JavaScript (using the D3, JQuery, crossfiter, and dc.js libraries).

3.8 Evaluation

We evaluated the MultiConVis interface in two different ways: 1) case studies with different domain experts, 2) a formal user study with regular blog readers. While the case studies provided qualitative evidence for the utility of the MultiConVis system, the user study allowed us to compare the system with a traditional interface. Note that ConVis, the interface for single conversations embedded in MultiConVis, had already been evaluated (described in Section 2.6 and Section 4.6.1), which showed that ConVis outperformed traditional interfaces along several subjective metrics (e.g., usefulness, enjoyable).

3.8.1 Case Studies

We conducted case studies with three users, whose professions are quite diverse, but who come from populations that could all arguably benefit from MultiConVis:

U1: a regular blog reader who visits the Macrumors blog site several times a week. Therefore, he was interested in exploring the conversations returned by our ‘iPhone bending’ query. His primary goal was to verify whether the problem of ‘iPhone bending’ reported by some customers was really serious or not.

U2: a graduate student in the school of Journalism, who contributes to local newspapers about recent political issues. He had strong interest in our dataset about the recent ‘ObamaCare health reform’. His primary goal was to understand and summarize the key opinions expressed by the participants about the ObamaCare health reform.

U3: a business analyst in a social media company, where she often needs to analyze a large amount of conversations to understand how customers react to newly released products. So, her goal in the study was to explore conversations about the 'iWatch release' to identify comments that express negative opinions about the product, which is a task that matches what she performs on a regular basis for her company.

For the purpose of case studies, we have collected three different datasets from two different blog sources: Macrumors [1] (a technology news related blog site dedicated to the discussion of recent news and opinion relating to the Apple Inc) and Daily Kos [3] (a political analysis blog site) between September to December 2014. To create each dataset, we provide a query to the blog site to retrieve the set of conversations that appear on the first page of the search results.

For each case study, we analyzed the results by triangulating between multiple data collection methods, including observations, notes taken by participants during the analysis session, and semi-structured interviews. In addition, we logged interface actions to better understand the usage patterns.

We now report the primary results of the case studies. The key findings were that: (a) all three users relied on the topic hierarchy to accomplish their task, (b) each user used the hierarchy differently, (c) all users found the topic hierarchy extremely useful. For instance, while the blog reader started his exploration by quickly scanning through the topics in the hierarchy and then going back and forth between topics and conversations, the journalist explored the topics in the hierarchy more systematically, exploring all the comments about one topic before moving to a new one. Still differently, the business analyst started by skimming through the titles of the conversations. But, as she was skimming through the conversations, she also kept an eye on the topics that were highlighted for each conversation in the topic hierarchy. In this way, she identified controversial topics that were intensely debated in recent conversations.

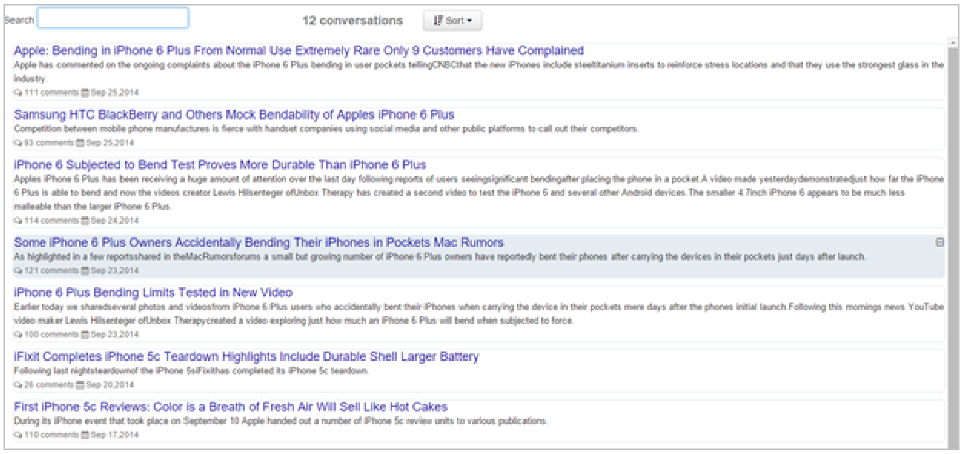
Overall, the semi-structured interviews revealed that users were very satisfied with the interface. In particular, U1 said *"The comments about that chemical acid bath was buried down in the middle of one conversation, which I don't think I would have noticed with a regular interface. Using MultiConVis, I was able to pick this topic from the hierarchy and then jumped into the related comments without having to read the entire conversations...."*. U2 found the topic hierarchy to be very helpful in supporting a systematic exploration of the conversations by organizing the key opinions into meaningful topical groups. More interestingly, he realized the potential utility of MultiConVis system for other exploratory tasks that he would like to perform, *"This tool could be not only useful when I want to write a story, but also to prepare for interviewing a policy maker, or a politician by quickly understanding what topics are triggering the most interesting or controversial discussions in the public spheres."* Finally, U3 anticipated that this tool could be very useful to understand what features of their products worked (or didn't work) and then revise the products accordingly, *"The MultiConVis interface would definitely help me to understand the requirements and needs of my customers more effectively. Our current way is just to skim through the comments, often missing the important feedback from customers ...but this interface can help me identify what are the biggest concerns from the customers and get clues about the ways to satisfy their needs."*

3.8.2 User Study

We ran a formal user study to evaluate the effectiveness and usability of the MultiConVis interface compared to an interface that represents the traditional interfaces for blog reading. The aim of the user study was to answer the following two questions: (1) When we compare MultiConVis with the traditional interface for exploring a set of conversations, is there any difference in user performance and subjective reactions? (2) What specific features of the MultiConVis interface are perceived as more/less beneficial by the potential users (e.g., Topic Hierarchy, Timeline)?

Methodology

Since the first research question requires comparisons among two different user interfaces, we conducted a summative evaluation through controlled experiments [78]. The study was designed with two interfaces as conditions: a) the traditional interface for blog reading, and b) *MultiConVis*. Here, the traditional interface shows a set of blog conversations as a linear list, where each item represents a set of metadata of the conversations, such as title, the number of comments, and



Apple: Bending in iPhone 6 Plus From Normal Use Extremely Rare Only 9 Customers Have Complained
Apple has commented on the ongoing complaints about the iPhone 6 Plus bending in user pockets tellingCNBCthat the new iPhones include steeltitanium inserts to reinforce stress locations and that they use the strongest glass in the industry.
111 comments Sep 25,2014

Samsung HTC BlackBerry and Others Mock Bendability of Apples iPhone 6 Plus
Competition between mobile phone manufactures is fierce with handset companies using social media and other public platforms to call out their competitors.
93 comments Sep 25,2014

iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus
Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.
114 comments Sep 24,2014

Some iPhone 6 Plus Owners Accidentally Bending Their iPhones in Pockets Mac Rumors
As highlighted in a few reportsshared in theMacRumorsforums a small but growing number of iPhone 6 Plus owners have reportedly bent their phones after carrying the devices in their pockets just days after launch.
121 comments Sep 23,2014

iPhone 6 Plus Bending Limits Tested in New Video
Earlier today we sharedseveral photos and videosfrom iPhone 6 Plus users who accidentally bent their iPhones when carrying the device in their pockets mere days after the phones initial launch.Following this mornings news YouTube video maker Lewis Hilsenteger ofUnbox Therapycreated a video exploring just how much an iPhone 6 Plus will bend when subjected to force.
100 comments Sep 23,2014

iFixit Completes iPhone 5c Teardown Highlights Include Durable Shell Larger Battery
Following last nightsteardownof the iPhone 5siFixithas completed its iPhone 5c teardown.
26 comments Sep 20,2014

First iPhone 5c Reviews: Color is a Breath of Fresh Air Will Sell Like Hot Cakes
During its iPhone event that took place on September 10 Apple handed out a number of iPhone 5c review units to various publications.
110 comments Sep 17,2014

Figure 3.9: The baseline interface, which initially presents the collection of conversations as a linear list, showing a set of metadata for each conversation in the list.

posting date (see Figure 3.9). The user can click on any conversation in the list, which results in showing all the comments of that conversation using an indented list representation. In addition, we provided a set of interactions that are common in most blog reading interfaces, i.e., searching for terms and sorting conversations by attributes (e.g., number of comments). A within-subject design was used with interface as the within-subject factor, allowing us to directly compare the measures of each participant with respect to both interfaces. Finally, all study aspects, including instructions and setup, went through several iterations of evaluation and pilot testing.

Procedure and task

At first, a pre-study questionnaire was administered to capture demographic information and prior experience with exploring blog conversations. Then, the participant went through the following steps for each of the two interfaces: 1) In a scripted warm-up session, the interface was introduced to the participant using a sample dataset. 2) The participant was then asked to perform a task based on a set of conversations. For each interface, a different set of conversations was provided.

Task: Considering the open-ended nature of blog reading, no specific set of questions was given. Instead, the participant was asked to explore a set of conversations about the given query and then write a single summary of what she thought were the major discussion points and most insightful comments within the conversations. The study lasted approximately 60 minutes and each participant was paid \$15 to participate.

We selected two different datasets crawled from the Macrumors site for testing ('iPhone bend' and 'iPad release'). The number of conversations in the datasets are kept the same (16 conversations in each dataset) to avoid potential variations due to the amount of conversational data. Also, to counterbalance any potential learning effects due to the order of exposure to specific interfaces and dataset, the order was varied using a 2 x 2 Latin square. During the study, we collected both quantitative data such as task completion time and qualitative data such as observations and questionnaires. Finally, a post-study questionnaire followed by a semi-structured interview were administered regarding the user's experience with two interfaces[5].

Participants

We conducted the study with 16 users (aged 18-37, 6 females) who have considerable experience with reading blogs. The participants held a variety of occupations ranging from journalists, engineers, system analysts and students from both graduate and undergraduate levels. They were recruited through emails and social networks (Facebook and Reddit posts).

Results Analysis

After completing the task with each interface, participants rated six different measures in the form of in-study questionnaires. Since these measures were rated using a standard 5 point Likert scale, standard parametric analysis was not suitable due to the lack of normality [71]. Instead we performed nonparametric analysis i.e., Mann-Whitney's U tests on the responses for each of these measures.

The results of these questionnaires are presented in Figure 5.6. The pairwise comparisons using Mann-Whitney's U tests indicate that MultiConVis was superior on five different measures out of six: usefulness ($Z = -1.823; p < .05$); enjoy-

[5] The study materials for the user study can be found in Appendix A

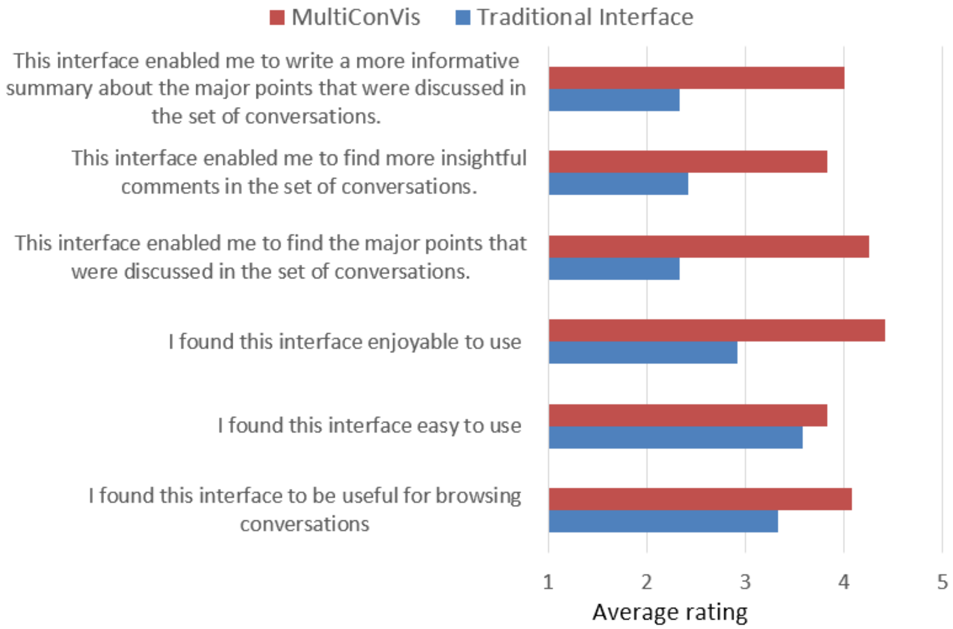


Figure 3.10: Average rating of interfaces by the participants on six different measures. Longer bars indicate better rating.

able to use ($Z = -3.697; p < .01$); find insightful comments ($Z = -3.95; p < .01$); find major points ($Z = -2.909; p < .01$); and enable to write more informative summary ($Z = -3.915; p < .01$). For the other measure i.e., ease of use, MultiConVis was still rated higher over the traditional interface, however the results was not significant. This is interesting, because MultiConVis appears to be as easy to use as the other interface in spite of its complex interface features.

Interface features: Each participant was also asked a set of questions regarding the usefulness of specific features of the MultiConVis interfaces. From Figure 3.11, we can readily see that the majority of the responses were dominated by positive ratings. Among the interface features, the Topic Hierarchy received the most positive ratings (strongly agree:9, agree:6), followed by the visual summary of each conversation, and interactive filtering by timeline.

Time: The average time required to complete the tasks was not significantly affected by the interfaces, with MultiConVis and the traditional interface requiring 1065 ± 249 and 1029 ± 204 secs respectively.

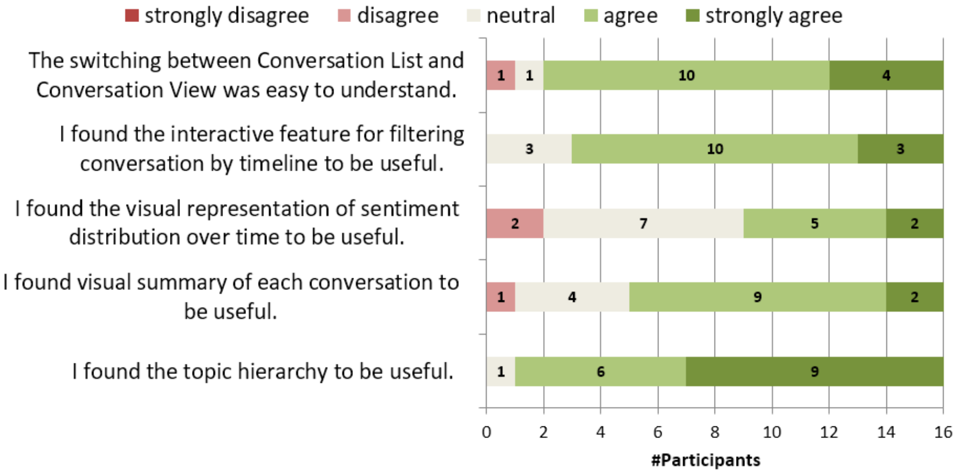


Figure 3.11: Responses to statements regarding specific features of the MultiConVis interface.

Overall Preference: In the post-study questionnaire, participants were asked which system they prefer for exploring a collection of conversations. 75% of the participants indicated a preference for MultiConVis, whereas 25% preferred the traditional interface. Many of the participants who chose MultiConVis indicated that the utility of Topic Hierarchy was the primary reason for their preference: *"By having a topic hierarchy of the relevant topics, as well as highlighting which conversation refers to which topic, it was very easy to filter out the blogs that were not relevant.'* (P8). They also found the visual summary provided for each conversation was very useful, *"The summary offered by this visualization is quite impressive and throws a lot of instant information."* (P2). Additionally, for the sentiment distribution over time *"...made it very easy to see how opinions changed over time. While investigating bend gate it was clear how the community opinion changed after the event had played out in the media"* (P4).

Those who preferred the traditional interface indicated that they like its familiarity *"I preferred the older style of interface mainly because it's what I'm more familiar with..."* (P1). They also pointed out that sometimes the topic hierarchy was inaccurate, for instance, topic labels did not always make sense to them: *"..maybe with better tagging I'd find it (MultiConVis) more useful..."* (P1), and *"the key-*

words weren't necessarily the most useful ones or the relevant ones" (P5). We have considered these comments to improve our approach in Chapter 4, by introducing a human-in-the-loop topic model.

3.9 Discussion

We now discuss the summary of findings from the user study, as well as the limitations of this type of user study.

3.9.1 Summary of Findings

Our case studies demonstrate that the system can be useful in a variety of contexts of use, while the formal user study provides evidence that the MultiConVis interface supports the user's tasks more effectively compared to traditional interfaces. In particular, all our participants, both in the case studies and in the user study, appear to greatly benefit from the topic hierarchy and the high-level overview of the conversations. The user study also shows that the MultiConVis interface is significantly more useful than the traditional interface, enabling the user to find insightful comments from thousands of comments, even when they were scattered around multiple conversations, often buried down near the end of the threads. More importantly, MultiConVis was preferred by the majority of the participants over the traditional interface, suggesting the potential value of our approach for combining NLP and InfoVis.

3.9.2 Evaluation Methodology

In this work, we conducted a lab-based user study to understand the potential effectiveness of the MultiConVis interface. Even though a controlled study is suitable for comparing different interfaces, it may not accurately capture real-world scenarios [78]. Although we carefully recruited participants who were frequent blog readers, still different settings were controlled to make a fair comparison among interfaces (e.g., they were not allowed to choose a conversation according to their own interest).

In order to enhance the ecological validity of our evaluations [18], a possible approach would be to perform Web-based studies to observe how the system is used by real users to satisfy their information needs. In Chapter 5, we will describe how we ran a user study in Web-based environments, where participants worked in their own settings performing their own task. This study was conducted among hundreds of users who performed information seeking tasks by exploring a set of conversations in a community question answering forum. This study also gives us the advantage of collecting interaction logs from a large number of users to get deeper insights that are arguably more generalizable than a lab study.

3.10 Summary

MultiConVis is an interactive visual text analytics system for exploring a collection of blog conversations. Unlike traditional systems, MultiConVis takes the unique characteristics of online conversations into account to tightly integrate NLP and InfoVis techniques. The resulting visual interface aggregates data across different levels, supporting a faceted exploration starting from a whole set of conversations, to a subset of conversations, to one conversation.

While the topic hierarchy was found to be very useful, still in a few cases the extracted topics were either noisy or did not match the user's current information needs. To deal with this problem, we have devised an interactive topic hierarchy revision approach, where the user can provide feedback to the system so that the revised topic hierarchy better matches her tasks and mental model. In the next chapter, we will discuss this interactive topic revision approach in details.

Chapter 4

Interactive Topic Modeling for Exploring Online Conversations

In Chapter 2 and 3, we presented two visual text analytics systems for exploring online conversations. In both systems, we have applied topic modeling techniques to summarize the primary themes discussed in a conversation (or a set of conversations). However, from the evaluations with real users, we found that the results of the topic model were sometimes noisy, or even if accurate did not match their current information needs.

To address this problem, in this chapter we propose novel topic modeling methods for asynchronous conversations that revise the model on the fly on the basis of users' feedback. We then integrate these methods within our visual interfaces (i.e., ConVis and MultiConVis) to create two new interfaces ConVisIT and MultiConVisIT, where IT stands for Interactive Topic modeling. The goal of incorporating the user' feedback within the visual interface is to support the user in exploring conversations, as well as in revising the topic model when the current results are not adequate to fulfill the user's information needs. Finally, we discuss two lab-based studies with real users that compared ConVisIT and MultiConVisIT with interfaces that do not support human-in-the-loop topic modeling[1].

[1] Portions of this work were published in *ConVisIT: Interactive Topic Modeling for Exploring Asynchronous Online Conversations*; Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI), pp. 169-180, 2016 [60]. An extended version of this paper has also been

4.1 Introduction

While topic models can provide an attractive solution to understanding large conversations, they may not always be useful to the end users [17, 59, 66]. This could be due to different reasons. For instance, the current information seeking tasks may require a topic model at a different level of granularity, e.g., if the user needs more specific information about 'ObamaCare' she might be interested in exploring its potential sub-topics such as 'health insurance', 'healthcare cost', and 'drugs'. Also, the interpretation of topics may vary among users according to their expertise and mental model. In a topic annotation study of blog conversations, human annotators sometimes disagreed on the number of topics and on the assignment of sentences to topic clusters [70]. For instance, for one of the conversations from their corpora, one annotator produced 22 topics, while another annotator reported only 8 topics. Furthermore, the results of automatic topic modeling can be simply incorrect, in the sense that the generated topics would not make sense to any user [22, 70]. For example, two semantically different topics 'Obama health policy' and 'job recession' might be wrongly grouped together with the misleading topic 'Obama recession'.

Similarly, when we organize topics into a hierarchy, the resulting organization may not be always useful to the users. For example, when a new product is launched, a business analyst may want to organize topics based on people's opinions about the 'sales' and 'customer service', whereas a potential buyer may want to categorize topics based on 'new features' of the product. In other cases, the topic organization might not be accurate i.e., two semantically different topics might be wrongly placed under the same parent topic. For example, a parent topic named 'iPhone bending vulnerability' may have two children'iPhone 6' and 'longer battery life', which might be completely unrelated in the context of the discussion.

In this chapter, we present an interactive topic modeling framework, that can support the user in exploring conversations by relying on topics that make sense to her, that are semantically coherent and match her expertise, mental model, and current task. In our framework the user can revise the topic model, while she is

appeared as a jounral paper: *Interactive Topic Modeling for Exploring Asynchronous Online Conversations: Design and Evaluation of ConVisIT*, by Enamul Hoque and Giuseppe Carenini; ACM Transactions on Interactive Intelligent Systems (TiiS),6(1):7:17:24, Feb. 2016 [61].

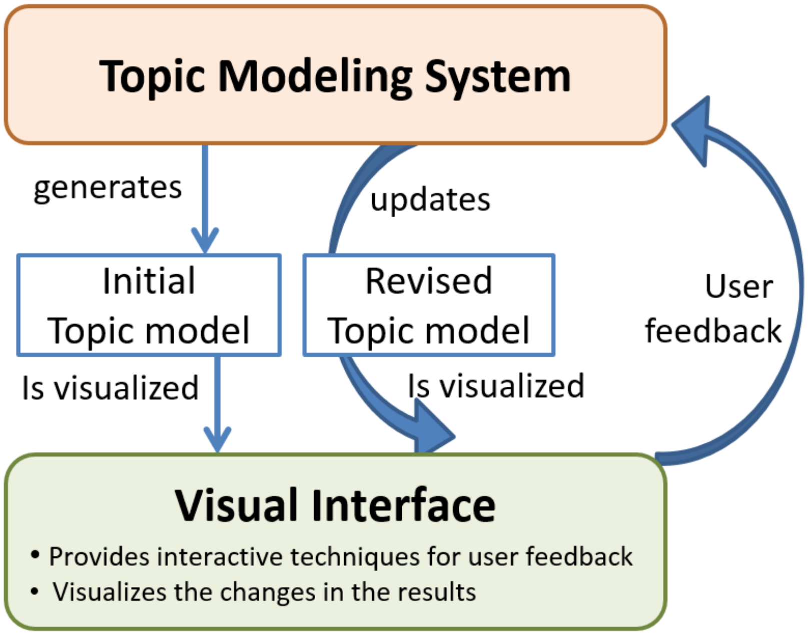


Figure 4.1: Interactive topic modeling framework for exploring asynchronous conversation.

exploring conversations. To achieve this, user feedback is incorporated within the topic modeling loop in real-time through the visual interfaces. In particular, we incorporated user feedback by extending the ConVis and MultiConVis interfaces. The re-designed interfaces support the user in revising both a list of topics as well as a topic hierarchy.

Figure 4.1 illustrates our interactive topic modeling framework. Given the asynchronous conversation(s), the system generates an initial topic model (a linear list of topics, or a set of topics organized into a hierarchy), which are presented in the visual interface along with other conversational data. The interface then supports the user in exploring the conversation. However, whenever the user realizes that the current topic model is not helping her, she can provide topic revision feedback to the system through interactions. Subsequently, the system updates the topic model accordingly and the new results are shown in the interface.

The primary contributions of our work are three-fold:

1) A novel interactive topic modeling approach specifically devised for asyn-

chronous conversations. Existing systems (e.g., [19, 66, 81]) were mainly devised for generic documents without considering the unique features of conversations. In contrast, we analyze the information seeking tasks in our target domain to select a minimum set of topic revision operations that are critical to the user. Then, we devise computational methods for each of these operations to be performed by the system.

2) We designed a set of interactive features that allow the user to revise the current topic model. In response, the interface updates and re-organizes the modified topics by means of intuitive animations, so that the user can better fulfill her information needs.

3) We conducted two lab-based summative studies to assess how user performance and experience change when a human-in-the-loop topic modeling approach is introduced in our visual interfaces for exploring both a single conversation and a collection of conversations.

The remainder of this chapter is organized as follows. First, we provide an overview of related research on interactive topic modeling. Next, we present our approach for interactive topic modeling and the specific interactive visualization features by which the user can revise the topic model. This is followed by a description of the user study along with a detailed analysis of the results. Finally, we discuss the overall findings and outline directions for future work.

4.2 Related Work

Several approaches and tools have been proposed in the literature for incorporating human feedback within the topic modeling generation process, which we discuss below.

4.2.1 Human-in-the-loop Topic Model

Since system-generated topic models can be noisy and and/or may not match users' current information needs, some recent works have investigated how user supervision can be introduced to improve the results. The main focus has been on answering the following two research questions: 1) How to revise the topic model given the user feedback? 2) How to best support the user in expressing such feedback within a visual interface?

To answer the first question– in the dominant LDA topic modeling framework, the original unsupervised LDA method was modified to allow for the introduction of human supervision [11, 66, 105]. For instance, Andrzejewski et al. incorporates user's domain knowledge in LDA by adding constraints in the form of must-link (enforces that sets of words must appear together in the same topic) and cannot-link (enforces that sets of words must be in different topics) using Dirichlet forest prior [11]. However, this method requires to rerun Gibbs sampling from scratch after a set of constraints is added, leading to high latency. Since such latency is undesirable for real-time interactions, Hu et al. propose a more efficient inference mechanism that aims to minimize user's waiting time [66]. More recently, another variant of LDA was proposed that also incorporated must-link and cannot-link constraints [130], however these constraints are applied at the document level instead of at the word level. The purpose of applying such constraints is to improve the topic model stability, by minimizing the changes to the topic assignments of old documents when the model is updated to take new documents into account.

Unfortunately, all these approaches were designed for non-conversational documents. In contrast, an asynchronous conversation has some unique features such as participants often reply to a comment or they quote a portion of comments, creating a conversational structure. It has been shown that by utilizing these unique features, the accuracy of the topic model on conversations can be improved over traditional document-centric topic model [70]. Therefore, we devise a new interactive topic modeling framework that is designed to take advantage of conversational features.

Previous work has addressed the question of how a visual interface can support the user in revising text analytic models, while exploring a set of documents [19, 23, 81]. In their seminal work, Pirolli et al. presented the *Scatter/Gather* system, where the user could select a document cluster and then ask the system to re-cluster it to analyze its sub-clusters [102]. More recently, Chuang et al. extend Termite [20], which visualizes the term-topic distributions produced by LDA, and allows the user to revise the model by clicking on words to promote or demote their inclusion/prominence in a topic [23]. Similarly, Lee et al. visualize topic modeling results from LDA, and allow the user to interactively manipulate the topical keyword weights and to merge/split topic clusters [81]. Even more recently, user feedback was incorporated through a scatter plot visualization, that steers a semi-supervised non-negative matrix factorization (NMF) method [100] for topic modeling [19]. The authors show that the NMF-based approach has faster empirical convergence and offers more consistency in the results over traditional LDA-based approaches. They visually present each topic cluster and then allow the user to directly manipulate the documents and keywords within each cluster to specify topic revisions. A fundamental limitation of most of these works is that the visual interfaces for interactive topic model were not evaluated with real users. Therefore, a set of critical research questions remained unanswered. For instance, would users be really interested in performing all the operations provided with such a complex interactive visualizations? What operations are actually useful to the users for performing exploratory tasks in a specific domain? To answer these questions, we applied a systematic design approach, where we first identified a set of topic revision operations which are most useful according to our tasks analysis, and then performed a user study to measure the utility of these operations.

4.2.2 Interactive Topic Hierarchy Revision

There have been some earlier works to design systems for revising hierarchical structures such as taxonomies and ontologies [10, 99]. For instance, ReTAX is a taxonomy revision system which takes a pre-established taxonomy as input and some new items, and then uses a set of consistency rules to determine the inconsistency in the hierarchy and generate refinement to the hierarchy to resolve the inconsistency. More recently, Nikitina et al. presents a technique for ontology revision, where the system presents a set of propositions, for example, *a* is a subclass of *b*, to the user, and then based on user's feedback revises the underlying ontology accordingly [99].

Unlike revising taxonomies or ontologies, the revision of a topic hierarchy has rarely been studied. One notable exception is the the work from Dou et al. [40], which allows the user to modify a hierarchical topic structure from a visual interface. However, such modification does not invoke the underlying topic modeling system; instead, the revised topic structure simply becomes visible in the interface.

Moreover, even though they considered split operation as critical for improving the current topic model, this operation was not supported in their work.

4.3 Interactive Topic Modeling System

As illustrated in Figure 4.1, our interactive topic modeling system performs two primary functions: 1) generating the initial topic model, 2) revising the topic model based on user feedback. We have already discussed how to generate the initial topic model in Chapter 2 and 3. We now discuss in detail how the system revises the topic model generated from a single conversation, as well as the topic model generated from a collection of conversations.

4.3.1 Interactive Topic Revisions of Topic Models for a Single Conversation

Although the initial topic model generated by our approach has been found to be more accurate than models generated by traditional methods for non-conversational text [70], still the extracted topics may not always match the user's information needs. Depending on the user's mental model and current tasks, the topic modeling results may not be adequate. For instance, more specific topics may be more useful in some cases, while more generic ones in other cases. Therefore, we incorporate a set of topic revision operations by which users can iteratively modify the initial topic model to better fulfill their information needs.

Since it may take some effort from the users to express different topic revision operations, it is important to identify the minimal set of operations that would be both intuitive and sufficient to support user's tasks [23]. For this purpose, we first considered eleven different possible topic revision operations listed in Table 5.1 based on reviewing existing work on interactive topic modeling [11, 19, 66, 81]. Next, we prioritized the operations based on the following criteria ordered by their importance: 1) *Task relevancy:* To what extent this operation is relevant to the tasks involved in exploring a conversation as identified in [59]? 2) *Topic model relevancy:* Is this operation applicable to our topic model approach? 3) *Redundancy:* Is this operation already covered by other operations, which are stronger on the previous two criteria?

No	Operation	Why?	Criteria			Reference
			Task relevancy	Topic Model relevancy	Redun-dancy	
1	Split a topic	This topic is too generic	high	yes	no	[19, 81]
2	Merge by joining	These topics are talking about similar things	high	yes	no	[19, 81]
3	Merge by absorption	A sub-topic is more related to a different topic than its current parent topic	high	yes	no	[19]
4	Split by keyword	This keyword should be separated into a new topic	medium	yes	yes	[19]
5	Change the overall granularity level of topics	Too few topics/ too many specific topics are generated	medium	yes	yes	[81]
6	Remove the topic from the display	This topic does not make any sense (i.e., off-topic)	low	yes	yes	[81]
7	Assign a label for this topic	The current label of this topic does not represent the actual topic	low	yes	yes	[43]
8	Increase the weight of this keyphrase	This keyphrase should be included in the topic label list	low	yes	yes	[43]
9	Apply must-link constraint	Those words must be in the same topic	low	no	no	[11, 66]
10	Apply cannot-link constraint	Those words must not be in the same topic	low	no	no	[11, 66]
11	Change keyword weights	This keyword is more related to the topic	low	no	yes	[19, 81]

Table 4.1: Different possible topic revision operations.

The three operations at the bottom of Table 5.1 (9-11) are eliminated based on both task and topic model relevancy criteria. Not only are these operations designed to fix the term-topic distribution, which is not applicable to our topic modeling approach; but more importantly, they are arguably not very useful to support the high-level exploratory reading tasks as identified in 2 and therefore the users may not be motivated to perform such operations. In the end, we selected the top three operations in Table 5.1 namely, 'split a topic', 'merge topics by join', and 'merge topics by absorption', because we identified them as the most relevant to our exploratory reading tasks in which the user may benefit from dynamically changing the granularity level of different topics. Also, by selecting them some other candidate operations with lower task relevancy become redundant and are therefore eliminated. These are 'change the overall granularity level of topics' (covered by topic splitting and merging) and 'split by keyword' (covered by topic splitting).

Figure 4.2 illustrates the three selected topic revision operations. In the remainder of the section, we describe how each of these operations supports the user's tasks, and how the underlying topic model is revised according to these operations.

Split a topic

Topic splitting allows the user to explore more specific sub-topics of a given topic, thus changing the topic granularity to a finer level. Consider an example, where initially the system creates a topic named 'military security'. As the user starts exploring this topic, she finds it to be too generic with respect to her information needs and therefore she wants to split it into more specific sub-topics.

Method: Assume that the user wants to split a topic A into multiple sub-topics, as shown in Figure 4.2. Upon user's request, the underlying topic model creates a sub-graph $G_A(V_A, E_A) \subset G(V, E)$ from the original graph $G(V, E)$ generated in the initial topic segmentation (see Section 2.4.1 for details), where V_A represents the vertices (sentences) of topic A, and each edge $w(x, y)$ in E_A represents the weighted edges of topic A.

Next, the system splits the chosen topical cluster A into further n sub-clusters $A_1, A_2, ..., A_n$, by applying the same graph partitioning algorithm used in the initial

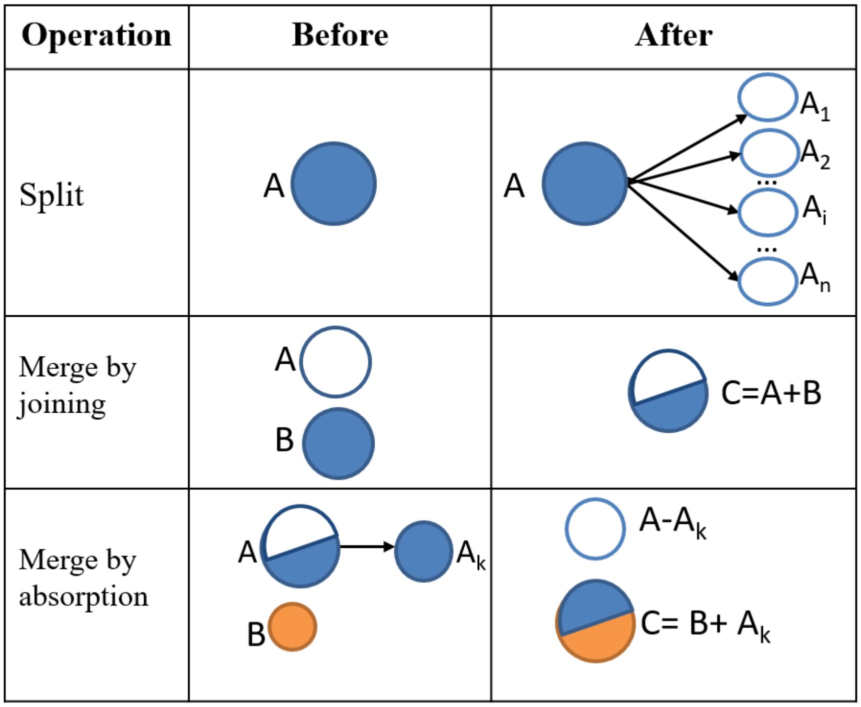


Figure 4.2: Three different user operations for topic revision

topic segmentation phase, i.e., approximate solution to n-Cut [118] on $G_A(V_A, E_A)$. Here, n is the optimal number of sub-topics, which is automatically determined by finding the value of n for which an objective function Q is maximized according to the formula proposed by Newman and Girvan [97],

$$Q_n(A) = \sum_{c=1}^{n} \frac{\sum_{x\in V_c, y\in V_c} w(x,y)}{\sum_{x\in V_A, y\in V_A} w(x,y)} - \left(\frac{\sum_{x\in V_c, y\in V_A} w(x,y)}{\sum_{x\in V_A, y\in V_A} w(x,y)}\right)^2. \tag{4.1}$$

$Q_n(A)$ measures the quality of a clustering of nodes in the graph $G_A(V_A, E_A)$ into n groups, where $\sum_{x\in V_c, y\in V_c} w(x,y)$ measures the within-cluster sum of weights, $\sum_{x\in V_A, y\in V_A} w(x,y)$ measures the sum of all edge weights in the graph, and $\sum_{x\in V_c, y\in V_A} w(x,y)$ measures the sum of weights over all edges attached to nodes in cluster c. In essence, according to Equation 4.1, the nodes in high-quality clusters should have much stronger connections among themselves than with other nodes in the graph.

We apply Equation 4.1 for increasing value of $n = 2, 3, 4, 5$ and select the value of n, for which $Q_n(A)$ is maximum. The highest possible value of n is capped to 5 because of time constraint imposed by the interactive nature of the operation. Notice, however, that this limitation is not too penalizing. Our analysis of the Slashdot corpus shows that in 86% cases of splitting a topic, the best value of $Q_n(A)$ is with $n \leq 5$ and in the cases for which this is not the case the improvement for $n > 5$ is minimal.

Once the parent topic is segmented into n different sub-clusters, representative keyphrases are generated for each sub-topic. This is done by running our topic labeling method, as described in Section 2.4.1, only on the sub-conversation covered by A.

Merge by joining

This operation allows the user to aggregate multiple similar topics into a single one. Opposite to topic splitting, the result is a topic with coarser granularity. Consider an example, where the initial topic model produces two different topics namely 'secure code' and 'simple sql server injection'. The user may find that both topics are too specific, therefore joining them into a more generic topic may help her to better perform subsequent tasks.

Method: Assume that the user decides to merge by joining two topics A and B (see Figure 4.2). To perform this operation, the topic modeling system creates another topic C and assigns its vertices as $V_C = V_A \bigcup V_B$ and edges as $E_C = E_A \bigcup E_B$. After that, a label for C is generated. This is done by running our topic labeling method, as described in Section 2.4.1, only on the sub-conversation covered by C.

Merge by absorption

If a sub-topic is more related to a different topic than its current parent topic, merge by absorption allows the user to separate this sub-topic from its current parent and merge it with the one to which it is more related. Unlike the previous merge operation (which joins two independent topics), this operation allows a sub-topic that is already placed under a topic to be absorbed by a different parent topic. Consider an example, where the sentences related to two different topics, namely

‘Obama health policy’ and ‘job recession’ are wrongly grouped together under the topic ‘Obama recession’. The user may realize that the sub-topic ‘job recession’ should be separated from its parent topic and merged with the ‘unemployment’ topic to which it is more related.

Method: Upon receiving a merge by absorption feedback from the user on A_k and B, the topic modeling system removes the sub-topic A_k from its current parent A and merge it with the topic B (see Figure 4.2). The system then creates a new parent topic C and then assigns vertices such that,

$$V_C = V_{A_k} \bigcup V_B, V_A = V_A \setminus V_{A_k}. \tag{4.2}$$

and edges such that,

$$E_C = E_{A_k} \bigcup E_B, E_A = E_A \setminus E_{A_k}. \tag{4.3}$$

Next, the topic labeling method takes the portion of the conversation that consists of the sentences in V_C, thus generating a label for C that potentially represents descriptive keyphrases from both topics A_k and B.

4.3.2 Interactive Topic Revisions of Topic Models for a Set of Conversations

In Chapter 3, we have shown that to support the exploration of multiple conversations it can be extremely useful to organize topics into a hierarchical structure. However, similarly to what we found out for simpler topic models for single conversations, the initial model generated by our system may not work well for the current user, because the resulting hierarchy is noisy and/or does not conform with the users mental model and current tasks.

To address these problems, we facilitate the user in modifying the topic hierarchy through a set of topic revision operations similarly to what was done for revising topic models for a single conversation. For this purpose, we analyze the feedback from the user study (as described in the previous chapter), as well as formative studies, to devise a set of operations for revising the topic hierarchy that is potentially useful. The list of operations is shown in Table 4.2.

The first two operations in Table 4.2 are intended to change the number of sub-topics of a parent topic depending on users information needs. The next two operations help the user to move a children topic from its current parent topic and place it under a more appropriate parent topic. The last two operations help the user to rename a topic node or simply remove it. While exploring a set of conversations, the user may apply a combination of different operations to organize topics into a hierarchy that is more accurate and matches her mental model and current tasks.

No	Operation	Why?
1	Show me fewer, more generic children topics	The current node has too many children topics
2	Show me more specific children topics	The current node has too few children topics
3	Add node as child	A child topic is wrongly placed under a different parent topic
4	Merge node as siblings	These topics cover similar content
5	Remove a node	This topic does not make any sense (i.e., garbage-topic)
6	Rename a node	The current label of this topic does not describe the topic properly

Table 4.2: A set of operations for revising the topic hierarchy.

Figure 4.3 provides a summary of what happens after applying each of the operations for revising the topic hierarchy. While some of these operations have similarities with the operations for revising topic models for a single conversation, we have designed a set of operations that are potentially more useful in the context of revising a topic hierarchy. In the remainder of the section, we provide a more detailed description on how each of these operations supports the user's tasks, and how the system revises the topic hierarchy according to each of these operations.

Show fewer, more generic sub-topics

In the initial topic hierarchy, some of the parent topics may have a large number of very specific children topics. If such a refined decomposition does not match the user mental model and/or information seeking task, it would be useful to allow her to change the topic hierarchy by showing only a few but more generic children

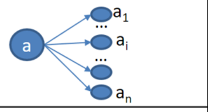
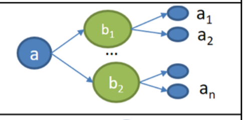
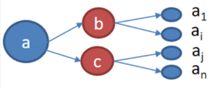
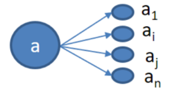
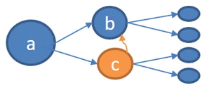
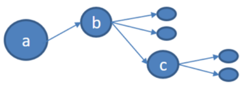
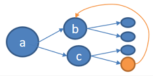
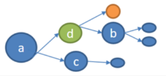
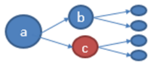
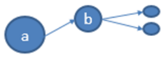
No	Operation Name	Operation type	Before operation	After operation
1	Show me fewer, more generic sub-topics	Unary		
2	Show me more specific sub-topics	Unary		
3	Add node as child	Binary		
4	Merge node as siblings	Binary		
5	Remove a node	Unary		
6	Rename a node	Unary	military security	cyber security

Figure 4.3: Illustrative examples of how the topic hierarchy changes as a result of applying different operations.

topics of that parent topic. In this way, the user can change the granularity level of a topic to be coarser. Note that this operation is similar to splitting a topic described in 4.3.1. However, the underlying computational method is slightly different in that, in this case, the system clusters the set of children topics of a node into a smaller set, as opposed to clustering the sentences of a topic into multiple sub-topics.

Method: Assume that a topic A has the following sub-topics $a_1, ..., a_n$. When the user requests for showing fewer sub-topics of A, the system clusters the set of sub-topics $a_1, ...a_n$ into a smaller set of sub-topics $b_1, ..., b_m$ where, $m < n$ (see Figure 4.3).

This is done by applying the same graph partitioning algorithm used in the generation of topic hierarchy, i.e., an approximate solution to n-Cut (see Section 3.5.2) [118]. More specifically, an undirected weighted graph $G(V, E)$ is constructed, where each node in $V = a_1, ..., a_n$ represents a sub-topic of A, and each edge $w(x, y)$ in E represents the similarity between two sub-topics. This similarity score is computed according to Equation 3.1. The number of clusters (i.e., $|m|$) in this clustering algorithm is automatically determined by maximizing a clustering objective function proposed in 4.1 [97].

Once the sub-clusters $b_1,...,b_m$ are created, representative keyphrases are generated for each sub-topic in $b_1,...,b_m$. This is done by running our topic labeling method only on the set of sentences covered by each sub-topic in $b_1,...,b_m$.

Show more specific sub-topics

This operation serves the opposite purpose of the previous operation, i.e., change the granularity level of a parent topic to a finer level by deriving and showing more specific children topics.

Method: Assume that the user requested to show more specific sub-topics of A. In response, the system removes the set of sub-topics in the immediate level (i.e., $b_1,...,b_m$) from the topic hierarchy and then links the sub-topics of $b_1,...,b_m$ i.e., $a_1,...a_n$ as the sub-topics of A (see Figure 4.3).

Add as Child

Sometimes a child topic node might be wrongly placed under a different topic instead of under a more appropriate parent topic. In such case, it may be useful to allow the user to move the child topic from its current topic and place it under a more appropriate parent. For instance, a topic namely ‘bending vulnerability’ was wrongly placed under the parent topic ‘iPhone5 models’ and the user realized that it would be more appropriate to add ‘bending vulnerability’ as a child under the parent topic ‘iPhone 6 bend’.

Method: When the user applies this operation on a sub-topic A_k, the system removes A_k from its current parent A and then assigns B as its new parent as suggested by the user (see Figure 4.3).

Merge as Siblings

This operation serves a similar purpose like the previous one, however, instead of adding a topic as a child to a node it is added as siblings to that node. The two siblings are then placed under a new parent node.

Method: The underlying computational method is similar to the 'merge by join' operation for revising the topic model for a single conversation. In essence, when the user applies this operation, the system removes the sub-topic A_k from its current parent A and merges it with the topic B (see Figure 4.3). The system then creates a new parent topic C and then assigns vertices such that,

$$V_C = V_{A_k} \cup V_B, V_A = V_A \setminus V_{A_k} \tag{4.4}$$

and edges such that,

$$E_C = E_{A_k} \cup E_B, E_A = E_A \setminus E_{A_k} \tag{4.5}$$

After that, the topic labeling method takes the portion of the conversation that consists of the sentences in V_C, thus generating a label for C that potentially represents descriptive keyphrases from both topics A_k and B.

Removing topics of a conversation

Sometimes, a topic is not relevant or interesting according to the user's current information needs. In such case, it may be useful to allow the user to remove such nodes and adjust the topic hierarchy accordingly.

Method: When the user applies this operation, the system removes the subtopic A_k, along with all the children nodes of A_k, from its current parent A and show the updated hierarchy to the user.

Renaming topics

If a topic doesn't accurately reflect its textual content, the user can rename the topic by giving a more appropriate label.

Method: The system updates the topic hierarchy by changing the label of the topic node and showing the updated label to the user.

4.4 Interactive Visualization for Topic Revision

We have extended the ConVis and MultiConVis interfaces to incorporate the topic revision operations described above. While doing so, we did not discard any existing features of these interfaces, rather we have complemented them with additional interactive features for revising the topics. We now discuss these extended interfaces (i.e., ConVisIT and MultiConVisIT) along with their interactive features.

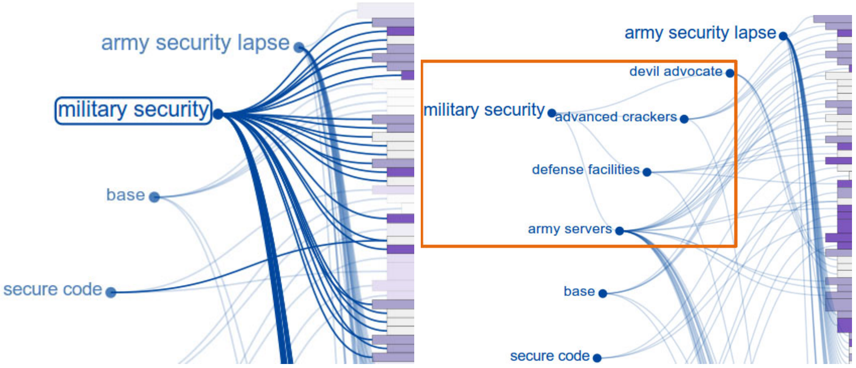
4.4.1 ConVisIT: Exploring a Single Conversation Using Interactive Topic Modeling

As the user explores a conversation, she may realize that the initial topic model is not helping her anymore, and may want to revise it. To support the user in such a situation, ConVisIT provides a set of interactive topic revision operations within the interface through some intuitive direct manipulation methods[2]. As the user performs these operations, the system updates the topic model and changes the visual encoding of the topic list from the initial flat list of topics into a multi-rooted tree organization. Such updates to the topic organization becomes visible to the user through perceptually meaningful animations, following the design guidelines of effective animation presented in [58]. In particular, we have devised staged animation for each operation, i.e., we break up the corresponding transition into a set of simple sub-transitions, allowing multiple successive changes to be easily observed.

For instance, when the user splits a topic by double clicking on it, the following sub-transitions occur (See Figure 4.4). First, the clicked topic A moves to the left along with its parent node(s) (if any), while existing nodes at the deepest level are pushed towards their new positions (up/down) around the circular layout to create angular spaces for the new sub-topics. Second, the new sub-topics A_1,A_2,...A_n appear and move from their parent's position (A) to their new positions. Third, labels appear for these sub-topics (see Figure 4.4b). Double clicking on A again causes it to collapse by following the exact reverse order of animation, i.e., the labels of the children move from their current positions to their parent and fade away, and then the parent moves to its previous position while other nodes move closer to the parent node to fill the gaps left by the removed children nodes.

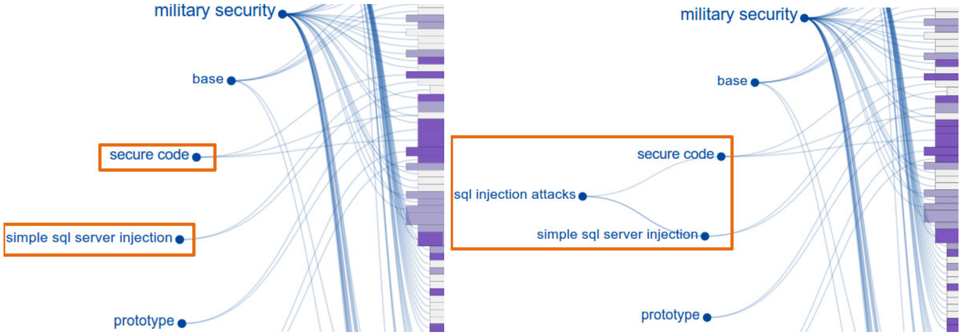
Merging of two topics can be performed by dragging a topic A over another topic B, which causes the system to update the topic model. As a result, a new

[2] A video demonstration of ConVisIT is available here: https://goo.gl/QALDvw



(a) Before splitting **(b)** After splitting

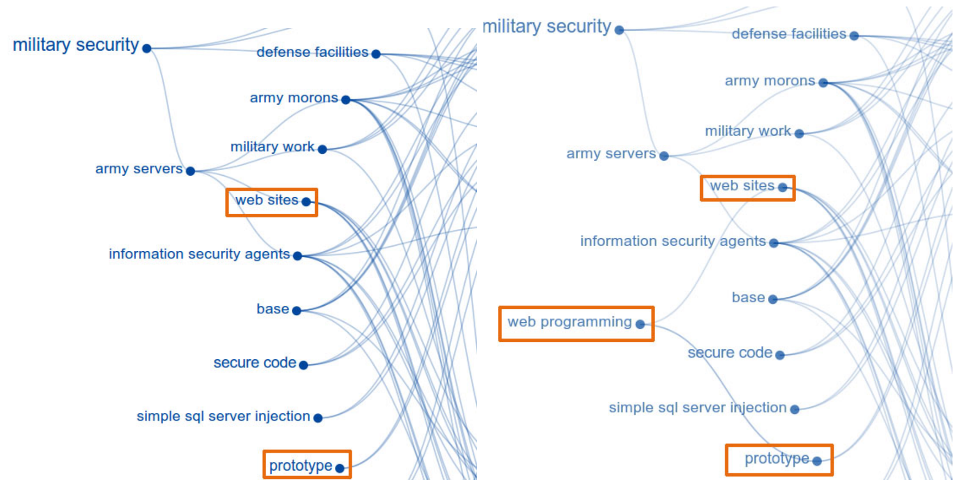
Figure 4.4: An example showing: (a) The user hovers over the topic 'military security' and decides to perform the split operation. (b) As a result, the topic moves to its left while the rest of the topics are pushed along the perimeter of the circular layout to create space for the new children.



(a) Before merge by joining **(b)** After merge by joining

Figure 4.5: An example showing: (a) The user decides to merge two topics by joining (indicated by orange color). (b) As a result, ConVisIT updates the topic organization where these two topic nodes are merged under the parent topic 'subject sql injection attack'.

parent topic C appears to the left and curved links are drawn from C to A and B to indicate parent-child relationship (see Figure 4.5). The user can subsequently double click on C to collapse it, which hides its sub-topics. Finally, if a child topic A_k is discovered to be wrongly placed under a topic A instead of under a more appropriate topic B, the user can drag A_k over B. As a result, the link of A_k with its



(a) Before merge by absorption

(b) After merge by absorption

Figure 4.6: An example showing: (a) The user realizes that the topic 'web sites' does not fit under the topic 'army server' and should be merged with the topic 'prototype'. (b) ConVisIT updates the topic organization where the previous link from 'web sites' to 'army server' is removed, and then 'web sites' is absorbed into a more generic parent topic 'web programming' along with 'prototype'.

parent A is removed and then a new parent node C appears that connects both A_k and B (see Figure 4.6).

As the user continues to perform interactive topic revisions, the topic organization can potentially grow quickly to multiple levels of hierarchy due to iterative splitting and merging. The current implementation can reasonably show a topic organization having a tree depth up to four levels, when the visualization is used on a 1920 x 1080 screen. This seems adequate for conversations with no more than a few hundreds comments, because the number of sub-topics grows exponentially with the depth of the topic hierarchy, and topics at the bottom of a hierarchy of depth four becomes so specific (i.e., cover so few sentences) that further splitting would be inappropriate. For instance, if we assume that the avg. branching factor in a single-rooted topic hierarchy is 3 and the conversation contains 300 sentences, each leaf of the topic hierarchy of depth 4 will contain on avg. $(300/3^4) = 3.7$ sentences.

4.4.2 MultiConVisIT: Exploring a Collection of Conversations Using Interactive Topic Modeling

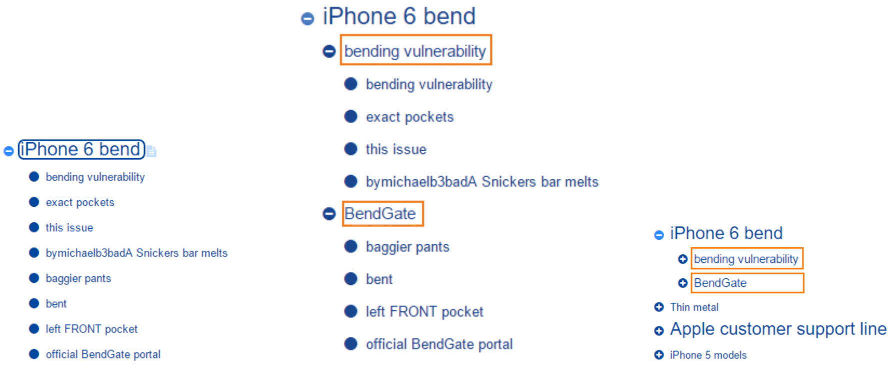
As the user explores the set of conversations, she may realize that the initial topic hierarchy does not match her mental model or information needs and may want to revise it. To support such situations, we developed a set of interactive techniques with respect to the topic revision operations listed in Table 4.2. We have also performed an informal user study to understand the potential usability problems and refine the topic revision operations to match with user's information needs. As the user performs these operations, the system revises the topic hierarchy and updates in the interface. Similar to ConVisIT, MultiConVisIT makes the revised topic organization visible to the user through staged animation [58] [3].

For instance, when the user asks the system to show fewer, more generic sub-topics of a topic *a*, the following sub-transitions occur, as shown in Figure 4.7. First, existing sub-topics of *a* (i.e., 'iPhone 6 bend') are moved vertically to provide space for their new parent topics. Second, the new parent topics appear. Third, the new parent topic nodes are collapsed so that the previous sub-topics become hidden. Double clicking on 'iPhone 6 bend' again results in showing more specific children nodes by removing the children in its immediate level.

If the user thinks that a topic is wrongly placed under a parent topic, she can change the topic assignment by dragging the topic over a different topic to which it is more related. The user can do this in two different ways: merge a topic as a sibling to another topic or place it as a child to another topic (see Figure 4.8). As the user drags a topic over another topic node, a dialog box appears that allows the user to decide whether she wants to merge the dragged topic as a sibling or as a child.

The user can also remove or rename topics. If the user feels that a topic is not relevant, or does not make any sense, she can drag that topic to the recycle bin. If a topic label does not represent its corresponding textual comments, the user can rename it. Finally, at any time the user can undo the latest topic revision operation that she made, by clicking on the 'Undo' button.

[3] A video demonstration of MultiConVisIT is available here: https://goo.gl/edT69x



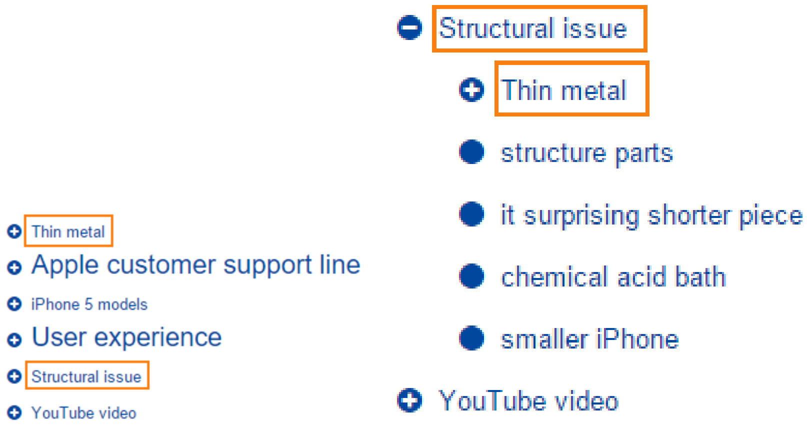
(a) Before the operation **(b)** New parents appear **(c)** Previous sub-topics are hidden

Figure 4.7: An example showing: (a) The user asks the system to show fewer, more generic topics of 'iPhone 6 bend'. (b) As a result, the sub-topics are moved vertically to create space for their new parents (indicated by orange color) and horizontally to move to the next level of the tree. c) the previous sub-topics of 'iPhone 6 bend' are hidden as their new parent nodes are collapsed.

4.5 Implementation

A server side component (in PHP) communicates with the topic modeling system (in Python) to produce the updated results. The visualization component, on the other hand, is implemented in JavaScript using the D3 and JQuery library, which is sufficiently fast to respond in real time to the user actions. The system runs on a laptop computer with a 2.4 GHZ processor and 16 GB RAM.

In ConVisIT, the average processing time for a topic splitting operation is 6.92 sec. and for a topic merging operation is 2.74 sec. (over the initial set of topics in our corpora). Similar processing time was observed in MultiConVisIT (7.34 sec. for *showing fewer, more generic topics* and 2.96 sec. for *adding a topic as sibling*). In order to increase the response time, topic split results were cached by the system for all the topics in the initial topic model, as well as for the sub-topics as soon as they are created upon topic revision operations. Similarly, the results for showing fewer, more generic topics were cached for the initial topic hierarchy.



(a) Before the operation **(b)** After the operation

Figure 4.8: An example showing: (a) The user decides to add 'Thin metal' as a child of the topic 'Structural issue' (indicated by orange color). (b) As a result, 'Thin metal' is now placed under the topic 'Structural issue'.

4.6 Evaluation

We now report two summative user studies that we have conducted in lab-based settings to compare ConVisIT and MultiConVisIT with interfaces that do not support human-in-the-loop topic modeling.

4.6.1 Study I

The goal of this study is to understand how the introduction of visual interfaces for exploring a single conversation may influence the user performance and subjective measures compared to more traditional interfaces. In this chapter, we have presented ConVisIT, which is highly interactive, providing the capability to revise topic models. Its precursor, ConVis (described in Chapter 2) is also an interactive visualization for exploring conversations, however, it does not support any topic revision operations. Finally, as a traditional interface for exploring conversation, we have re-implemented the interface to the popular Slashdot blog. The user study aims to answer the following questions:

(1) When we compare ConVisIT, ConVis, and the Slashdot interface, is there any difference in user performance and subjective measures?

- Does one interface help to find more insightful comments in a conversation?
- Is one interface perceived as more useful and easy to use?
- Is reading behavior influenced by the interfaces? If the answer is 'Yes' then how?

(2) What specific visualization features/components of the three different interfaces are perceived as more/less beneficial by the potential users (e.g., interactive topic revision, Thread Overview, and relations between facets)?

Methodology

We performed a laboratory-based summative study to compare among interfaces [78]. The study was designed with three interfaces as conditions: *Slashdot*, *ConVis*, and *ConVisIT*. The Slashdot interface follows a typical blog reading interface design and it serves as a suitable baseline for our experiment. It provides an indented list-based representation of the whole conversation as well as common functionalities of blog interfaces such as scrolling up and down, collapsing a sub-thread, and searching for terms. The primary reason for including ConVis as an interface condition was to verify whether any potential improvements in performance and user behaviour over a typical blog reading interface are due to the visualization features common between ConVis and ConVisIT, or due to the interactive topic revision feature (which is only present in ConVisIT). For fair comparison, different interface parameters such as screen size and font size were kept the same across all the interfaces. Moreover, a within-subject design was used for this experiment with interface as the within-subject factor, allowing us to directly compare the performance and subjective measures of each participant with respect to all three interfaces. Finally, all study aspects, including instructions and setup, went through several iterations of evaluation and pilot testing with two users, who did not participate in the actual study.

Participants

20 subjects (aged 19-43, 8 females) with considerable prior experience of reading blogs participated in the study. Figure 4.9 shows responses to statements regard-

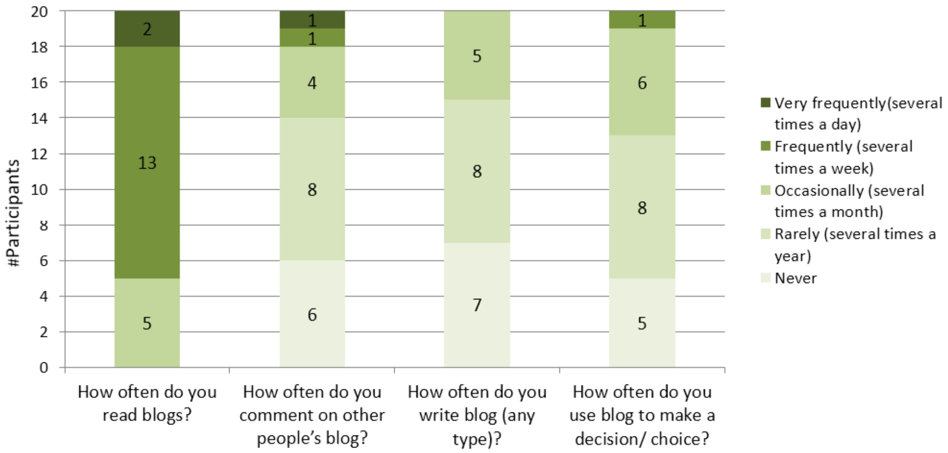


Figure 4.9: Responses to statements regarding the prior experience.

ing the prior experience of the participants. Notice that 75% of the participants reported that they read blogs at least several times a week. Moreover, 70% of the participants post comments on other people's blogs at least several times a month. The subjects held a variety of occupations including engineers, software developers, and university students mostly with strong science background. They were recruited through emails and Reddit posts.

Procedure and task

At the beginning, a pre-study questionnaire was administered to capture demographic information and prior experience with blog reading. Then, the user went through the following steps for each of the three interfaces: 1) In a scripted warm-up session, the interface was introduced to the participant. A sample conversation was shown using the given interface and the experimenter explained the interface actions by following the written script. 2) The participant was then asked to perform a task on a given conversation (a different conversation was provided for each interface). Rather than asking some specific questions, we provided an open-ended task to reflect the exploratory nature of blog reading. We asked the participant to explore the conversation according to her own interests using the given interface and write down a summary of the keypoints found while exploring the conversation. The study lasted approximately 90 minutes and each participant was paid $20 to participate.

We carefully selected three different conversations from the Slashdot blog corpora having similar number of comments (89, 101, and 89) to avoid potential variations due to the conversation length or complexity of the thread. Also, to counterbalance any potential learning effects due to the order of exposure to specific interfaces and conversations, the order was varied using a 3 x 3 Latin square.

During the study, we collected both quantitative data such as task completion time and qualitative data such as observations and questionnaires. After completing the task with each interface, participants rated the following aspects on a 5 point Likert scale in an in-study questionnaire: 1) *usefulness*: 'I found this interface to be useful for browsing conversations'; 2) *easeofUse*: 'I found this interface to be easy to use'; 3) *enjoyable*: 'I found this interface enjoyable to use'; and 4) *findInsightfulComments*: 'This interface enabled me to find more insightful comments'. At the end of the study, post-study questionnaires followed by a semi-structured interview were administered regarding the interfaces overall as well as their individual features[4]. Finally, we logged interface actions to better compare the usage patterns of the three different interfaces.

Results analysis

In-study questionnaires: The results of the in-study questionnaires are presented in Figure 4.10, showing the average rating expressed by the participants on four different measures. Since the data was collected using a standard 5 point Likert scale, the standard parametric analysis is not suitable due to the lack of normality [71]. Instead, we perform nonparametric analysis i.e., Mann-Whitney's U tests on the responses for each of these measures. Finally, all reported pairwise comparisons are corrected with the Bonferroni adjustment.

The analysis reveals that the interfaces significantly affected *findInsightfulComments*, with pairwise comparisons showing that ConVisIT was perceived to help them in finding more insightful comments than the ConVis and the Slashdot interfaces (see Figure 4.10). This is an important result because it supports

[4] The study materials for the user study can be found in Appendix B

Measures	Slashdot vs ConVis	Slashdot vs ConVisIT	ConVis vs ConVisIT
usefulness	$\boldsymbol{U = 82.5; p < 0.001}$	$\boldsymbol{U = 70.0; p < 0.001}$	$U = 182.5; p = 0.575$
easeofUse	$U = 178.0; p = 0.518$	$U = 177.0; p = 0.511$	$U = 195.0; p = 0.885$
findInsightful-Comments	$\boldsymbol{U = 62.0; p < 0.001}$	$\boldsymbol{U = 27.0; p < 0.001}$	$\boldsymbol{U = 131.0; p < 0.05}$
enjoyable	$\boldsymbol{U = 67.5; p < 0.001}$	$\boldsymbol{U = 119.0; p < 0.05}$	$U = 159.0; p = 0.24$

Table 4.3: Statistical analysis (Mann-Whitney's U test) on *usefulness*, *easeofUse*, *enjoyable* and *findInsightfulComments* measures (2-tailed p values).

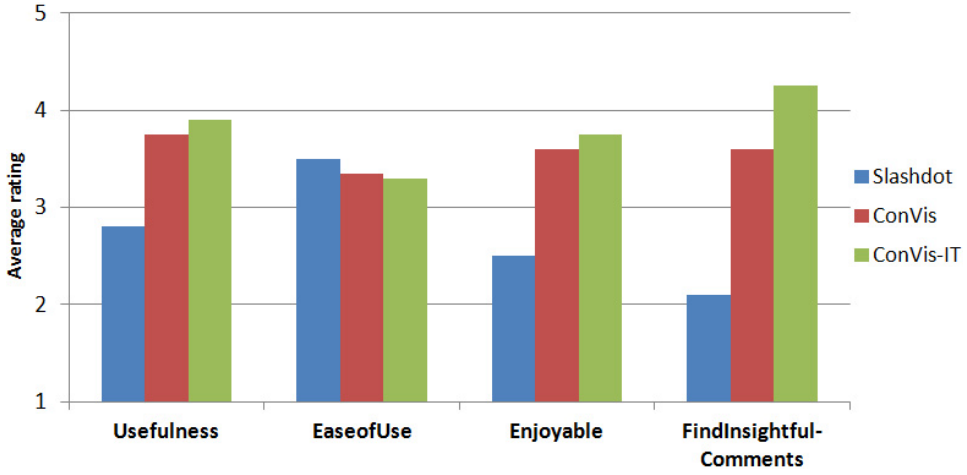


Figure 4.10: Average rating of the three interfaces by the participants for the following measures: *usefulness*, *easeofUse*, *enjoyable* and *findInsightfulComments* . Longer bars indicate better rating.

our intuition that by allowing the user to dynamically modify the topic organization (in ConVisIT), we enable her to find more insightful comments. There were also significant effects of interface on *usefulness* as shown in Table 4.3, with pairwise tests showing that ConVisIT and ConVis were perceived to be significantly more useful than the Slashdot interface. Moreover, ConVisIT was rated slightly more useful than ConVis, although the difference was not significant. Similar results were obtained on the *enjoyable* measure, where ConVis and ConVisIT were rated significantly higher than Slashdot (see Figure 4.10). Finally, the *easeofUse* measure is not significantly affected by the interfaces, indicating that none of the interfaces was superior on this measure. However, this is a favorable outcome for ConVisIT in that even though its interactive features are more complex than in ConVis, the participants did not report ConVisIT as being significantly more difficult to use. Similarly, it is also a favorable outcome for both ConVisIT and ConVis, since, in spite of their complexity, they were found to be as easy to use as the simpler traditional blog interface.

Interface features: The in-study questionnaire also included a number of questions regarding the usefulness of specific features of the three interfaces. To complement this data, we also analyzed the interaction log data of ConVis, ConVisIT, and Slashdot. The quantitative results of the subjective ratings are provided in Figure 4.11. We can readily see that the majority of the responses regarding features of the Slashdot interface range from strongly disagree to neutral. In contrast, responses regarding ConVis and ConVisIT features are dominated by strongly positive to neutral ratings.

Regarding topic revision operations, *Split* was found to be more useful (35% strongly agree and 40% agree) than *Merge* (20% strongly agree and 25% agree). This is also evident from the usage of these operations, as the split operation was used more frequently (5.3 times on average) than merge (1.6 times on average). Moreover, 16 out of 20 users performed split operation prior to performing any merge operation. A possible explanation is that participants generally found the initial topic model results to be too coarse grained with respect to their information needs, expertise and mental model, and therefore they tended to apply split operation both earlier on and more frequently than the merge operation so that they could read at finer topic granularity.

An interesting observation from the log data is that even though some features were common in both ConVis and ConVisIT, they were used more frequently with ConVisIT. For example, participants hovered and clicked on topics and comments more times on average using ConVisIT than using ConVis, as shown in Figure 4.12. A possible explanation is that due to the presence of interactive topic revision features, the participants could create topics that were more useful to them and therefore they relied on topics more frequently in their exploration.

Time: Interestingly, the average time required to complete the tasks was not significantly affected by the interfaces, with Slashdot, ConVis, and ConVisIT re-

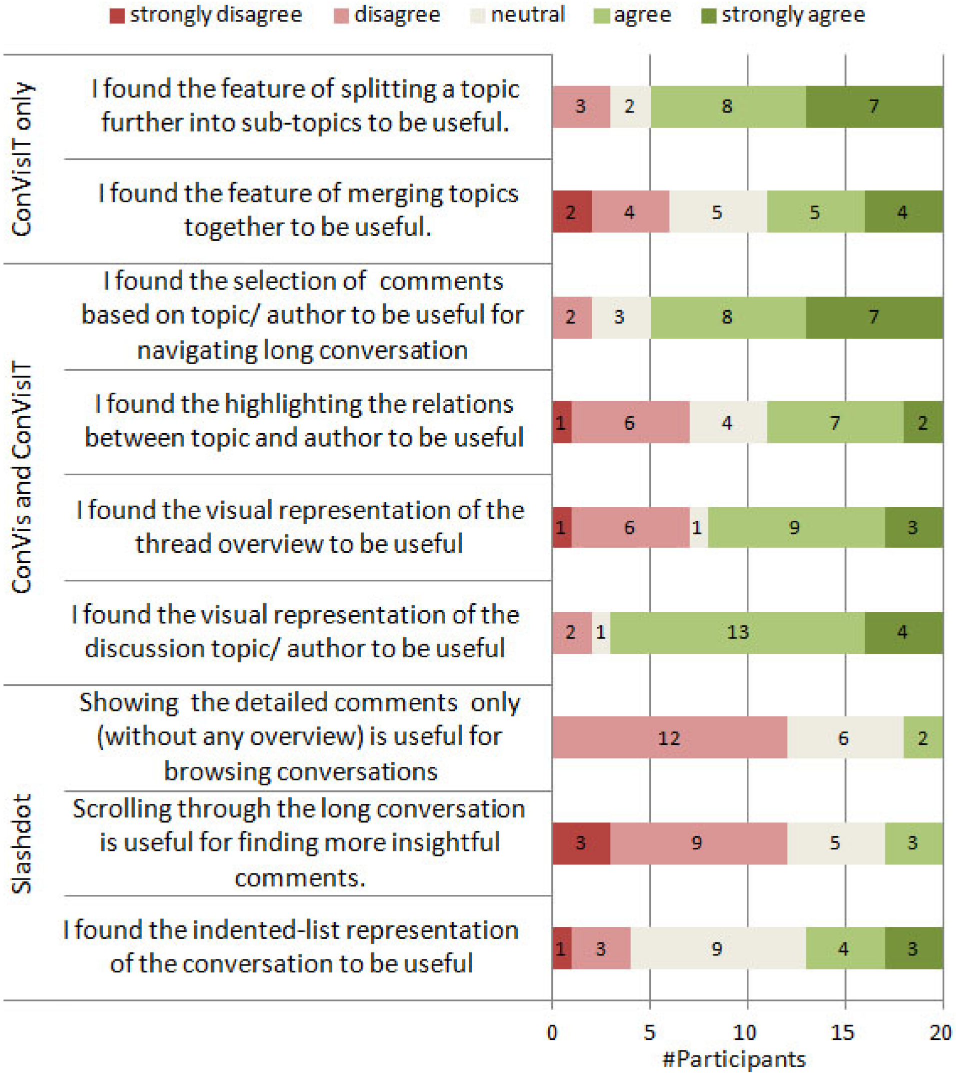


Figure 4.11: Responses to statements regarding specific features of the three interfaces under investigation.

quiring on average±*sd* 1056 ± 479, 1240 ± 486 and 1159 ± 604 secs respectively. This result is rather promising, because it indicates that participants were not slowed down by the fact that they were unfamiliar with the topic revision operations and by the overhead involved in performing those operations.

User-generated summaries: Recall that during the study, each participant was

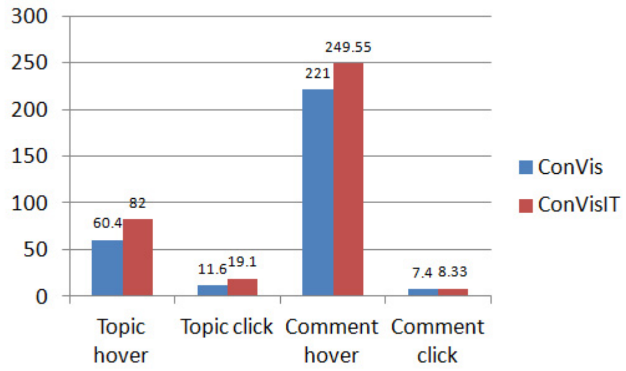


Figure 4.12: Some interaction log statistics for interactions that are common between ConVis and ConVisIT (based on avg. values among 20 participants)

asked to write down a summary of the keypoints she had found after exploring a conversation using the given interface. We have analyzed these summaries to verify whether the three different interfaces for exploring conversations had an effect on the user's ability to write high-quality summaries [5].

Evaluation Protocol: For the purpose of evaluating the summaries, we recruited two human raters with research experience in natural language processing, but who were not involved in this research in any way. For each of the three conversations, a set of summaries were presented to the raters. The raters were also told that the original blog conversation from the Slashdot corpora was not available anymore and so they would have to rate each of these summaries on a 5-point likert scale according to their overall satisfaction with its content. Note that the summary raters did not know which interface was used to produce each summary[6].

While rating the summary the rater was asked to consider the following three criteria: 1) How informative this summary is (the more informative the better); 2) How insightful this summary is (the more insightful the better); 3) Whether there is any redundant information within the summary (the less redundant the better). The rater was also told that the focus of this evaluation was about the content of the

[5] Although there were 20 participants, for 2 participants the user-generated summaries were missing, therefore we had 54 (18 x 3) summaries with an equal number of task-system pairs.

[6] The instructions for rating blog summaries are provided in Section B.1.3

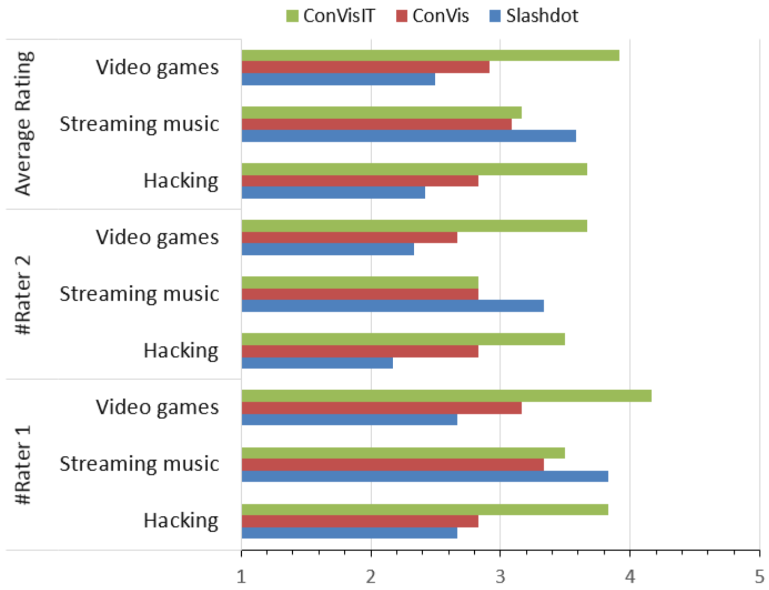


Figure 4.13: Average ratings for user generated summaries based upon two human raters

summary, not about whether it was grammatically correct and/or fluent, therefore they were told to ignore these linguistic aspects while rating. Finally, the rater was allowed to revise the ratings of the summaries she had already assessed, as she moved down the list and saw more and more summaries.

We converted the Likert responses from a scale of ‘Extremely Poor’ to ‘Excellent’ to a scale of 1 to 5, with 1 corresponding to ‘Extremely Poor’, and 5 to ‘Excellent’. The weighted Kappa coefficient was then computed with linear weights to determine the level of agreement between the two raters. The set of weights chosen was $[0, 0.25, 0.50, 0.75, 1.0]$. The resultant weighted Kappa coefficient was 0.449, which represents good agreement [46].

Results: Figure 4.13 shows the results of the evaluation of the user-generated summaries created with the support of three different interfaces. These results suggest that the interface used by the participants for reading a conversation influences their ability to write good summaries. In particular, for two conversations (‘Hacking’ and ‘Video games’), the summaries created with the support of the ConVisIT interface received considerably higher ratings by both human raters compared to

Conversation	Rater	Slashdot vs ConVis	Slashdot vs ConVisIT	ConVis vs ConVisIT
Hacking	R1	$U = 15.5; p = 0.665$	$\boldsymbol{U = 5.5; p < 0.05}$	$\boldsymbol{U = 6.5; p < 0.05}$
	R2	$U = 11.0; p = 0.207$	$\boldsymbol{U = 6.5; p < 0.05}$	$U = 12.0; p = 0.301$
Streaming music	R1	$U = 13.5; p = 0.450$	$U = 13.5; p = .423$	$U = 16.5; p = 0.799$
	R2	$U = 12.0; p = 0.283$	$U = 10.0; p = 0.092$	$U = 17; p = 0.849$
Video games	R1	$U = 13.5; p = 0.452$	$\boldsymbol{U = 6.0; p < 0.05}$	$\boldsymbol{U = 6.5; p < 0.05}$
	R2	$U = 13.0; p = 0.388$	$\boldsymbol{U = 6.5; p < 0.05}$	$\boldsymbol{U = 7.0; p < 0.05}$

Table 4.4: Statistical analysis (Mann-Whitney's U test) on summary ratings (2-tailed p values)

the two interfaces that do not support human-in-the-loop topic model. Pair-wise tests show that such differences are significant (Table 4.4). For the other conversation ('Streaming music'), there was not any significant difference between the three interfaces. This finding may be due to differences in conversation length (the 'Streaming music' conversation was longer) and/or to differences in the amount of noise/error in the initial topic models for the different conversations.

Overall preference: At the end of the study, participants were asked to indicate their overall preference for a blog reading interface and then justify their choice. 60% of the participants indicated a preference for ConVisIT, 25% for ConVis, and 15% for Slashdot. Most of the participants who chose ConVisIT felt that the topic revision operations were very helpful in finding relevant comments: *"ConVisIT is the most convenient interface because of its splitting and merging features. Using this interface to understand the conversation, I really did not have to go through all the comments" (P19).* It was also evident that when the granularity level of the topics did not match the user's information needs, ConVisIT was especially helpful for navigation: *"Sometimes the first-level keywords are way too generic, so it's better to navigate via second-level categories (P11)"*. However, participants did become frustrated in a few cases, when ConVisIT could not accurately split the topics into meaningful list of sub-topics as mentioned by P13: *"...I enjoyed the ability to split apart topics, though I think it would benefit from better categorization of topics as I felt like some were misclassified"*.

Those participants who chose the ConVis interface over its counterparts emphasized the utility of its visual components, i.e., the visual representation of the thread and highlighting the relations between topics and comments, which *"...makes it easier to find out which comments are more interesting"*, and *"...allowed me to see more of what was going on, how comments were inter-related, as well as kept me interested and focused on the thread as a whole." (P4).* The primary reason for preferring ConVis over ConVisIT was that sometimes the revised topic organization became too cluttered or made the navigation too complex: *"... drilling down to a sub-topic made the graph look too cluttered up. Sometimes, it was harder to figure out if two topics were at the same level or not based on the layout." (P7)*, and *"It felt like a good mix, others were too complex (ConVisIT) or too simple (Slashdot)." (P2).*

Three participants who preferred the Slashdot interface felt that it was easier to use, although one said *"...it is not giving me the structural information that I am interested about " (P16).* Another reason was that they were so much familiar with this interface: *"Scrolling through the conversation was good enough for me to find important topics in it, maybe because I am used to reading things this way." (P15).*

4.6.2 Study II

While in study I we compared interfaces for exploring a single conversation, in study II we focus on investigating the effectiveness of interactive topic modeling for a set of conversations, namely on the utility of topic hierarchy revision operations. The study was designed in a similar way to the study I (Section 4.6.1), with the primary difference being that unlike Study I in which ConVisIT was compared with both ConVis and a traditional interface, here we compare MultiConVisIT with MultiConVis only. Recall, however, that MultiConVis was already compared to a traditional interface in a separate study described in Chapter 3, which showed that MultiConVis outperformed its counterpart along several subjective measures (e.g., usefulness, enjoyable) and was preferred by the majority of participants. We decided to run two separate studies, mainly because the task for exploring and analyzing a set of conversations usually require significantly more time than exploring a single conversation; therefore, running a within-subject study with three interfaces would have been less feasible.

In this study, we aim to answer the following questions:

(1) When we compare MultiConVisIT with MultiConVis for exploring a set of conversations, is there any difference in user performance and subjective measures?

(2) What specific topic revision features of the MultiConVisIT interface are perceived as more/less beneficial by the potential users?

Methodology

Similarly to what was done in study I (see Section 4.6.1), we designed a summative evaluation through controlled experiments with two interfaces as conditions: *MultiConVis*, and *MultiConVisIT*. A within-subject design was used with interface as the within-subject factor, allowing us to directly compare the measures of each participant with respect to both interfaces. We again refined all study aspects, including instructions and setup, through several iterations of pilot study with three users, who did not participate in the actual study.

Participants

We conducted the study with 16 users (aged 18-28, 9 females) who have considerable experience of reading blogs. The participants held a variety of occupations ranging from journalists, engineers, and students from both graduate and undergraduate levels. They were recruited through emails and social networks (Facebook and Reddit posts).

Procedure and Tasks

At first, a pre-study questionnaire was administered to capture demographic information and prior experience with exploring blog conversations. Then, the participant went through the following steps for each of the two interfaces: 1) In a scripted warm-up session, the interface was introduced to the participant using a sample dataset. 2) The participant was then asked to perform a task based a set of conversations. For each interface, a different set of conversations was provided.

Task: Similarly to what was done in study I (Section 4.6.1), we provided an open-ended task of exploring the set of conversations and write the summary of major discussion points. However, since MultiConVis and MultiConVisIT show a set of conversations along with a large number of topics organized into a hierarchy, we asked the participant to summarize the major points under the most appropriate corresponding topic in the hierarchy, rather than creating a plain summary. In this way, we were able to test whether the the interface has an effect on the user's ability to find the most insightful and informative comments and then able to summarize them in a coherent way under different topics.

The participant was provided a task scenario where she would work as a business analyst for Apple and needed to analyze the set of conversations form a given dataset,so that later on she could discuss her insights with her colleagues. For example, when the dataset on iPhone bending was provided, the participant was given the following task:

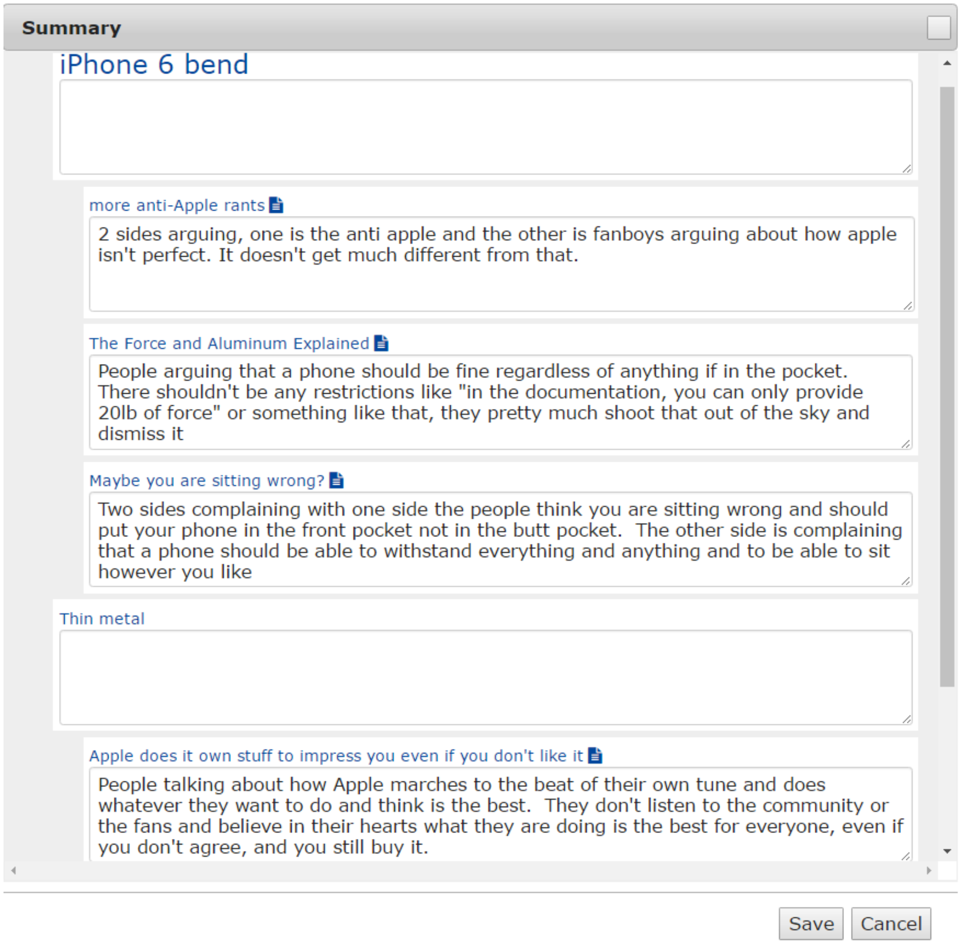
The issue of iPhone bending went viral on social media after the iPhone 6 was launched in September 2014. Soon after the product was released, some people claimed that this new phone can easily bend in the pocket while sitting on it. This incident triggered a huge amount of discussions in Macrumors, a blog site that regularly publishes Apple related news and allows participants to make comments.

You are working for Apple as a business analyst. Your task is to find the major discussion points about the iPhone bending issue and summarize each of them under the most appropriate corresponding topic. The final outcome will be a summary of the conversations organized according to a topic hierarchy that you will have to show and discuss with your colleagues. So you want to make sure that the topic hierarchy and the summary of major discussion points are as informative and as clear as possible.

To facilitate the above task, the interface allowed the user to click on a 'summary' button adjacent to each topic node, so that the user could enter the summary of a topic within a text box. At any time, the user could click on the 'Show summary view' button to review what summary has been added so far for different topic nodes. Figure 4.14 shows such a summary view, where the user has entered summaries under the corresponding topics.

We selected two different datasets crawled from the Macrumors site for testing ('iPhone bend' and 'iPad release')[7]. The number of conversations in the datasets was kept the same (16 conversations in each dataset) to avoid potential variations

[7] These are the same datasets that were used for the user study described in Chapter 3



2 sides arguing, one is the anti apple and the other is fanboys arguing about how apple isn't perfect. It doesn't get much different from that.

People arguing that a phone should be fine regardless of anything if in the pocket. There shouldn't be any restrictions like "in the documentation, you can only provide 20lb of force" or something like that, they pretty much shoot that out of the sky and dismiss it

Two sides complaining with one side the people think you are sitting wrong and should put your phone in the front pocket not in the butt pocket. The other side is complaining that a phone should be able to withstand everything and anything and to be able to sit however you like

Thin metal
Apple does it own stuff to impress you even if you don't like it

People talking about how Apple marches to the beat of their own tune and does whatever they want to do and think is the best. They don't listen to the community or the fans and believe in their hearts what they are doing is the best for everyone, even if you don't agree, and you still buy it.

Figure 4.14: An example of summaries of different topics created by a user during the study.

due to the amount of conversational data. Also, to counterbalance any potential learning effects due to the order of exposure to specific interfaces and dataset, the order was varied using a 2 x 2 Latin square. During the study, we collected both quantitative data such as task completion time and qualitative data such as observations and questionnaires. Finally, a post-study questionnaire was administered regarding the user's experience with two interfaces[8]. The study lasted approxi-

[8] The study materials for the user study can be found in Appendix B

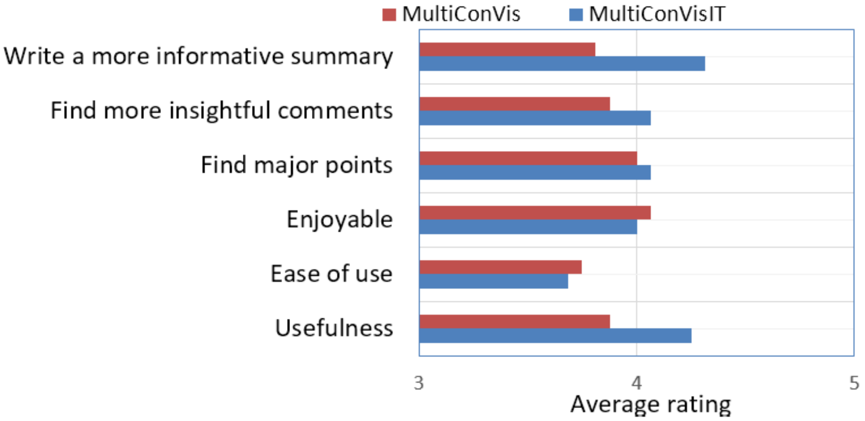


Figure 4.15: Average rating of the two interfaces by the participants for the following measures: *usefulness*, *easeofUse*, *enjoyable*, *findInsightfulComments* and *writeInformativeSummary*. Longer bars indicate better rating.

mately 90 minutes and each participant was paid $20 to participate.

Analysis of results

After completing the task with each interface, participants rated six different measures in an in-study questionnaire. The results of this questionnaire are presented in Figure 4.15. The pairwise comparisons using Mann-Whitney's U tests indicate that MultiConVisIT is superior on two different measures out of six: *usefulness* ($U = 75; p < 0.05$) and enable to write a more informative summary i.e., *writeInformativeSummary* ($U = 76.5; p < 0.05$). For the other measures, MultiConVisIT is still superior over its counterpart except for *easeofUse*, however, the results of these measures are not significant. Overall, this is a promising result as it suggests that in general participants found the MultiConVisIT interface to be more useful and they felt that it helped them to write a more informative summary about the set of conversations.

Interface features: The in-study questionnaire also included a number of questions regarding the usefulness of specific features of the two interfaces. To complement this data, we also analyzed the interaction log data collected during the user experiments. The quantitative results of the subjective responses are pro-

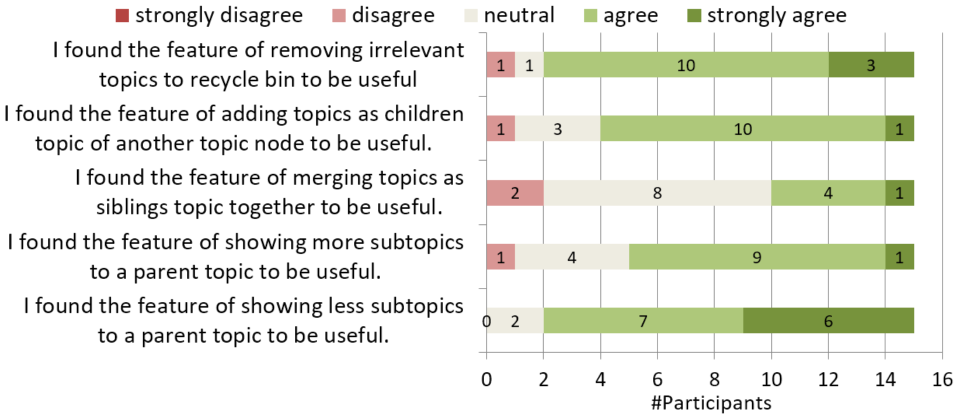


Figure 4.16: Responses to statements regarding specific features of the MultiConVisIT interface.

vided in Figure 4.16[9]. We can readily see that the majority of the responses regarding topic revision operations range from strongly agree to neutral. These results suggest that most of the users found the ability to organize topics through different topic revision operations to be useful. This is also evident from the usage of these operations, as users applied them quite frequently ('Show me fewer, more generic children topics' operation was used 3.9 times on average; 'Show me more specific children topics' was used 2.1 times on average; and 'remove a node' was used 3.4 times on average).

Time: The average time required to complete the tasks was not significantly affected by the interfaces, with MultiConVis and MultiConVisIT requiring on average$\pm$*sd* 1490 ± 369 and 1610 ± 321 secs respectively.

Overall Preference: In the questionnaire, participants were also asked if they would prefer MultiConVisIT over its counterpart. 68.75% of participants indicated a preference for MultiConVisIT, while 31.25% indicated their preference for MultiConVis.

Many of the participants who chose MultiConVisIT indicated that the ability to organize the topic hierarchy according to their own mental model and current tasks was the primary reason for their preference: *"Sometimes the topics were not organized in the way I expected. By organizing the topics into categories according*

[9] Among 16 participants, one participant's questionnaire data was missing.

to my own way was very useful for browsing. it makes the navigation easier...' (P7).

Another participant mentioned that *"The additional features are somehow better when it comes to getting the main topics. I found the collapsing of topics together very interesting because if one wants to look at a very specific part of the discussion, that is enabled* (P4)".

Those who preferred the MultiConVis interface indicated that they found it easier to learn compared to MultiConvisIT and few of them thought that the existing topic hierarchy was already sufficient for them *"...it took me quite a while to get used to the added features (of MultiConVisIT)"* (P11).

User generated summaries: Similarly to the analysis in Study I (Section 4.6.1), we have compared the summaries written by the users to verify whether the interfaces for exploring conversations had an effect on the user's ability to write high-quality summaries. For this purpose, we employed two human raters who were not involved in this research in any way. For each of the two set of conversations, a set of summaries were presented to the raters. While rating the summary the rater was asked to consider the three criteria as stated in Section 4.6.1 (informativeness, insightfulness, and redundancy of the summary content). Again, the rater was told to focus on the content of summaries rather then the grammatical correctness while rating the summary[10].

We converted the Likert responses from a scale of 'Extremely Poor' to 'Excellent' to a scale of 1 to 5, with 1 corresponding to 'Extremely Poor', and 5 to 'Excellent'. The weighted Kappa coefficient was then computed with linear weights to determine the level of agreement between the two raters. The set of weights chosen was $[0, 0.25, 0.50, 0.75, 1.0]$. The resultant weighted Kappa coefficient was 0.471, which represents good agreement [46].

Results: Figure 4.17 shows the results of the evaluation of the user-generated summaries created with the support of MultiConVis and MultiConVisIT. One could readily notice that for both sets of conversations ('iPhone bend' and 'iPad release'), the summaries created with the support of the MultiConVisIT interface received considerably higher ratings by both human raters compared to its counterpart that do not support human-in-the-loop topic model. Pair-wise tests show that in three

[10] The instructions for rating blog summaries are provided in Section B.2.3

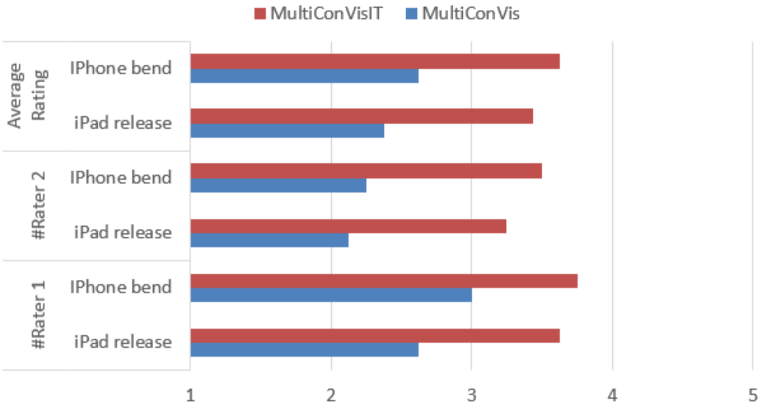


Figure 4.17: Average ratings for user generated summaries based upon two human raters

Dataset	Rater	MultiConVis vs MultiConVisIT
iPhone bending	R1	$U = 21.0; p = 0.23$
	R2	$\boldsymbol{U = 11.0; p < 0.05}$
iPad release	R1	$\boldsymbol{U = 14.0; p < 0.05}$
	R2	$\boldsymbol{U = 12.0; p < 0.05}$

Table 4.5: Statistical analysis (Mann-Whitney's U test) on summary ratings (2-tailed p values).

out of four cases of comparisons between the two interfaces, the differences are significant (see Table 4.5).

4.7 Discussion

We now discuss the summary of findings from the two user studies, as well as the limitations of our interactive topic modeling approach.

4.7.1 Summary of Findings

Based on our analysis of the study results, we now revisit our research questions. The first user study reveals that overall ConVisIT was the most preferred interface, and was rated higher over its counterparts on the *findInsightfulComments* measure. Similarly, the summaries created by the participants after reading the conversation using ConVisIT tend to receive at least similar or higher ratings compared to the other two interfaces. In contrast, Slashdot was the least preferred interface, and it received significantly lower rating on three different measures. As for ConVis, it seems to provide a middle ground between the other two interfaces and its topic organization, although static, was found to be visually less cluttered than the one of ConVisIT. In general, this shows that while interactive topic model can be beneficial to the user, such feature may introduce visual clutter and interaction costs at least for some users. Finally, there were no significant differences among the interfaces in terms of *easeOfuse* and *time to task completion*, in spite of the higher complexity of ConVis and ConVisIT.

In the first study, we also analyzed what specific visualization features/components of the interfaces are perceived as more/less beneficial by the potential users (e.g., interactive topic revision, Thread Overview, relations between facets). We found that in general, the visualization features of ConVis and ConVisIT received higher rating than the ones of Slashdot. Interestingly, we found that subjective reactions about different features of the interfaces such as split, merge, and clicking on topic directly correlates with their frequency of use. More importantly, we found that not all interactive topic revision operations were equally received. For example, the split operation was used more frequently than its counterparts. Although we have proposed some possible explanations, this issue needs to be further investigated.

From the second study, we again found the interactive topic modeling approach to be effective for exploring and analyzing a set of conversations. More specifically, MultiConVisIT was found to be more useful, although few users reported that they needed more time to learn the additional interactive topic revision features. The interface was also preferred over its counterpart that does not provide interactive topic revision operations. Moreover, it enhanced the participants' perceived ability to write a more informative summary over MultiConVis. Such a finding is also supported by the objective evaluation of user-generated summaries, as summaries produced by the user with MultiConVisIT were rated higher by external raters (based on how informative, insightful, and non-redundant the summary is). Finally, the majority of the participants rated most of the topic hierarchy revision features positively.

4.7.2 Limitations

In the second study, to avoid increasing complexity in the design, we have only tested the interactive topic features for revising the topic hierarchy in MultiConVisIT. However, using MultiConVisIT, the user can drill down to a single conversation with the ConVis view, therefore we could also provide the topic revising features for a linear topic model, as described in Section 4.3.1. In future work, we may investigate how these two topic revision approaches could be combined within the MultiConVisIT interface, and whether (and how) that may lead to potential differences in user performance and subjective measures.

Another interesting point is that in both studies, the system collected the topic revision feedback from each individual user which is not shared with other potential users. Arguably, it could be useful for a user to share her refined topic models so that other potential users exploring the same dataset might benefit. Therefore, a promising direction would be to incorporate topic revision feedback from multiple users with the aim of building more accurate, shareable topic models.

4.8 Summary

In this chapter, we presented and evaluated a novel human-in-the-loop topic modeling approach to support the exploration of online conversations. We devised a set of topic revision operations specifically for asynchronous online conversations and incorporated them into our visual text analytics systems (i.e., ConVis and MultiConVis). By utilizing our interactive topic revision approach, users can explore and revise the topic model to better fulfill their information needs.

The user studies reported in this chapter reveal that both ConVisIT and MultiConVisIT were preferred by the majority of the participants over their counterparts that do not support interactive topic revision. Moreover, our analysis shows that summaries written by participants during the exploration of conversations received higher (or at least equal) ratings by human raters, when ConVisIT and MultiConVisIT were the interfaces used to explore online conversations. In essence, the results from the studies indicate that users benefit from getting more control over the topic modeling process while exploring conversations.

Chapter 5

Tailoring Our Visual Text Analytics Solutions to a Community Question Answering Forum

So far, we have presented a set of systems covering two different dimensions of our visual text analytics design space, namely single vs. a set of conversations and static vs. human-in-the-loop model. While developing these systems, we did not restrict our solutions to any specific domain problem faced by users. Furthermore, our evaluations were limited to either case studies or lab studies.

In this chapter, we are interested in understanding how our generic visual text analytics solutions can be applied and tailored to a specific domain problem. To answer this question, we present a design study in a community question answering (CQA) forum, where our visual text analytics solutions were simplified and tailored to support information seeking tasks for a user population possibly having low visualization expertise.

A crucial aspect of this work is that unlike the evaluations we have presented in previous chapters, we evaluated the new system in a more ecologically valid way, by deploying it in a real-world environment in which it was tested with hundreds of real users. Through this large-scale online study, we gained deeper insights about the potential utility of the system, as well as learned generalizable lessons for designing visual text analytics systems for the CQA forums and similar domains of conversations [1].

5.1 Introduction

Community question answering forums, such as StackExchange, Yahoo! Answers, and Quora are becoming more and more popular these days.[2] They represent effective means for communities of users around particular topics to share information and to collectively satisfy their information needs. CQA forums typically organize their content in the form of multiple topic-oriented *question–comment threads*, where a question posed by a user may be answered by a possibly long list of comments from other users.

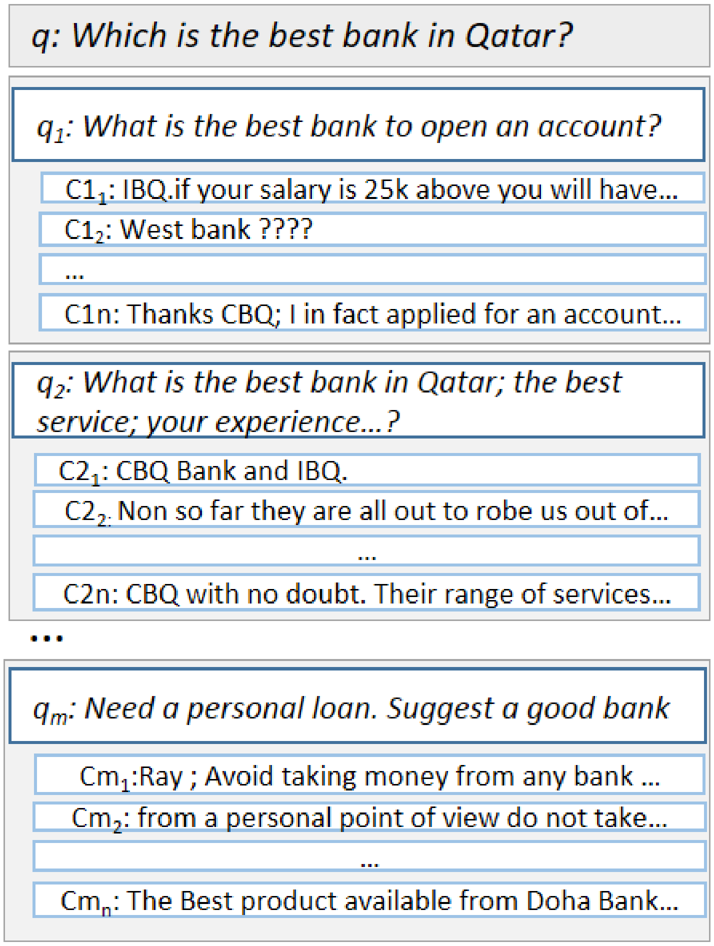
Many such online forums are not moderated, which often results in very noisy and redundant content. Users tend to deviate from the original question and engage in discussions on completely irrelevant or only loosely related topics. At the same time, similar questions may be posted repeatedly with minor variations. This near-duplicity is difficult to track for users, who are usually offered only simple search capabilities by the forum interface. When relevant answers to user questions are scattered around multiple related conversations and buried among a large number of comments, the user is facing a challenging information processing task, which, without proper support, leads to information overload.

For example, consider John, who is an expatriate, just arrived in Qatar and is seeking recommendations for a good bank. When he searches for 'Which is the best bank in Qatar?' in the Qatar Living forum[3], a very popular site in Qatar, it returns about a dozen previously asked questions, such as ' What is the best bank to open an account?' or 'What is the best bank in Qatar for small business?' (see Figure 5.1). Each of these questions is followed by a set of comments, resulting in

[1] This chapter is a modified version of our paper *CQAVis: Visual text analytics for community question answering*, by Enamul Hoque, Shafiq Joty, Lluís Màrquez and Giuseppe Carenini; in Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI), pp. 161-172, 2017 [63].

[2] stackexchange.com, answers.yahoo.com, quora.com

[3] http://www.qatarliving.com/forum



q: Which is the best bank in Qatar?

q_1: What is the best bank to open an account?

$C1_1$: IBQ.if your salary is 25k above you will have...

$C1_2$: West bank ????

...

C1n: Thanks CBQ; I in fact applied for an account...

q_2: What is the best bank in Qatar; the best service; your experience...?

$C2_1$: CBQ Bank and IBQ.

$C2_{2.}$ Non so far they are all out to robe us out of...

...

C2n: CBQ with no doubt. Their range of services...

...

q_m: Need a personal loan. Suggest a good bank

Cm_1:Ray ; Avoid taking money from any bank ...

Cm_2: from a personal point of view do not take...

...

Cm_n: The Best product available from Doha Bank...

Figure 5.1: An example of a new question q asked by a user which is shown at the top, followed by a set of related thread questions $q_1, \ldots, q_n$ and their comments.

hundreds of comments in total. Given the large number of comments from multiple related threads, it would be very difficult and time-consuming for John to identify and make sense of useful comments using a traditional interface.

In this chapter, we present CQAVis, an intelligent visual interface specifically tailored to help users find comments that provide good answers to a new question (i.e., never asked in exactly this form before) in community-created forums. CQAVis was designed by simplifying and tailoring our generic solutions for blog conversations, to take into account specific features of CQA data and tasks. The resulting interface allows the user to start with a new question, then to explore the related threads to find the ones that seem to be most relevant to her information needs, and eventually to navigate through the comments of a thread in search for relevant answers to her question. The underlying text analytic module dynamically ranks potential answers to a new question by combining two relevant measures: (*i*) how good or useful the comment is with respect to the thread question, such as q_1, q_2 in Figure 5.1, and (*ii*) how similar the thread question is with respect to the new question (q).

Our system was deployed in the Qatar Living forum site to evaluate our interface among hundreds of real users. Qatar Living forum was suitable for our study, because it represents the type of forums where the information overload problem, as described above, could be more prevalent due to unmoderated noisy content. Moreover, a large number of its users have limited expertise in using visual interfaces, which poses critical challenges to designing interfaces.

The primary contributions of our work include: 1) characterization of the CQA forums by identifying user tasks and some key design needs; 2) design of CQAVis that demonstrates how our generic approach for integrating NLP and InfoVis techniques presented in Chapter 2 and 3 can be applied and tailored to meet these specific user needs; 3) the evaluation of the tool in the wild in an ecologically valid testing by deploying the system among real forum readers, which in turn reveal that the overall approach for combining NLP and InfoVis techniques presented in this dissertation can be effective for a diverse range of user population; and 4) generalizable lessons learned from the study that can be useful to design visual interfaces for online conversations in other domains such as news comments and health forums, as well as to design for user populations possibly having low visualization literacy.

5.2 The Design Process

Our design study process followed the nine-stage framework proposed by Sedlmair et al. [111]. In particular, we focused on four core phases of the design framework, i.e., discover, design, implement, and deploy:

1) Discover: In this stage, we analyzed the needs, problems, and requirements in the domain of CQA forums through literature review and conducting in-depth interviews with Qatar Living forum users and administrators.

2) Design: After reaching a shared understanding of the CQA domain, we explored the design space, by analyzing the CQA data and tasks and how our current interfaces can be re-designed to support those tasks. We applied an iterative design approach, starting with paper prototyping, followed by prototyping on a limited annotated dataset which led us to the final prototype on the whole forum datset.

3) Implement: We developed both client and server side components in collaboration with the Qatar Living administrators.

4) Deploy: Following several pilot studies and corresponding refinements of the prototype, we deployed the tool as a beta version in the Qatar Living website and gathered feedback about its use in the wild.

5.3 User Requirements Analysis

In order to understand the requirements of users, we have analyzed existing literature characterizing the CQA domain and conducted in-depth interviews with the Qatar Living admin and target users.

5.3.1 Domain Characterization

To characterize the domain of question answering (QA) forums, we analyzed existing literature in the areas of human-computer interaction and computer supported collaborative work, focusing on what types of questions are asked [51, 64, 88], who answers and why [41, 88] and what are the predictors for answer quality [51].

Subjective nature of questions: Researchers have found that there are more subjective and opinion-based questions than factual questions [56, 64, 88]. Morris et al. surveyed QA users and found that only 17% of the questions they asked were seeking factual information, while the most common categories of questions were requests for recommendation (29%) and opinions (22%) [88]. Similar results were found for a social-question-answering system, with 64.7% of the queries were found to be subjective [64]. Due to the nature of these questions, e.g., 'best Italian restaurant in Doha', often any particular subjective answer, for example 'I like

Di Capri Ristorante', may not satisfy the information needs, therefore the user interface should effectively support browsing various answers from multiple related threads.

Variability in answer quality: Previous work also analyzed the characteristics of good answers. Harper et al. conducted a controlled field study to analyze different predictors of answer quality across several QA sites [51]. They found that while QA site like Yahoo! Answers provide lots of high-quality answers, users should also expect substantial variability in the quality of individual answers. To address this issue, it may be useful to apply an automatic approach to identifying high quality answers and help users to navigate through these answers.

Slower response: Some researchers have explored the factors affecting answer quality and response time on QA sites. Raban and Harper identified both intrinsic factors such as perceived ownership of information and gratitude and extrinsic factors such as reputation systems that motivate CQA users to answer questions [104]. However, even when motivated people are available to answer, their response times tend to be long [65, 88]. For example, the average time to receive a response to a question posted to Microsoft's Live QA site was 2 hours and 52 minutes [65].

Interviews: In addition to analyzing existing literature, we also conducted two semi-structured interviews and a number of follow-up interviews with our collaborator at Qatar Living. We also interviewed five users in our early design process, who regularly visit Qatar Living forum. The goal was to understand more specific needs and requirements for the type of forum that Qatar Living represents to directly inform our design process.

Many naive users: Qatar Living is one of the most popular sites in Qatar, with over 550,000 visitors per month and over 19 million page views a month from Qatar. The Qatar Living forum is very popular in Qatar, especially among the expatriates. It is actively visited by hundreds of users everyday, who mainly try to fulfill their information needs in their topics of interest. However, a large portion of the forum users are naive and they are not proficient with sophisticated user interfaces. Therefore, an important design goal was to make the interface simple and intuitive. In addition to naive and non-expert users, there is a dedicated small group of forum moderators having higher level expertise about the topics. These users actively browse the new questions posted in the forum and try to answer them depending on their expertise. While we have mainly focused on supporting the former group of users, we argue that moderators can also benefit from our text analytics and visual interface for their tasks.

Searching for previous questions rather than asking new ones: Our collaborator and Qatar Living users pointed out that usually the readers try to get their questions answered as quickly as possible. So, they often prefer to use the search feature within the forum to find similar questions to their current question, rather than posting their questions and waiting for answers.

However, they have difficulty in exploring threads of comments associated with similar questions, due to a large volume of comments they need to read, which is time consuming and cumbersome using the existing search interface. This suggests a pressing need for improving the search interface for the forum to enhance the user's ability to find good answers.

Difficulty in finding good answers: Like many other CQA sites, Qatar Living forum contents are often noisy and redundant. Users tend to use very informal language, often writing very long stories with small pieces of relevant text only. Due to noisy and redundant content, the question threads can become longer with only a few relevant answers. As a result, searching for relevant answers often leads to the information overload problem. To make matters worse, although there is an upvoting/downvoting feature, most users either do not know how to use this feature or they do not bother to do it. This was also confirmed by the users who were interviewed during the early design. Based on this observation, our collaborators agreed that an automatic comment classifier that is reasonably accurate can be effective in identifying good answers. More importantly, the interface should facilitate the user to find the good answers, which may be scattered among the large amount of comments from multiple different threads.

In summary, while designing the system for supporting the information seeking tasks in forums like Qatar Living (that are typically unmoderated, contain near duplicate questions and lot of noisy comments), we should consider following user requirements: 1) The interface should effectively support the user in identifying multiple good answers from related question threads. 2) To address the variability in the answer quality, a classifier should be introduced to identify useful comments. 3) The interface should introduce interactive visualization components to enhance the user's ability to find good answers from large volume of comments. 4) To support users having lower visualization expertise, the interface should be simple and intuitive.

5.3.2 Data and Task Abstractions

Tasks: In our conversations with the Qatar Living admin, we learned several use-cases and tasks of the forum users. We analyzed these tasks according to a visualization task typology [92] in order to inform our design. At the high level, users are primarily interested in seeking information with the goal to *discover* new information or knowledge. At this level, the user may ask questions like "Which is the best bank in Qatar?", or " Where can I find a good Chinese restaurant in Qatar?". Once the user is presented with some related questions to her new questions, the next level task is to *search* for the most related questions of interest by *browsing* the list of questions presented to her. When they find the most related questions from the list, next they focus on *identifying*, *comparing*, and *summarizing* the most useful answers to her original question.

Data: Based on our user requirements and tasks analysis, we derive how the data should be abstracted for visualizing to the user. As illustrated in Figure 5.1, an example dataset consists of a question asked by a user with the set of related questions found by the system. We encode the relatedness of a question to the new question by a rank value (ordinal). Each related question is also followed by a set of comments that tried to answer that question. For each of these comments, we derive the *goodness* score provided by a classifier with respect to its related question and represent as a normalized quantitative value between [0.0,1.0], by passing the score through a sigmoid function. We also assign each comment into one of six equally sized bins depending on its classification score to help the user understand how relevant a particular comment is. Based on this binning, we also compute the distribution of comments for each related question thread by counting how many comments fall into a particular bin. We compute this distribution, because it can effectively convey to the user how many comments are useful among all the comments of a question.

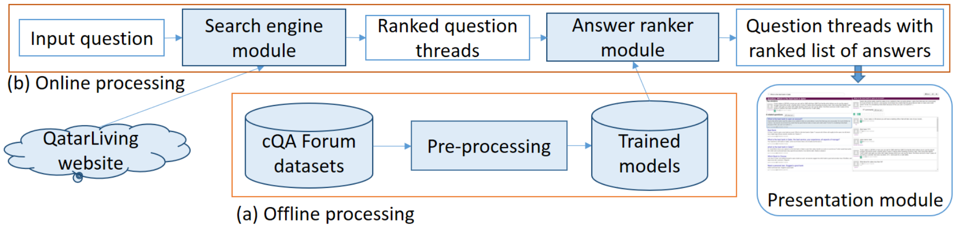


Figure 5.2: Overview of our interactive system for supporting community question answering.

5.4 System Overview

Figure 5.2 presents an overview of our system, which is organized in two parts. In the offline step (Figure 5.2a), we pre-process the datasets and we train a comment classifier. In the online regime (Figure 5.2b), the user enters a question as input, and the system performs three steps on the fly: retrieving the top *n* related question threads, ranking all the answers, and visualizing the results. We briefly discuss these steps below.

5.4.1 Offline Processing

To build the system, we used a dump of the Qatar Living forum from March 2016, and we performed several pre-processing steps including the conversion of the XML dump to JSON format that our interface can process. This dump contains $202,304$ conversations and $2,043,022$ comments. On an average, each conversation consists of 10.21 comments.

We also used the datasets on CQA from SemEval-2016 Task 3 (subtask A) [94], where the comments in the threads are manually annotated with `good` vs. `bad` labels, indicating how well the comments answer the question in the thread. Using this dataset, we extracted a collection of features and we trained a Support Vector Machine (SVM)-based comment classifier that scores each comment in a thread regarding its *goodness*.

5.4.2 Online Processing

When a user types a new question q, the system performs the following three steps on the fly: (*i*) *Retrieve related questions*, where Google local search is invoked to retrieve the top-n question threads in the Qatar Living forum that are most similar to q, $\{q_i\}_{i=1}^{n}$; (*ii*) *Rank the answers*, where all the comments from these top-n question threads are ranked based on their relevance with respect to q. (*iii*) *Visualize the results*, where the presentation module takes the related questions' threads together with the ranked lists of comments and the overall best selected answer, and presents them to the user.

5.5 Text Analytics

The answer ranker module computes the relevance score of a comment c in a question thread q_i with respect to the new question q by combining two scores: (*i*) $\sigma(q, q_i)$, the similarity of q_i to q; and (*ii*) $\gamma(c, q_i)$, the goodness score for c with respect to q_i. Formally, the relevance score $\rho(c, q, q_i)$ is computed by:

$$\rho(c,q,q_i) = \sigma(q,q_i) \times \gamma(c,q_i) \tag{5.1}$$

We use the inverse rank in the list returned by the Google search engine as $\sigma(q, q_i)$, and $\gamma(c, q_i)$ is computed by a comment classifier, indicating how well comment c answers q_i. The resulting score is used to rank all the comments from the retrieved question threads to obtain the best overall answer to the input question q. Intuitively, if a comment is a good comment with respect to the thread question, and the thread question is related to the new question, then the comment is likely to be a relevant answer to the new question.[4] The core NLP component of this architecture is the comment classifier, which is briefly described below.

Comment Classifier Given a question q and a list of comments associated with it $\{c_i\}_{i=1}^{m}$, the task of the classifier is to assign a relevance score to each of the comments according to their *goodness* at answering the question. This very problem

[4] As discussed in the SemEval-2016 Task 3 description paper [94], this is a very simple way to obtain good results for the general task of ranking answers for new questions.

was set at SemEval-2016 Task 3 [94], subtask A. We trained an SVM classifier on that dataset to distinguish between `good` and `bad` comments.[5] The dataset is split into training, development and test sets, with 2,669, 500, and 700 questions, and 17,900, 2,440, and 7,000 answers, respectively. The kernel function in our SVM is a linear combination of four functions: two linear kernels over numeric features and embeddings, and two tree kernels over shallow syntactic trees.

Numeric Features These features are inspired by [12]. They include three types of information: (*i*) a variety of textual similarity measures computed between the question and the comment; (*ii*) several Boolean features capturing the presence of URLs, emails, positive/negative words, acknowledgments, forum categories, long words, etc.; (*iii*) a set of global features modeling dialogue and user interactions in the thread.

Embedding Features Higher level abstract features learned automatically by deep neural networks have proved to be quite beneficial for learning semantic similarity between two texts [39, 103, 115, 117]. We learn embeddings for questions and answers by training a convolutional neural network (CNN) on the comment classification task following the approach of [115]. Specifically, the input to the CNN is formed by two matrices containing word embeddings for the question and for the answer, respectively. The CNN performs a *convolution* and a *max-pooling* operations on the word embeddings and on the convoluted feature maps, respectively, to produce the question embedding q_E and the answer embedding c_E. These embeddings are then combined to produce a similarity value using a similarity matrix. The similarity and the embeddings along with other additional similarity features are then passed through a hidden layer and next to the output layer for classification. q_E and c_E are learned by backpropagating the (cross entropy) errors from the output layer. q_E and c_E vectors are finally concatenated and used as features in our SVM model.

Tree kernels Tree kernels provide effective ways to learn by comparing syntactic structures of two texts in the SVM framework, which has been shown to give state-of-the-art results in CQA [98]. First, we produce shallow syntactic trees for

[5] The conversations in the SemEval dataset were written in the same language (English) as the material on the QatarLiving forum site.

the question and for the comment using the Stanford parser. Following [114], we link the two trees by connecting nodes such as NP, PP, VP, when there is at least one lexical overlap between the corresponding phrases of the trees, and we mark those links using a specific tag. The kernel function K is defined as: $K((q_1,q_2),(c_1,c_2)) = TK(q_1,c_1)+TK(q_2,c_2)$, where $TK(q,c)$ is a tree kernel function operating over a pair of question (q) and comment (c) trees.[6]

Classification Performance Our comment classifier was evaluated on the SemEval-2016 test set with the official scorer, obtaining the following results: MAP=77.66, AvgRec=88.05, MRR=84.93, F_1=66.16, Acc=75.54. Compared to the participant systems at SemEval-2016, our system scores in second position regarding the official MAP evaluation metric (-1.5 points below the best). In contrast, our system achieves better F_1 (+1.8) and better Accuracy (+0.4) than the top system. For a full description of the results from SemEval-2016, see [94].

5.6 CQAVis Design

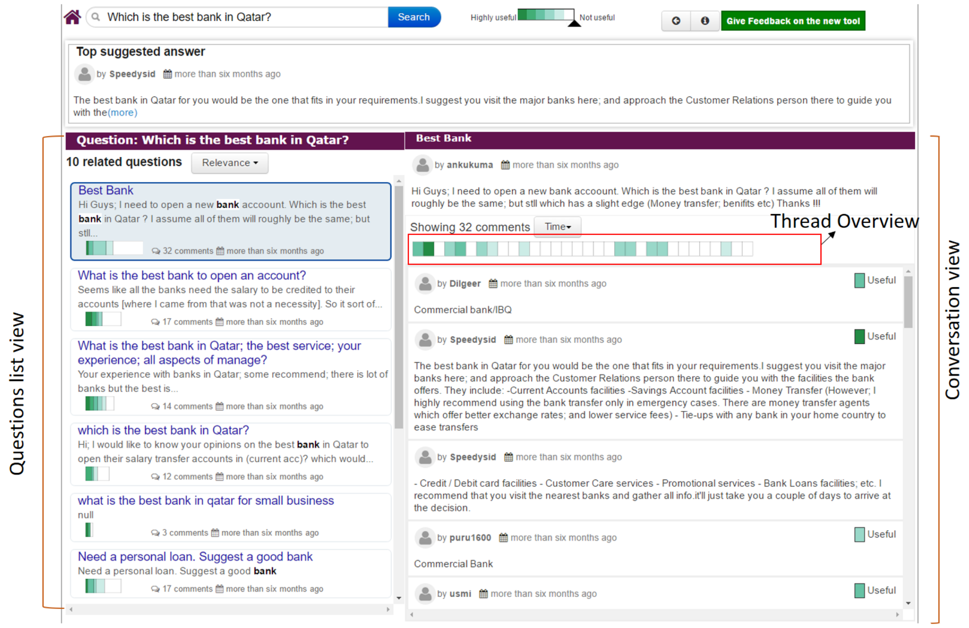
In order to explore a large number of design choices, we carried out an iterative design process, starting from early mockups and prototypes using paper and Powerpoint. We then developed a mid-level prototype which works on a small CQA annotated corpus [94], where the comments are annotated with `good` vs. `bad` labels by human experts. Finally, we developed a fully functional system and deployed within a real CQA site. Throughout the design process, we performed formative evaluations [78] to identify potential usability issues and to iteratively refine the prototype. We now present the final design of the CQAVis interface[7] [8], along with justifications for the key design decisions based on our user requirements analysis and the InfoVis literature.

The design of our visual interface was influenced by our previously developed interfaces for exploring a set of conversations (i.e., ConVis and MultiConVis); however, in this new design we took into account specific features of CQA data and tasks. A high level design decision for the interface was to follow an overview+detail approach, where the overview represents the *question list view*

[6] We use Partial Tree Kernel and Syntactic Tree Kernel [26, 89] to instantiate TK.

[7] A live demo of CQAVis is available at iyas.qcri.org

[8] A video demonstration of CQAVis is available here https://goo.gl/IM3Gez



Top suggested answer

by **Speedysid** more than six months ago

The best bank in Qatar for you would be the one that fits in your requirements.I suggest you visit the major banks here; and approach the Customer Relations person there to guide you with the(more)

Question: Which is the best bank in Qatar?

10 related questions

Best Bank
Hi Guys; I need to open a new **bank** accooount. Which is the best **bank** in Qatar ? I assume all of them will roughly be the same; but stll...
32 comments more than six months ago

What is the best bank to open an account?
Seems like all the banks need the salary to be credited to their accounts [where I came from that was not a necessity]. So it sort of...
17 comments more than six months ago

What is the best bank in Qatar; the best service; your experience; all aspects of manage?
Your experience with banks in Qatar; some recommend; there is lot of banks but the best is...
14 comments more than six months ago

which is the best bank in Qatar?
Hi; I would like to know your opinions on the best **bank** in Qatar to open their salary transfer accounts in (current acc)? which would...
12 comments more than six months ago

what is the best bank in qatar for small business
null
3 comments more than six months ago

Need a personal loan. Suggest a good bank
Need a personal loan. Suggest a good **bank**
17 comments more than six months ago

Best Bank

by **ankukuma** more than six months ago

Hi Guys; I need to open a new bank accooount. Which is the best bank in Qatar ? I assume all of them will roughly be the same; but stll which has a slight edge (Money transfer; benifits etc) Thanks !!!

Showing 32 comments

by **Dilgeer** more than six months ago — Useful

Commercial bank/IBQ

by **Speedysid** more than six months ago — Useful

The best bank in Qatar for you would be the one that fits in your requirements.I suggest you visit the major banks here; and approach the Customer Relations person there to guide you with the facilities the bank offers. They include: -Current Accounts facilities -Savings Account facilities - Money Transfer (However; I highly recommend using the bank transfer only in emergency cases. There are money transfer agents which offer better exchange rates; and lower service fees) - Tie-ups with any bank in your home country to ease transfers

by **Speedysid** more than six months ago

- Credit / Debit card facilities - Customer Care services - Promotional services - Bank Loans facilities; etc. I recommend that you visit the nearest banks and gather all info.it'll just take you a couple of days to arrive at the decision.

by **puru1600** more than six months ago — Useful

Commercial Bank

by **usmi** more than six months ago — Useful

Figure 5.3: A screenshot of the interface showing the top answer and related questions for a user's question. As the user selects a related question marked by the blue rectangular boundary, the interface shows the corresponding thread in the conversation view.

showing the top-most relevant questions to the user's question; and the detail view (i.e., *conversation view*) showing the question followed by the answers for a particular question thread (see Figure 5.3). We made this choice because this allows users to browse comments concerning a specific question, while still having the context of the other related questions, and also because this approach has been found to be more effective for text comprehension tasks than other approaches such as zooming and focus+context [24].

Questions list view After the system finds the related questions to the user's question, it presents an overview of the ranked list of relevant questions in a scrollable list view (see Figure 5.3, left). Each item within the *questions list view* represents a question thread, showing a set of metadata i.e., the original question, the posting date, the total number of comments, as well as a stacked bar with the distribution of useful comments. Since we are representing an ordered sequence of values, we used a set of six sequential colors by varying monotonically on the green color channel ranging from dark green (highly useful) to white (not useful). In this way, the user can quickly get a sense of which threads seem to be more relevant and which threads may contain the most useful answers.

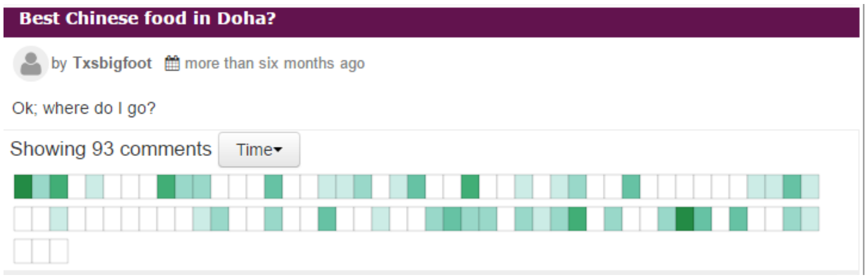
Notice that encoding the distribution of useful comments using colors within a stacked bar is analogous to how the sentiment distribution was represented within a conversation (in MultiConVis) and within a comment (in ConVis). However, here instead of diverging color we used sequential colors, as the normalized usefulness score ranges from 0 (not useful) to useful (1).

The questions are ordered by their relevance rank by default, but the user can change this order by selecting criteria from the popup menu 'Order by'. For instance, she can order the question threads based on the number of useful answers within each of these threads.

Another important feature of the interface is that at any time the user can filter out comments with low usefulness score by using the slider of the widget (containing sequence of colored rectangles) at the top, as shown in Figure 5.3. In this way, the user can quickly narrow down the set of less useful comments of different question threads and focus on the ones that are potentially good answers to her question.

Note that at the top of the question list view, the interface also shows the comment that has received the best score with respect to the new question ("Top suggested answer"). This feature was designed to support the user in finding a very good answer to her question immediately, without having to open any question thread and then navigating to answers within that thread. This was motivated by the user requirements analysis, from which we learned that users would like to find some very good answers quickly, therefore showing the top ranked answer right away could be very useful.

Conversation view When the user selects a particular question thread from the list, the system presents the corresponding thread in the conversation view, as shown in Figure 5.3. Again, we followed an overview+detail approach, where at the top we show a visual overview of the entire thread along with the question, fol-



Best Chinese food in Doha?

by Txsbigfoot more than six months ago

Ok; where do I go?

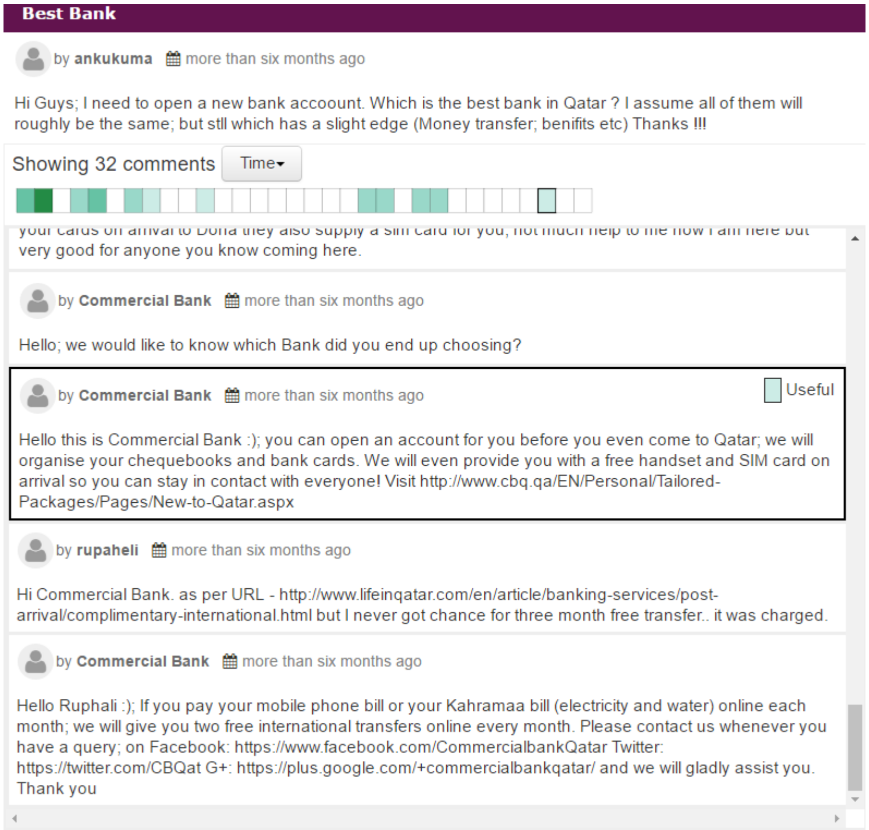
Showing 93 comments

Figure 5.4: An example of a thread overview that splits a large number of comments into multiple rows to deal with horizontal space constraints.

lowed by a detail view containing the list of comments. Here, the thread overview visually encodes the comments using a sequence of rectangles from left to right, where each rectangle represents a comment. The color within each rectangle encodes the classification score of the comment represented by that rectangle. If the horizontal space is not sufficient for showing all the comments, then it shows the rectangles in multiple rows as shown in Figure 5.4. In this way, the thread overview visualization can scale with hundreds of comments, which is sufficient for a typical CQA forum conversation.

From the thread overview, the user can quickly notice which comments are more useful and then immediately navigate to a particular comment by clicking on the rectangle representing that comment (see Figure 5.5). Note that the two views are coordinated, i.e., hovering on a rectangle in the thread overview highlights the corresponding comment in the detailed view (by scrolling if needed) and vice-versa, thus providing the user a sense of where s/he is in the current thread and what to expect next. Finally, the user can reorder the comments of a thread based on their classification score to quickly go through the most useful answers.

Throughout the design of CQAVis, an important goal was to make the interface simple and intuitive for the naive users, who constitute a large portion of users of Qatar Living and similar forums. To achieve this goal we focused on using visualization metaphors that are common and easily understood (e.g., bar graph based visualization and sequence of rectangles) and a small set of simple, low cost interactions [77] that can be easily triggered and reversed without requiring much cognitive overload.



Best Bank

by **ankukuma** more than six months ago

Hi Guys; I need to open a new bank accoount. Which is the best bank in Qatar ? I assume all of them will roughly be the same; but stll which has a slight edge (Money transfer; benifits etc) Thanks !!!

Showing 32 comments

your cards on arrival to Doha they also supply a sim card for you, not much help to me now I am here but very good for anyone you know coming here.

by **Commercial Bank** more than six months ago

Hello; we would like to know which Bank did you end up choosing?

by **Commercial Bank** more than six months ago

Hello this is Commercial Bank :); you can open an account for you before you even come to Qatar; we will organise your chequebooks and bank cards. We will even provide you with a free handset and SIM card on arrival so you can stay in contact with everyone! Visit http://www.cbq.qa/EN/Personal/Tailored-Packages/Pages/New-to-Qatar.aspx

by **rupaheli** more than six months ago

Hi Commercial Bank. as per URL - http://www.lifeinqatar.com/en/article/banking-services/post-arrival/complimentary-international.html but I never got chance for three month free transfer.. it was charged.

by **Commercial Bank** more than six months ago

Hello Ruphali :); If you pay your mobile phone bill or your Kahramaa bill (electricity and water) online each month; we will give you two free international transfers online every month. Please contact us whenever you have a query; on Facebook: https://www.facebook.com/CommercialbankQatar Twitter: https://twitter.com/CBQat G+: https://plus.google.com/+commercialbankqatar/ and we will gladly assist you. Thank you

Figure 5.5: When the user clicks on a rectangle in the thread overview representing a comment, the interface scrolls to that comment (marked by black color) in the conversation view.

As one could easily notice that the design of the conversation view in CQAVis was strongly influenced by ConVis. For example, both interfaces visually encode the thread overview using a sequence of colored bars to represent comments. Moreover, in both interfaces, the thread overview and detailed view are coordinated, so that any interaction in one view reflects in the other view. Yet there are a few notable differences between the two interfaces. First, recall that in ConVis, the topics and authors of the conversation were connected to the comments in the thread overview via explicit links. In contrast, in CQAVis, we created a compact representation of the thread overview with a sequence of rectangles positioned horizontally and removed the representation of topics and authors along the thread view. We did this because our data and task abstractions obviated the necessity of presenting the topics and authors and also this helps us simplifying the interface considerably. Second, unlike what was done in ConVis design, we did not encode the comment length using the height/width of the rectangle representing a comment. The primary reason for removing this feature is that in a pilot study we found that either users did not understand what this encoding means and even if they understand this encoding they did not find it to be useful (see later in Section 5.8.3).

5.7 Implementation

The system is implemented as a Java Web application and runs on an Apache Tomcat Server. The back-end of the system is developed using Java. The presentation module, on the other hand, is implemented in JavaScript using the D3 and JQuery libraries. It should be noted while implementing the presentation module, we were able to reuse parts of the implementations from our previously developed interfaces, making it faster to design the fully functional prototype.

Furthermore, our system was designed to be sufficiently fast to respond in real time to the user's actions. A key factor for the efficiency is the fact that we precomputed and stored the goodness scores for all the comments in all the question-threads from the static snapshot of the Qatar Living database. In this way, at running time there is no need to classify the comments of the already stored question-comment threads.

5.8 Web-based User Study

To better understand the potential utility of our approach in real world scenarios we undertook a large-scale, Web-based study. The primary aim of our study was to empirically examine how real users would use CQAVis and what their impressions would be to such a visual search interface. The main research questions were: 1) What are the possible benefits and limitations of the CQAVis interface in supporting the task of information seeking? 2) When we compare CQAVis with a typical interface for forum search, as instantiated by Qatar Living forum, is there any difference in subjective measures?

5.8.1 Methodology

While a lab-based user study would allow us to have more control over the users and tasks, realism would be largely lost [18]. Therefore, we decided to run the study in a Web-based environment to enhance its ecological validity, since participants can then work in their own settings performing their own tasks [78]. It also gives us the advantage of collecting interaction logs from a large number of users to get deeper insights that are arguably more generalizable than a lab study.

5.8.2 Study Setup and Procedure

In order to run the user study, we discussed with our collaborators at Qatar Living, who agreed to incorporate our web-based tool as a beta version of the forum site. Our system was deployed at a server and then a Web-link of the system was made available on the forum search page for the real users of the Qatar Living forum. To avoid compatibility issues, we tested our interface on the Web browser versions of Mozilla Firefox, Apple Safari, and Google Chrome to ensure that we could support a wide range of participants.

Participants were guided through three main steps of the study: 1) *Introduction*: In the home page, some background information and example queries were provided to get started, along with an invitation to use the interface, as shown in Figure C.1. The page also contained a short video (duration of 78 seconds) to demonstrate the main features of the interface. 2) *Interaction*: The main part of the study was the interaction with CQAVis. Here, users were not asked to complete any specific task; instead they could perform their own set of information seeking tasks. 3) *Feedback*: Participants were free to fill out a post-study questionnaire at any time during their interaction by clicking on the 'Give feedback on the new tool' button at the top. The form also allowed them to provide free-form comments and suggestions, as shown in Figure C.2. Finally, the questionnaire sought voluntary information about the age, gender, and Web experience of participants (see Figure C.3) [9]. Throughout the sessions, we logged interface actions along with their timestamps in a completely non-intrusive way to better understand the usage patterns of the CQAVis tool.

[9] The study materials for the user study can be found in Appendix C

5.8.3 Pilot Study

Before making the beta version publicly available and running the online study, all study aspects, including instructions and setup, went through several iterations of a pilot study. We ran this pilot study in a lab-based setting with six participants, where we collected the data in the form of questionnaires, interviews, and observations.

The pilot study helped us in refining both the study procedure and the prototype. For example, the pilot study suggested that background questions should be asked at the end of the study instead of at the beginning, because participants wanted to immediately explore the system without requiring to fill out the questionnaire. We also modified the types of questions being asked (e.g., we provided fewer open-ended questions). The pilot study also led us to simplify the interface, by avoiding visually encoding less useful data, such as the comment length, which was originally encoded using the width of the rectangle encoding the comment itself.

5.8.4 Participants

Our online study attracted 768 participants over a period of 18 weeks. The users were recruited through the beta version link of Qatar Living, as well as through publicizing in the online social networks Facebook and Twitter, and via mailing lists.

Those participants who chose to provide their background information (5.3% of total participants) held a variety of occupations, including students, expatriates working as engineers, architects and consultants, researchers and professors in universities etc. The majority of participants were young (85% of them were below 45). Among those who indicated their gender information, 65% were male participants. In general, most of the participants were quite familiar with using the Web, with 72% of them indicating that they visit the Web quite frequently (several times a day). However, when it comes to uses of online forums, the responses were mixed, ranging from rarely to very frequently with 37% users mentioning that they occasionally visit forums to their questions answered (i.e., several times per month).

5.8.5 Analysis of Results

We now present both our quantitative and qualitative analysis, as well as the results based on the data collected from the user logs and questionnaires.

Sessions and queries: During the study, we captured quantitative data regarding 1,122 queries from 768 users. A summary of the queries and sessions is provided in Table 5.1. From the table, we can see that based on the medians, a typical participant spent 142 seconds with the system and issued just 1 query per visit. The average session lengths are considerably larger, as some participants engaged with the system for much longer time periods.

	Min	**Median**	**Mean**	**St. Dev**	**Max**
Length of session (seconds)	1.79	142	416	666.45	3,327
Queries per session	0	1	1.47	1.54	16
Query length (characters)	1	20	22.81	14.4	200

Table 5.1: Overview of user study sessions and queries.

Question Type	**Percent**	**Example**
Recommendation	21.82	Where can I find Italian restaurants in Doha ?
Opinion	18.51	Is QnB a good bank?
Factual knowledge	31.21	When does Ramadan start in 2016?
Rhetorical	0.55	How it is to live in Qatar?
Invitation	0.83	need tennis partner
Others	27.07	razor racing car

Table 5.2: Breakdown of query types along with examples

We categorized the questions asked by the participants by following [88] to understand the nature of information needs that were prevalent among our target users. The distribution of question types is shown in Table 5.2. Here, both *opinion* and *recommendation* questions are subjective in nature; *opinion* questions ask for a rating of a specific item whereas *recommendation* questions ask for open-ended requests and suggestions. In contrast, *factual* questions expect objective answers.

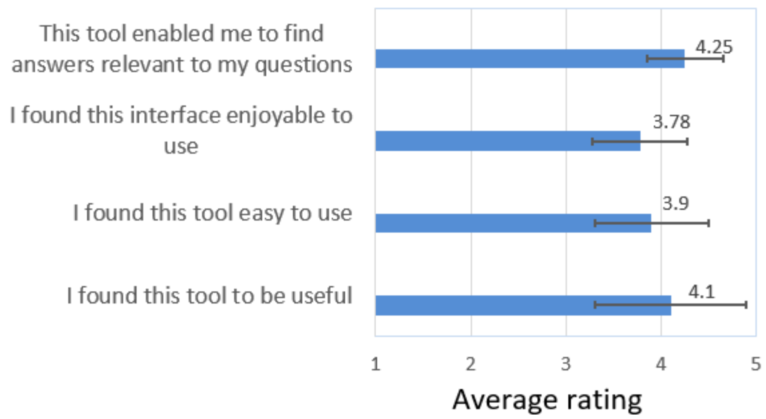


Figure 5.6: Average rating of interfaces by the participants on four different measures. Longer bars indicate better rating.

Rhetorical questions are intended to promote discussions as opposed to eliciting specific answers. An *invitation* asks for attending an event. Finally, the *other* category consists of queries, that do not fall into any of the previous categories.

The distribution of questions was similar to what has been found in the existing literature [88], with subjective questions (i.e., *opinion* and *recommendation*) being strongly prevalent among participants (41%). This justifies the rationale for tailoring our interface to deal with subjective questions which may require the user to read many useful comments to get the answers from various perspectives.

Subjective ratings: After interacting with the interface the user could chose to provide feedback by clicking on the Feedback button. 56 users chose to provide feedback on the tool. In the feedback form, participants rated four different measures on a standard 5 point Likert scale: 1) ‘I found this tool to be useful’; 2) ‘I found this tool easy to use’; 3) ‘I found this interface enjoyable to use’; 4) ‘This tool enabled me to find answers relevant to my questions’.

The results of these questionnaires are presented in Figure 5.6. From the Figure, we can readily see that the majority of the responses were dominated by positive ratings. In particular, most users agreed that the tool is useful and it enabled them to find answers relevant to their questions.

Preference: In the questionnaire, participants were also asked if they would prefer this tool over their regular forum search tool. 68.75% of participants indicated a preference for CQAVis, with only a small fraction of them (6.25%) choosing the regular one. 25% indicated that they were indifferent between the two interfaces.

Interaction patterns: In addition to questionnaires, we analyzed the log data to get insights into the interaction patterns of users. Figure 5.7 shows the percentage of users who used each interactive features of the interface at least once. As expected, almost all the participants typed at least one query during the interaction. Similarly, most of them hovered and clicked on conversations in the question list view. When interacting with the conversation view, over 54% of the users hovered on the thread view and 39.4% clicked on rectangles in the thread view representing comments. This result is rather encouraging, because despite being completely new visualization features, they were used by a large portion of users. Finally, sorting and filtering comments were used by a smaller number of users (12% and 9% respectively). A possible explanation is that many participants did not notice these features, while interacting with the interface. Another reason could be that users were able to fulfill their information needs with other interactive features.

Qualitative data analysis: I analyzed the free-from text provided by 39 participants to gain insight into the users' experience with the interface. In order to make sense of these comments and suggestions, I carried out a bottom-up coding approach: first, read all the free-form texts to gain an overview of the participants' feedback; second, find common themes and associate codes accordingly; and finally, categorize the themes into the main types of feedback. Once I had identified the common themes, I read all the free-form texts again in a second pass to analyze how appropriate their associated codes were and to count frequency of occurrences. In this second pass no new themes emerged. All the themes resulting from my analysis are described below:

General feedback (21 participants): From this analysis, it was found that the feedback towards CQAVis was positive (67%), but there was also some negative (24%) and neutral (9%) feedback. More specifically, those who were positive towards the CQAVis interface expressed that the interface was simple and easy to use, which was an important design goal. According to participant P20, *"The design of this tool is very simple and easy to use. I am impressed with the tool's accessibility and how intuitive it was..."*. Also, some participants' perceived speed

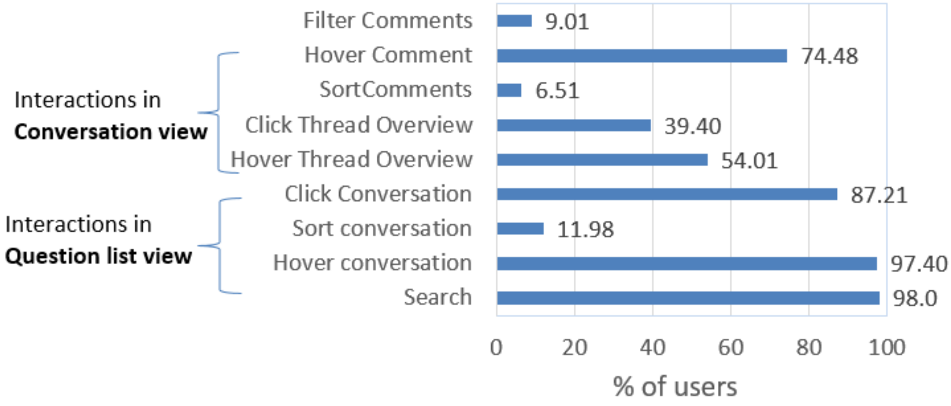


Figure 5.7: Interface features used by the participants.

of task completion was enhanced by the interface as pointed out by P29 *"Quick and reliable"*.

A number of participants thought that the system was able to satisfy their information needs effectively. For instance, P13 mentioned that *"It gave me the answer I was looking for in a straightforward way, which is what you want from a search tool. No need to scroll through lots and lots of Google pages..."*. Similarly, P1 liked the idea of finding high quality comments from similar question, *"I like that you can get similar questions and their corresponding high quality answers immediately, without having to read all the comments"*. Some people also compared their positive search experience with the traditional Qatar Living search tool: *"Qatar Living is difficult to search but with this tool it gets much easier"* [P22].

Those who were critical about the interface mentioned that the text analytics techniques need to be more accurate '*"Need more accuracy for the result"* [P6]. Some people also questioned the reliability of the comments and suggested a way to filter out spam comments: *"It could be made better by filtering out spam comments. Some of the information has no actual basis...*" [P21]. Also, one participant suggested that for time sensitive questions, the system should consider the timestamp of answers for ranking, *"I asked: when does Ramadan start? But the top answer was actually posted few years ago*" [P18].

Reactions to interface features (16 participants): There were also recurring comments on particular features of the interface. For instance, some participants liked the idea of having the question view and the conversation view side-by-side: *"It is nice to have the questions and answers load quickly side to side without having to open many tabs in the browser"* [P22].

Several people were impressed by the visual thread overview and the color coding to represent the usefulness of a comment. *"I liked the color coding idea of the comments in the tool. It is very useful"* [P24]. However, learning this feature requires sometimes for one participant *"At the beginning it was not clear what the colored squares are..."* [P8]. One possible explanation is that very few participants (2%) actually watched the video tutorial provided on the introduction page.

Suggestions for improvement (4 participants): A few participants felt that the user interface needs some improvements in general. There were also few specific suggestions about the components, for instance, the size of the slider at the top needs to be increased, so that it can be easily noticed [P10] and the interface should show the textual label 'not useful' explicitly for the comments that fall into the least useful bin [P34].

5.9 Discussion

We now discuss the implications of our results and generalizable lessons we have learned from the design study.

5.9.1 Summary of Findings

Based on our analysis of the results, we now revisit our research questions mentioned at the beginning of the previous section. The first question was what are the possible benefits and limitations of CQAVis in supporting information seeking tasks. From the feedback data, the majority of participants who filled up the questionnaires found the interface to be useful and felt that it enabled them to find the relevant answers to their questions. Also, the qualitative feedback from participants suggests that their overall impression was quite positive.

With regards to the second research question, when the participants were asked to indicate a preference, the majority of them chose CQAVis over the traditional forum search tool. However, recall that the questionnaire data was filled up only by a fraction of participants, which may have introduced a positive bias. While this prevents us from making strong claims from the questionnaire data alone, we complement the analysis with qualitative observation based on the free-form comments as well as from the interaction log data to get a deeper understanding of both positive and negative aspects of the interface features and their usage. In particular, we have emphasized critiques of the interface and suggestions for improvements that we got from those users, because this neutral or negative feedback is much more likely to generalize to the whole target population (as they are coming from a subset of the population with a likely positive bias).

We should also consider that the interaction log data was analyzed over all the participants, thus arguably reflects overall usage patterns. In particular, the log data reveals that not all the interface features were equally used. While some of the new interface features such as the thread overview were used by a fair number of participants, still some participants did not use them. A possible explanation is that some participants might prefer the traditional way of scrolling through the comments of the thread, while still having a situational awareness by looking at the thread view. In the future, capturing users' eye gaze data could shed more light on this aspect.

5.9.2 Lessons Learned

We now reflect upon on our design and evaluation of the CQAVis interface to summarize the lessons learned that can arguably be generalizable to other conversational domains.

Design

Most target users in our domain did not have enough familiarity with complex interactive visualization. To support such users, we have focused on following design principles which can be applicable to other domains where users have similar expertise level.

Less is more: In our early prototypes, we considered some advanced features, such as visually encoding additional data (e.g., comment length) and more complex interactions (e.g., navigate through the related-question-graph) with an aim to better support users. However, the feedback from users throughout the pilot studies led us to simplify the interface iteratively, eliminating such kinds of interactive visualization features. Based on our experience, we suggest that when designing for similar populations in the domain of conversations, the designer should simplify interface features iteratively to retain features that are not only useful but also simple and intuitive.

Enhance learnability: We found that in our study users do not tend to spend time reading the instructions or watch the video tutorial to learn the new interface. Therefore, the interface should enhance the learnability by providing self-explanatory components by adding more textual labels and tooltips.

Introduce familiar visualizations: During the prototyping stage, we realized that novice users in the Qatar Living forums often find it difficult to understand complex visualizations. Therefore, the interface should use the visualization components that are easily understood by most people, such as bar graph based visualization.

Evaluation

While we argue that the web-based online study enhanced the ecological validity by evaluating with real forum readers performing their real tasks, it also posed several challenges. For instance, it was difficult to collect sufficient amount of quantitative and qualitative feedback from a large number of participants. While it is common to collect users' background and demographic information in the form of a pre-study questionnaire, in a pilot study we found that participants were reluctant to fill out the questionnaire. Therefore, the questionnaire was included in the feedback form that the user could fill out, after they had interacted with the interface.

Even then the challenge was how to get feedback from a large amount of participants who have interacted with the interface. While a button for providing feedback was available at the top of the interface, some participants did not even notice it. To further enhance the likelihood of obtaining some feedback, we introduced a pop-up screen that would appear reminding the user to submit the feedback when they move their cursor at the top of the screen. To provide further incentives to the user, the message mentioned that the participant will be entered into a lottery of winning 50 QAR gift cards. While all of the above techniques helped us to receive more feedback, we call for more research on how to get a rich number of feedback from a large number of participants in a Web-based study.

5.10 Summary

In this chapter, we presented an interactive system for exploring CQA forums as an example of how our visual text analytic solutions can be simplified and tailored to specific domain problems. The resulting system, CQAVis, supports users to find good answers to a newly-posed question, by combining a novel set of NLP and InfoVis techniques, informed by an understanding of the user requirements in the domain of CQA. The underlying NLP methods automatically retrieve and rank a set of comments with respect to the new question, (*i*) by selecting a set of question threads that are relevant to the user question, (*i*) by assigning a goodness score to the comments within these threads, and (*iii*) by measuring the similarity between the new question and the thread questions. The visual interface, which was simplified and tailored from the MultiConVis interface, helps users in rapidly navigating through the useful comments, even if they are scattered around multiple different threads.

Our large-scale Web study underlines the potential for tightly integrating NLP and InfoVis, offering the users a new way of information seeking in CQA forums. An important finding from the study is that although a large portion of the user population did not have visualization expertise, the primary interactive visualization features were still widely used by participants. This suggests that by careful consideration of the target user characteristics and by iteratively simplifying the visual encodings of the interface, it is possible to tailor a visual text analytic system to a target population with possibly low visualization literacy - not for just those who have strong visualization expertise. It also reveals important lessons for designing and studying such systems for a user population with varying levels of expertise, which can arguably be generalizable for other conversational domains.

Chapter 6

Reflection and Conclusion

In this dissertation, we explored how to identify and leverage critical synergies at the intersection between natural language processing and information visualization to support users in exploring a large amount of online conversations more effectively. Our work was motivated by the challenges arising from the volume and complexity of conversational data and the shortcomings of existing approaches in dealing with such challenges.

To address the information overload problem, we explored a design space covering two dimensions: the scale of conversational data i.e., single vs. a set of conversations; and the underlying text analytic model i.e., static vs. human-in-the-loop. We developed and evaluated a set of systems, each addressing a different aspect of this design space, which are presented in Chapter 2, 3, and 4. Subsequently, we conducted a design study to demonstrate that our solutions can be successfully tailored to develop a new system for addressing specific domain problems, such as problems faced by users in a community question answering forum, as described in Chapter 5.

In this final chapter, we revisit our approach for designing visual text analytics systems (Section 6.1), reflect upon the research impact of these systems (Section 6.2), and indicate open research questions and directions for future work (Section 6.4). We conclude the dissertation with some closing remarks about visual text analytics for online conversations.

6.1 Reflection on the Design Approach

After presenting the design studies that focused on how to tightly integrate NLP and InfoVis for exploring online conversations, we take a step back to reflect on the wider context of designing visual text analytics. In particular, we critically analyze the role of the designer in the process of integration between NLP and InfoVis techniques to derive lessons that are broadly applicable for designing visual text analytics systems.

Within the visualization community, there has been a significant advancement in the field of design study methodologies [86, 91, 111], which provides guidelines on how to perform design activities and how to validate different stages of design. However, when developing a visual text analytics system, in addition to designing the visualizations, it is also necessary to devise a set of *text analysis methods* and validate the interpretability, accuracy, and usefulness of the output generated by these methods. Arguably, devising suitable text analysis methods is just as critical as visualization design in determining the effectiveness of a visual text analytics system. Therefore, within a user-centered design approach, a designer must consider what should be the most suitable text analysis methods and how to iteratively modify these methods when their output is not sufficiently interpretable, accurate, and useful.

Unfortunately, when designing visual text analytics systems, many researchers treat text analysis models as black boxes without considering whether they are the most suitable models for the target domain problem. For example, many text visualizations select the terms to be displayed based on their frequency [37], or their TF-IDF scores [124], even though more sophisticated techniques are available [52] that could select better descriptive keyphrases. By not considering the most suitable text analysis methods, often the system fails to effectively support the real world tasks.

Contrary to this trend, Chuang et al. opened the black-box and focused on how to devise an interpretable and trustworthy text analysis model by aligning the model with the tasks and user expertise in a particular problem domain [21]. From their experience of designing a dissertation browser, they distilled the following process-oriented guidelines: *align* the model with the tasks, user expertise, and visual encoding; *verify* the modeling decisions to assess how well they fit an analyst's goals; iteratively *modify* a model when its output is incorrect or incomplete; and *progressively disclose* data at multiple levels of abstractions, so that analysts can switch between different levels of abstractions to interpret and verify the model's output. Within their guidelines, they discussed *what* are the possible approaches to improve a candidate model, such as modify model parameters, modify the model structure, add more training data and leverage interactive machine learning techniques. Unfortunately, no adequate guidelines were provided on *when* and *how* to choose a particular model modification approach. For instance, even though they acknowledge that modifying the model by introducing interactive machine learning techniques is a challenging problem, no further guidelines were provided on when and how the designer should devise such techniques so that the resulting system improves significantly.

Echoing the call for aligning the model with the tasks and visual encoding [21], we focused on the problem domain characterization and task abstractions first and then devised the suitable text analysis models. Based on our data and task abstractions, we also made modeling choices, i.e., choosing the suitable topic modeling and sentiment analysis methods.

In the next step, we *verified* the performance of the topic model with end users. For instance, after designing ConVis, we attempted to *verify* the performance of the topic models by running an informal evaluation with a small number of real users, as discussed in Chapter 2. Through this evaluation, we identified that the current model sometimes does not match the user's mental model and current tasks. Therefore, we pondered *how* the model could be modified so that it can support users in performing their tasks more effectively. Since our analysis revealed that the perceived usefulness of the topic model depends on user's mental model and current tasks, allowing the user to revise the model was deemed to be more promising than other alternative approaches, such as modifying the model parameters, and the model structure.

Once we had decided to introduce interactive topic modeling (as described in Chapter 4), we faced another important challenge of *how* to devise a minimal set of interactive topic revision operations, that real users would find useful. To approach this issue, we first identified a set of candidate operations from existing

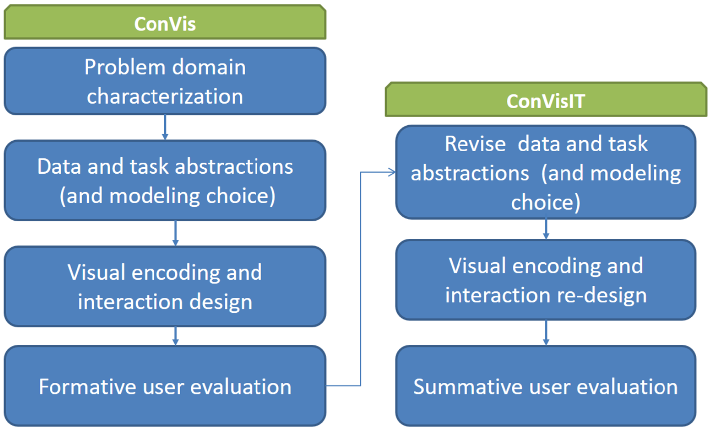

ConVis
Problem domain characterization
Data and task abstractions (and modeling choice)
Visual encoding and interaction design
Formative user evaluation
ConVisIT
Revise data and task abstractions (and modeling choice)
Visual encoding and interaction re-design
Summative user evaluation

Figure 6.1: Design stages of ConVis and ConVisIT.

literature on topic revision and then we prioritized the operations based on three criteria, i.e., 1) task relevancy, 2) topic model relevancy, and 3) redundancy. These criteria helped us to tie the model modification process with the task abstractions and the current topic model. As a result, we were able to design a new interactive topic modeling method that better matches the goals of users engaged in exploring and analyzing conversations. An overview of our design process is illustrated in Figure 6.1. A similar design process was successfully applied when we introduced the human-in-the-loop topic modeling approach in MultiConVisIT.

Based on our experience and current literature, we summarize the following guidelines for a designer of a visual text analytics system: We suggest that rather than treating text analysis models as black boxes, the designer should consider how to tailor and adapt these models based on a detailed analysis of specific user needs and requirements in the target domains. Furthermore, the designer should iteratively analyze the performance of text analysis methods to determine whether to introduce human-in-the-loop in the computation process or focus on improving the model without considering human supervision. Finally, if the designer decides to introduce human-in-the-loop computation, the type of interactive feedback operations for modifying the model should be derived based on the data and tasks abstractions in the target domain. In essence, we call for applying the user-centered design approach to inform and iteratively refine both the text analysis methods and interactive visualizations design.

6.2 Impact of Our Visual Text Analytics Systems

Since our visual text analytics systems have been made publicly available, they have been tailored and adopted in a variety of domains, both in our work as well as in other research projects. In Chapter 5, we have already reported a design study for a community question answering forum, where our visual text analytics solutions were simplified and tailored to support information seeking tasks.

In addition to our work, several other researchers have applied or partially adopted the data abstractions and visual encodings of MultiConVis and ConVis in a variety of domains, ranging from news comments [32, 107], to online health forums [76, 84], to educational forums [47]. We now analyze these recent works and discuss similarities and differences with our systems.

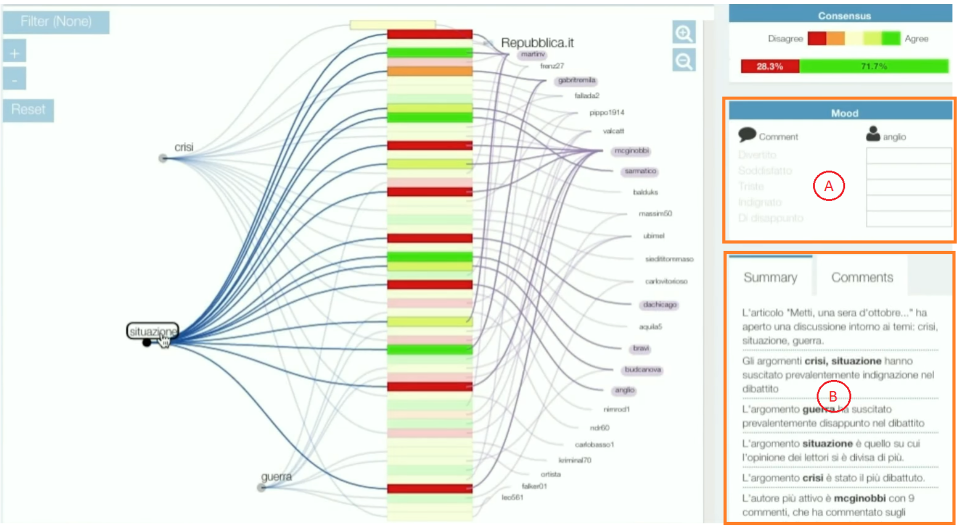
News comments: SENSEI[1] is a research project that was funded by the European Union and was conducted in collaboration with four leading universities and two industry partners in Europe. The main goal of this project was to develop summarization and analytics technology to help users make sense of human conversation streams from diverse media channels, ranging from comments generated for news articles to customer-support conversations in call centers.

After the research work on developing ConVis was published and the tool was made publicly available, the SENSEI project researchers expressed their interest in adopting our system. Their primary objective was to evaluate their text summarization and analytics technology by visualizing the results with ConVis, with the final goal of detecting end-user improvements in task performance and productivity.

In their version of the interface[2], they kept the main features of ConVis, namely the topics, authors, and thread overview; and then added some new features to show text analytics results specific to their application, as shown in Figure 6.2 [107]. In particular, within the thread overview, for each comment they encoded how much this comment agrees or disagrees with the original article, instead of showing the sentiment distribution of that comment. Another additional interactive feature is

[1] www.sensei-conversation.eu

[2] A video demo of their version of the interface is available at www.youtube.com/watch?v=XIMP0cuiZIQ



L'articolo "Metti, una sera d'ottobre..." ha aperto una discussione intorno ai temi: crisi, situazione, guerra.

Gli argomenti **crisi, situazione** hanno suscitato prevalentemente indignazione nel dibattito

L'argomento **guerra** ha suscitato prevalentemente disappunto nel dibattito

L'argomento **situazione** è quello su cui l'opinione dei lettori si è divisa di più.

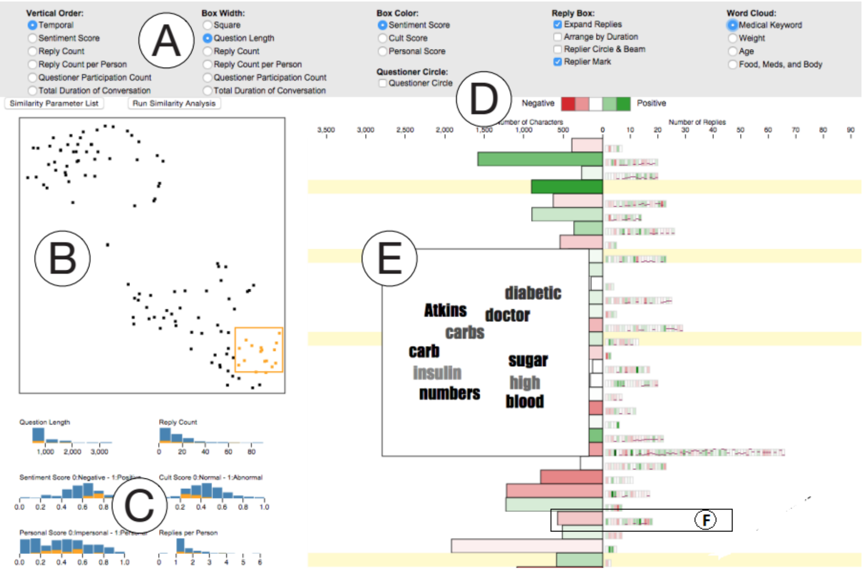
L'argomento **crisi** è stato il più dibattuto.

L'autore più attivo è **mcginobbi** con 9 commenti, che ha commentato sugli

Figure 6.2: A screenshot of the modified ConVis interface used in the SENSEI project. The interface shows the results of some additional text analysis methods, namely the degree of agreement/disagreement between a comment and the original article (within the thread overview), the predicted mood of the corresponding author (A), and the textual summary of the conversation (B) [107].

that clicking on an author element results in showing the predicted mood of that author (using five different modes, i.e., amused, satisfied, sad, indignant, and disappointed). Furthermore, they added a summary view that shows a textual summary of the conversation in addition to the detailed comments. Finally, they introduced some new interactive features, such as zooming and filtering to deal with the conversations that are very long with several hundreds of comments.

Online health forums: Kwon et al. developed VisOHC [76], a visual analytics system designed for administrators of online health communities (OHCs). In this paper, they cite our work and discuss the similarities as well as the differences between VisOHC and ConVis. For instance, similar to the thread overview in ConVis, they represented the comments of a conversation using a sequence of rectangles and used the color encoding within those rectangles to represent sentiment (see Figure 6.3). However, they encoded additional data in order to support the specific domain goals and tasks of OHC administrators. For instance, they



Vertical Order:
● Temporal
○ Sentiment Score
○ Reply Count
○ Reply Count per Person
○ Questioner Participation Count
○ Total Duration of Conversation

Box Width:
○ Square
● Question Length
○ Reply Count
○ Reply Count per Person
○ Questioner Participation Count
○ Total Duration of Conversation

Box Color:
● Sentiment Score
○ Cult Score
○ Personal Score

Questioner Circle:
□ Questioner Circle

Reply Box:
☑ Expand Replies
□ Arrange by Duration
□ Replier Circle & Beam
☑ Replier Mark

Word Cloud:
● Medical Keyword
○ Weight
○ Age
○ Food, Meds, and Body

Figure 6.3: VisOHC visually represents the comments of a conversation using a sequence of rectangles (F), where color within each rectangle represents sentiment expressed in a comment. Additionally it shows a scatter plot (B), and a histogram view (C) (The figure is adapted from [76]).

used a scatter plot to encode the similarities between discussion threads and a histogram view to encode various statistical measures regarding the selected threads, as shown in Figure 6.3.

Mamykina et al. analyzed how users in online health communities collectively make sense of the vast amount of information and opinions within an online diabetes forum, called TuDiabetes [84]. Their study found that members of TuDiabetes often value a multiplicity of opinions rather than consensus. From their study, they concluded that in order to facilitate the collective sensemaking of such diversity of opinions, a visual text analytics tool like ConVis could be very effective. They also mentioned that in addition to topic modeling and sentiment analysis, some other text analysis methods related to their health forum under study, such as detection of agreement and topic shift in conversation, should be devised and incorporated into tools like ConVis.

Educational forums: More recently, Fu et al. presented iForum, an interactive visual analytics system for helping instructors in understanding the temporal patterns of student activities and discussion topics in a MOOC forum [47]. They mentioned that while the design of iForum has been inspired by tools such as ConVis, they have tailored their interface to the domain-specific problems of MOOC forums. For instance, like ConVis, their system also provides an overview of topics and discussion threads, however, they focused more on temporal trends of an entire forum, as opposed to an individual conversation or a set of conversations related to a specific query.

Beyond online conversations: Recently, Shen et al. present NameClarifier [116], a visual analytics system that supports the user to interactively disambiguate author names in bibliographic citation records. In this system, they partially adopted our visual encoding technique for arranging the facets namely topics and authors and exposing the relations between them. More specifically, they encoded the relations between a list of ambiguous authors and a list of confirmed authors via subtle curves, similarly to how ConVis visualizes the relations between topics and authors by linking them to the corresponding comments in the thread overview.

6.3 Summary of Contributions

The contributions of this dissertation can be summarized as follows:

- A user requirements analysis based on extensive literature review in the domain of blogs to inform our interface design for both a single conversation as well as a set of conversations (Chapter 2 and 3).

- Adoption of a topic modeling method for mining topics from a single conversation by taking advantage of the conversational features (Chapter 2). We also extended this method for creating a topic hierarchy for a collection of conversations, by organizing the topics extracted from each conversation in the collection (Chapter 3).

- The design of ConVis and MultiConVis, which visualize both *topic* and *opinion* mining results along with a set of metadata, such as authors and position of the comments. We also proposed a way to seamlessly integrate the two interfaces to allow users to switch from exploring a collection of conversations to a single conversation (Chapter 2 and 3).

- Results from a series of user studies, namely an informal evaluation, a formal lab-based study, and three case studies, which revealed the differences in user performance and subjective opinions when our systems were compared to traditional blog interfaces for exploring conversations (Chapter 2 and 3).
- A novel interactive topic modeling approach specifically devised for asynchronous conversations. We developed: i) a system that revises the topic model generated from a single conversation, as well as the topic model generated from a collection of conversations (Chapter 4) and ii) interactive features to help the user in performing a set of topic revision operations (Chapter 4).
- Results from two lab-based user studies which revealed the potential utility of our human-in-the-loop topic modeling approaches (Chapter 4).
- Demonstration of how our generic solutions for integrating NLP and InfoVis techniques presented in Chapter 2 and 3 can be simplified and tailored to the information seeking tasks of a community question answering forum users (Chapter 5).
- Evaluation of our new community question answering forum tool in the wild in an ecologically valid testing by deploying the system among real forum readers (Chapter 5).
- Generalizable lessons that can inform the design of visual interfaces for online conversations in other domains, as well as to design for user population possibly having low visualization expertise (Chapter 5).

6.4 Limitations and Future Work

While this thesis has made some significant progress in supporting the tasks of exploring online conversations, it also raises further challenges, open questions, and ideas for future work. Here we discuss the key challenges and opportunities for future research.

How can we scale up our systems for big data? As social media conversational data is growing in size and complexity at an unprecedented rate, new challenges have emerged from both computational and visualization perspectives. In particular, we need to address the following aspects of big data, while designing visual text analytics for online conversations:

Volume: Most of the existing visualizations are inadequate to handle very large amounts of raw conversational data. For example, ConVis scales with conversations with hundreds of comments; however, it is unable to deal with a very long conversation consisting of more than a thousand comments. To tackle the scalability issue, we will investigate computational methods for filtering and aggregating comments, as well as devise interactive visualization techniques such as zooming to progressively disclose the data from a high-level overview to low-level details.

Velocity: The systems that we have developed do not process streaming conversations. Yet in many real-world scenarios, conversational data is constantly produced at a high rate, which poses enormous challenges for mining and visualization methods. For instance, immediately after a product is released a business analyst may want to analyze text streams in social media to identify problems or issues, such as whether customers are complaining about a feature of the product. In these cases, timely analysis of the streaming text can be critical for the company's reputation. For this purpose, we aim to investigate how to efficiently mine streaming conversations and how to visualize the extracted information in real time to the user.

How can we support the user in tailoring our systems to a specify conversational genre, a specific domain, or tasks? In Section 6.2, we already discussed how our current visual text analytics systems have been applied and tailored to various domains. However, in these systems, the user does not have flexibility in terms of the choice of the datasets and the available interaction techniques. Therefore, it may take a significant amount of programming effort to re-design the interface for a specific conversational domain. For example, when we tailored our system to a community question answering forum with a specific user population in mind, we had to spend a considerable amount of time modifying the existing code in order to re-design the interface for the new conversational genre.

In this context, can we enable a large number of users - not just those who have strong programming skills to author visual interfaces for exploring conversations in a new domain? To answer this question, we need to research how to construct an interactive environment that supports custom visualization design for different domains without requiring the user to write any code. Such interactive environment would allow the user to have more control over the data to be represented and the interactive techniques to be supported. To this end, we will investigate current research on general purpose visual authoring tools such as Lyra [110] and IVisDesigner [106], which provide custom visualization authoring environments, to understand how we can build a similar tool, but specifically for conversational data.

How can the system adapt to a diverse range of users? A critical challenge of introducing a new visualization is that the effectiveness of visualization techniques can be impacted by different user characteristics, such as visualization expertise, cognitive abilities, and personality traits [27]. Unfortunately, most previous work has focused on finding individual differences for simple visualizations only, such as bar and radar graphs [122]. It is still unknown how the individual differences might impact a more complex visualization like ConVis, that not only requires coordinations between text and visualization but also supports more complex interactive techniques. In this regard, we will examine what aspects of a visual text analytics system are impacted by user characteristics and how to dynamically adapt the visualization to such characteristics.

How can we leverage text analysis and visualization techniques to develop advanced storytelling tools for online conversations? Data storytelling has become increasingly popular among InfoVis practitioners such as journalists, who may want to create a visualization from social media conversations and integrate it into their narratives to convey critical insights. Unfortunately, even sophisticated visualization tools like Tableau [3] offer only limited support for authoring data stories, requiring users to manually create textual annotations and organize the sequence of visualizations. More importantly, they do not provide methods for processing

[3] www.tableau.com

the unstructured or semi-structured data generated in online conversations.

In this context, we aim to investigate how to leverage NLP and InfoVis techniques for online conversations to create effective semi-automatic authoring tools for data storytelling. More specifically, we need to devise methods for generating and organizing the summary content from online conversations and choosing the sequence in which such content is delivered to users. To this end, we will investigate current research on narrative visualization [67, 112].

6.5 Final Remarks

The overarching goal of this dissertation was to combine text analysis and interactive visualization techniques to support users in exploring online conversations. To that aim, we posed a set of research questions in Chapter 1 that guided the development of our visual text analytics systems. These research questions were answered by synthesizing design study methodologies in information visualization, text analysis methods specifically designed to deal with conversational data, and human-in-the-loop computation to deal with noisy text analysis results.

We applied these considerations to the design, implementation, and evaluation of a variety of text analytics systems. Our first system, ConVis addresses the challenges of exploring and analyzing an asynchronous conversation, by offering a visual overview of topics, authors, and the thread structure of a conversation (Chapter 2). Next, MultiConVis moves beyond visualizing a single conversation to a collection of conversations related to a given query (Chapter 3). It combines a novel hierarchical topic modeling technique with interactive visualization in order to support users in understanding the discussions and allow them to seemingly switch from exploring a collection of conversations to a single conversation. We conducted a series of user studies through informal evaluation, case studies, and lab-based studies, which revealed significant improvements in user performance and subjective measures when our systems were compared to traditional blog interfaces. The outcomes from these studies also motivated us to introduce an interactive topic modeling approach. The resulting systems, ConVisIT and MultiConVisIT empower the user in revising the underlying topic models through an intuitive set of interactive features when the current models are noisy and/or insufficient to support their information seeking tasks (Chapter 4). Finally, the online deployment of CQAVis, a visual interface for supporting information seeking in community question answering forums demonstrates that our systems can be effectively tailored to a specific domain problem – a critical finding that indicates the generality and applicability of our approach (Chapter 5).

Despite the tremendous advances in NLP and InfoVis, only little effort has been devoted to combining sophisticated text analysis and interactive visualization techniques in a synergistic way to address information overload problems. This dissertation demonstrates that by tightly integrating advanced text analysis and InfoVis techniques, guided by a human-centered design approach, we can effectively support users in dealing with these problems in a variety of contexts.

We believe that exploring online conversations is just one example that can be supported more effectively by combining techniques from text analysis and information visualization, guided by a user-centred design approach. Beyond online conversations, there are many other types of text collections, such as scientific documents, news articles, and literature, where creating a strong synergy between these two research areas is critical in addressing the information overload problem. Therefore, we envision that a similar approach for combining NLP and InfoVis could also help users in exploring these text collections more efficiently and effectively.

Bibliography

[1] Macrumors, 2016 (accessed December 28, 2016). http://macrumors.com/. → pages 3, 50, 52, 69

[2] Slashdot, 2016 (accessed January 28, 2016). http://slashdot.com/. → pages 35, 52

[3] Daily Kos, 2017 (accessed February 01, 2017). http://dailykos.com/. → pages 35, 69

[4] Alexa's Internet traffic rating service, 2017 (accessed February 25, 2017). http://www.alexa.com/topsites. → pages 1

[5] ColorBrewer, 2017 (accessed February 25, 2017). http://colorbrewer2.org/. → pages 37

[6] Pew Research, 2017 (accessed February 25, 2017). http://www.pewinternet.org/2016/11/11/social-media-update-2016/. → pages 1

[7] Wordpress user activities, 2017 (accessed February 25, 2017). https://wordpress.com/activity/. → pages 1

[8] Illinois Part of Speech Tagger, 2017 (accessed March 09, 2017). http://bit.ly/1xtjFHe. → pages 57

[9] R. Aggarwal, R. Gopal, R. Sankaranarayanan, and P. V. Singh. Blog, blogger, and the firm: Can negative employee posts lead to positive outcomes? *Information Systems Research*, 23(2):306–322, 2012. → pages 25, 26

[10] E. Alberdi and D. H. Sleeman. Retax: A step in the automation of taxonomic revision. *Artifical Intelligence*, 91(2):257–279, Apr. 1997. → pages 82

[11] D. Andrzejewski, X. Zhu, and M. Craven. Incorporating domain knowledge into topic modeling via Dirichlet forest priors. In *Proceedings of the International Conference on Machine Learning*, pages 25–32, 2009. → pages 81, 83, 84

[12] A. Barrón-Cedeño, S. Filice, G. Da San Martino, S. Joty, L. Màrquez, P. Nakov, and A. Moschitti. Thread-level information for comment classification in community question answering. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference of the Asian Federation of Natural Language Processing (ACL-IJCNLP)*, pages 687–693, 2015. → pages 128

[13] E. Baumer, M. Sueyoshi, and B. Tomlinson. Exploring the role of the reader in the activity of blogging. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1111–1120, 2008. → pages 25, 26, 29, 55

[14] D. Bawden and L. Robinson. The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2):180–191, 2009. → pages 2

[15] D. Boyd. A blogger's blog: Exploring the definition of a medium. *Reconstruction*, 6(4), 2006. → pages 11

[16] N. Cao, D. Gotz, J. Sun, Y.-R. Lin, and H. Qu. SolarMap: Multifaceted visual analytics for topic exploration. In *Proceedings of the IEEE Conference on Data Mining (ICDM)*, pages 101–110, 2011. → pages 32, 33

[17] G. Carenini, G. Murray, and R. Ng. *Methods for Mining and Summarizing Text Conversations*. Morgan Claypool, 2011. → pages 1, 3, 10, 11, 78

[18] S. Carpendale. Evaluating information visualizations. In *Information Visualization: Human-Centered Issues and Perspectives*, pages 19–45. Springer, 2008. → pages 75, 135

[19] J. Choo, C. Lee, C. K. Reddy, and H. Park. Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 19(12):1992–2001, 2013. → pages 20, 80, 81, 82, 83, 84

[20] J. Chuang, C. D. Manning, and J. Heer. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the International*

Working Conference on Advanced Visual Interfaces (AVI), pages 74–77, 2012. → pages 81

[21] J. Chuang, D. Ramage, C. Manning, and J. Heer. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 443–452. ACM, 2012. → pages 146, 147

[22] J. Chuang, S. Gupta, C. Manning, and J. Heer. Topic model diagnostics: Assessing domain relevance via topical alignment. In *Proceedings of the Conference on Machine Learning*, pages 612–620, 2013. → pages 78

[23] J. Chuang, Y. Hu, A. Jin, J. D. Wilkerson, D. A. McFarland, C. D. Manning, and J. Heer. Document exploration with topic modeling: Designing interactive visualizations to support effective analysis workflows. In *NIPS Workshop on Topic Models: Computation, Application and Evaluation*, pages 1–4, 2013. → pages 81, 83

[24] A. Cockburn, A. Karlson, and B. B. Bederson. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR)*, 41(1):2, 2008. → pages 36, 130

[25] C. Collins, S. Carpendale, and G. Penn. Docuburst: Visualizing document content using language structure. In *Computer graphics forum*, volume 28, pages 1039–1046. Wiley Online Library, 2009. → pages 6

[26] M. Collins and N. Duffy. Convolution kernels for natural language. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems (NIPS)*, pages 625–632. MIT Press, 2002. → pages 129

[27] C. Conati, G. Carenini, E. Hoque, B. Steichen, and D. Toker. Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. *Computer Graphics Forum (Proceedings of EuroVis)*, 33(3):371–380, 2014. → pages 155

[28] W. Cui, S. Liu, Z. Wu, and H. Wei. How hierarchical topics evolve in large text corpora (Proceedings of InfoVis). *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2281–2290, 2014. → pages 53, 54

[29] C. Culy and V. Lyding. Double Tree: An advanced kwic visualization for expert users. In *14th International Conference Information Visualisation*, pages 98–103, 2010. → pages 6

[30] K. Dave, M. Wattenberg, and M. Muller. Flash forums and forumReader: navigating a new kind of large-scale online discussion. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW)*, pages 232–241, 2004. → pages 25, 26, 31, 55

[31] M. De Choudhury and H. Sundaram. Why do we converse on social media?: an analysis of intrinsic and extrinsic network factors. In *Proceedings of the ACM SIGMM International Workshop on Social Media*, pages 53–58. ACM, 2011. → pages 55

[32] P. Deokgun, S. Simranjit, D. Nicholas, , and E. Niklas. Supporting comment moderators in identifying high quality online news comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1114–1125, 2016. → pages 2, 149

[33] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 115–122, 2010. → pages 54

[34] J. Donath. A semantic approach to visualizing online conversations. *Communications of the ACM*, 45(4):45–49, 2002. → pages 31

[35] J. Donath, K. Karahalios, and F. Viégas. Visualizing conversation. *Journal of Computer-Mediated Communication*, 4(4):1–9, 1999. → pages 25

[36] M. Dörk, S. Carpendale, C. Collins, and C. Williamson. VisGets: Coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 14(6):1205–1212, 2008. → pages 52

[37] M. Dörk, D. Gruen, C. Williamson, and S. Carpendale. A visual backchannel for large-scale events. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 16(6):1129–1138, 2010. → pages 53, 146

[38] M. Dörk, N. H. Riche, G. Ramos, and S. Dumais. PivotPaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 18(12):2709–2718, Dec. 2012. → pages 32, 33

[39] C. dos Santos, L. Barbosa, D. Bogdanova, and B. Zadrozny. Learning hybrid representations to retrieve semantically equivalent questions. In

Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) and the International Joint Conference on Natural Language Processing (ACL-IJCNLP), pages 694–699, 2015. → pages 128

[40] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. HierarchicalTopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 19(12):2002–2011, 2013. → pages 53, 82

[41] N. B. Ellison, R. Gray, J. Vitak, C. Lampe, and A. T. Fiore. Calling all Facebook friends: Exploring requests for help on Facebook. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*, pages 155–164. → pages 122

[42] J. L. Elsas. *Leveraging collection structure in information retrieval with applications to search in conversational social media*. PhD thesis, Carnegie Mellon University. → pages 11

[43] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 473–482. ACM, 2012. → pages 84

[44] S. Faridani, E. Bitton, K. Ryokai, and K. Goldberg. Opinion Space: a scalable tool for browsing online comments. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1175–1184, 2010. → pages 32

[45] N. FitzGerald, G. Carenini, G. Murray, and S. Joty. Exploiting conversational features to detect high-quality blog comments. In *Proceedings of the Canadian Artificial Intelligence*, pages 739–744, 2007. → pages 46

[46] J. L. Fleiss, B. Levin, and M. C. Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, 2013. → pages 106, 114

[47] S. Fu, J. Zhao, W. Cui, and H. Qu. Visual analysis of MOOC forums with iForum. *IEEE Transactions on Visualization and Computer Graphics (Prooceedings of VAST)*, 23(1):201–210, 2017. → pages 149, 152

[48] T. Furukawa, Y. Matsuo, I. Ohmukai, K. Uchiyama, and M. Ishizuka. Social networks and reading bahavior in blogosphere. In *Proceedings of the International AAAI Conference on Weblogs and Social Media(ICWSM)*, 2007. → pages 25

[49] M. Galley, K. McKeown, E. Fosler-Lussier, and H. Jing. Discourse segmentation of multi-party conversation. In *Proceedings of the Annual Meeting on Association for Computational Linguistics (ACL)*, pages 562–569, 2003. → pages 33, 59

[50] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg. Exploring the placement and design of word-scale visualizations. *IEEE Transactions on Visualization and Computer Graphics (Proocedings of InfoVis)*, 20(12): 2291–2300, Dec 2014. → pages 64

[51] F. M. Harper, D. Raban, S. Rafaeli, and J. A. Konstan. Predictors of answer quality in online Q&A sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 865–874, 2008. → pages 122, 123

[52] K. S. Hasan and V. Ng. Automatic keyphrase extraction: A survey of the state of the art. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1262–1273, 2014. → pages 146

[53] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. ThemeRiver: visualizing thematic changes in large document collections. *Transactions on Visualization and Computer Graphics*, 8(1):9–20, 2002. → pages 53

[54] M. A. Hearst. TileBars: visualization of term distribution information in full text information access. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 59–66, 1995. → pages 6

[55] M. A. Hearst. Design recommendations for hierarchical faceted search interfaces. In *Proceedings of the SIGIR Workshop on Faceted Search*, pages 26–30, 2006. → pages 32

[56] M. A. Hearst. 'Natural' search user interfaces. *Communications of the ACM*, 54(11):60–67, Nov. 2011. → pages 122

[57] M. A. Hearst, M. Hurst, and S. T. Dumais. What should blog search look like? In *Proceedings of the ACM Workshop on Search in Social Media*, pages 95–98, 2008. → pages 25, 55

[58] J. Heer and G. G. Robertson. Animated transitions in statistical data graphics. *Transactions on Visualization and Computer Graphics*, 13(6): 1240–1247, 2007. → pages 93, 96

[59] E. Hoque and G. Carenini. ConVis: A visual text analytic system for exploring blog conversations. *Computer Graphics Forum (Proceedings EuroVis)*, 33(3):221–230, 2014. → pages v, 23, 54, 66, 78, 83

[60] E. Hoque and G. Carenini. MultiConVis: A visual text analytics system for exploring a collection of online conversations. In *Proceedings of the ACM Conference on Intelligent User Interfaces (IUI)*, pages 96–107, 2016. → pages v, vi, 49, 77

[61] E. Hoque and G. Carenini. Interactive topic modeling for exploring asynchronous online conversations: Design and evaluation of ConVisIT. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 6(1):7:1–7:24, Feb. 2016. → pages vi, 78

[62] E. Hoque, G. Carenini, and S. R. Joty. Interactive exploration of asynchronous conversations: Applying a user-centered approach to design a visual text analytic system. In *Proceedings Workshop on Interactive Language Learning, Visualization, and Interfaces (ILLVI 2014), in conjunction with the ACL-2014*, pages 45–52, 2014. → pages v

[63] E. Hoque, S. Joty, M. Lluís, and G. Carenini. CQAVis: Visual text analytics for community question answering. In *Proceedings of the ACM conference on Intelligent User Interfaces (IUI)*, pages 161–172, 2017. → pages vi, 119

[64] D. Horowitz and S. D. Kamvar. The anatomy of a large-scale social search engine. In *Proceedings of the International Conference on World Wide Web (WWW)*, pages 431–440. ACM, 2010. → pages 122

[65] G. Hsieh and S. Counts. mimir: A market-based real-time question and answer service. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 769–778, 2009. → pages 123

[66] Y. Hu, J. Boyd-Graber, B. Satinoff, and A. Smith. Interactive topic modeling. *Machine Learning*, 95(3):423–469, 2014. → pages 20, 78, 80, 81, 83, 84

[67] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 17(12):2231–2240, 2011. → pages 156

[68] Indratmo, J. Vassileva, and C. Gutwin. Exploring blog archives with interactive visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, pages 39–46, 2008. → pages 31, 50, 52

[69] Q. Jones, G. Ravid, and S. Rafaeli. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research*, 15(2):194–210, 2004. → pages 3, 4, 24, 25, 48

[70] S. Joty, G. Carenini, and R. T. Ng. Topic segmentation and labeling in asynchronous conversations. *Journal of Artificial Intelligence Research*, 47:521–573, 2013. → pages 7, 9, 11, 24, 28, 33, 35, 36, 46, 50, 78, 81, 83

[71] M. C. Kaptein, C. Nass, and P. Markopoulos. Powerful and consistent analysis of likert-type ratingscales. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 2391–2394, 2010. → pages 72, 101

[72] B. K. Kaye. Web side story: An exploratory study of why weblog users say they use weblogs. AEJMC Annual Conference, 2005. → pages 25, 26, 55

[73] D. A. Keim and D. Oelke. Literature fingerprinting: A new method for visual literary analysis. In *IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 115–122, 2007. → pages 6

[74] D. Kim and T. J. Johnson. Political blog readers: Predictors of motivations for accessing political blogs. *Telematics and Informatics*, 29(1):99–109, Feb. 2012. → pages 25, 26

[75] R. Kiros, Y. Zhu, R. R. Salakhutdinov, R. Zemel, R. Urtasun, A. Torralba, and S. Fidler. Skip-thought vectors. In *Advances in neural information processing systems*, pages 3294–3302, 2015. → pages 61

[76] B. Kwon, S.-H. Kim, S. Lee, J. Choo, and J. Y. Jina Huh. Visohc: Designing visual analytics for online health communities. *IEEE Transactions on Visualization and Computer Graphics*, 2015. → pages 2, 149, 150, 151

[77] H. Lam. A framework of interaction costs in information visualization. *Transactions on Visualization and Computer Graphics*, 14(6):1149–1156, 2008. → pages 30, 39, 132

[78] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012. → pages 8, 41, 47, 62, 70, 75, 99, 129, 135

[79] S. Laqua and M. A. Sasse. Exploring blog spaces: a study of blog reading experiences using dynamic contextual displays. In *Proceedings of the British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, pages 252–261, 2009. → pages 11

[80] B. Lee, G. Smith, G. G. Robertson, M. Czerwinski, and D. S. Tan. FacetLens: Exposing trends and relationships to support sensemaking within faceted datasets. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1293–1302, 2009. → pages 32

[81] H. Lee, J. Kihm, J. Choo, J. Stasko, and H. Park. iVisClustering: An interactive visual document clustering via topic modeling. In *Computer Graphics Forum (Proceedings of EuroVis)*, volume 31, pages 1155–1164, 2012. → pages 20, 80, 81, 82, 83, 84

[82] S. Liu, X. Wang, J. Chen, J. Zhu, and B. Guo. TopicPanorama: a full picture of relevant topics. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 183–192, 2014. → pages 53

[83] C. Macdonald, R. L. Santos, I. Ounis, and I. Soboroff. Blog track research at TREC. *SIGIR Forum*, 44(1):58–75, 2010. → pages 25, 27

[84] L. Mamykina, D. Nakikj, and N. Elhadad. Collective sensemaking in online health forums. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 3217–3226, 2015. → pages 149, 151

[85] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. TwitInfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 227–236, 2011. → pages 54

[86] S. McKenna, D. Mazur, J. Agutter, and M. Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 20(12):2191–2200, 2014. → pages 146

[87] G. Mishne. Information access challenges in the blogspace. In *Workshop on Intelligent Information Access (IIIA)*, 2006. → pages 25, 26

[88] M. R. Morris, J. Teevan, and K. Panovich. What do people ask their social networks, and why?: A survey study of status message Q&A behavior. In

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), pages 1739–1748, 2010. → pages 122, 123, 137, 138

[89] A. Moschitti. Efficient Convolution Kernels for Dependency and Constituent Syntactic Trees. In J. Fürnkranz, T. Scheffer, and M. Spiliopoulou, editors, *Machine Learning: ECML 2006*, volume 4212 of *Lecture Notes in Computer Science*. 2006. → pages 129

[90] S. A. Munson and P. Resnick. Presenting diverse political opinions: how and how much. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1457–1466, 2010. → pages 25, 26

[91] T. Munzner. A nested model for visualization design and validation. *Transactions on Visualization and Computer Graphics (Proocedings of InfoVis)*, 15(6):921–928, 2009. → pages 8, 24, 41, 146

[92] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014. → pages 125

[93] G. Murray, E. Hoque, and G. Carenini. Opinion summarization and visualization. In F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, editors, *Sentiment Analysis in Social Networks*, pages 171 – 187. Morgan Kaufmann, 2017. → pages 2

[94] P. Nakov, L. Màrquez, A. Moschitti, W. Magdy, H. Mubarak, A. A. Freihat, J. Glass, and B. Randeree. SemEval-2016 task 3: Community question answering. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval)*, 2016. → pages 126, 127, 128, 129

[95] S. Narayan and C. Cheshire. Not too long to read: The tldr interface for exploring and navigating large-scale discussion spaces. In *Hawaii Conference on System Sciences (HICSS)*, pages 1–10, 2010. → pages 27, 29, 31

[96] B. A. Nardi, D. J. Schiano, M. Gumbrecht, and L. Swartz. Why we blog. *Communications of the ACM*, 47(12):41–46, 2004. → pages 11

[97] M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004. → pages 59, 61, 86, 91

[98] M. Nicosia, S. Filice, A. Barrón-Cedeño, I. Saleh, H. Mubarak, W. Gao, P. Nakov, G. Da San Martino, A. Moschitti, K. Darwish, L. Màrquez,

S. Joty, and W. Magdy. QCRI: Answer selection for community question answering - experiments for Arabic and English. In *In Proceedings of the International Workshop on Semantic Evaluation (SemEval)*, 2015. → pages 128

[99] N. Nikitina, S. Rudolph, and B. Glimm. Interactive ontology revision. *Web Semantics: Science, Services and Agents on the World Wide Web*, 12: 118–130, 2012. → pages 82

[100] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994. → pages 82

[101] V. Pascual-Cid and A. Kaltenbrunner. Exploring asynchronous online discussions through hierarchical visualisation. In *IEEE Conference on Information Visualization*, pages 191–196, 2009. → pages 6, 31

[102] P. Pirolli, P. Schank, M. Hearst, and C. Diehl. Scatter/gather browsing communicates the topic structure of a very large text collection. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 213–220. ACM, 1996. → pages 81

[103] X. Qiu and X. Huang. Convolutional neural tensor network architecture for community-based question answering. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1305–1311, 2015. → pages 128

[104] D. Raban and F. Harper. Motivations for answering questions online. *New Media and Innovative Technologies*, 73, 2008. → pages 123

[105] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings Conference on Empirical Methods on Natural Language Processing (EMNLP)*, pages 248–256, 2009. → pages 81

[106] D. Ren, T. Höllerer, and X. Yuan. iVisDesigner: Expressive interactive design of information visualizations. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 20(12):2092–2101, 2014. → pages 155

[107] G. Riccardi, C. Balamurali, A R, B. Fabio, Favre, F. Carmelo, A. Funk, R. Gaizauskas, and V. Lanzolla. Report on the summarization views of the sensei prototype. In *Technical report*, 2015. → pages 149, 150

[108] A. Ritter, C. Cherry, and B. Dolan. Unsupervised modeling of twitter conversations. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 172–180. Association for Computational Linguistics, 2010. → pages 11, 50

[109] W. Sack. Conversation Map: an interface for very-large-scale conversations. *Journal of Management Information Systems*, 17(3):73–92, 2000. → pages 7, 31

[110] A. Satyanarayan and J. Heer. Lyra: An interactive visualization design environment. 33(3):351–360, 2014. → pages 155

[111] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012. → pages 8, 121, 146

[112] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6): 1139–1148, 2010. → pages 156

[113] C. Seifert, B. Kump, W. Kienreich, G. Granitzer, and M. Granitzer. On the beauty and usability of tag clouds. In *IEEE Conference on Information Visualisation*, pages 17–25, 2008. → pages 6

[114] A. Severyn and A. Moschitti. Structural relationships for large-scale learning of answer re-ranking. In *Proceedings of the International ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 741–750. ACM, 2012. → pages 129

[115] A. Severyn and A. Moschitti. Learning to rank short text pairs with convolutional deep neural networks. In *Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR)*, pages 373–382, 2015. → pages 128

[116] Q. Shen, T. Wu, H. Yang, Y. Wu, H. Qu, and W. Cui. Nameclarifier: A visual analytics system for author name disambiguation. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 23(1):141–150, Jan 2017. → pages 152

[117] Y. Shen, W. Rong, Z. Sun, Y. Ouyang, and Z. Xiong. Question/answer matching for cqa system via combining lexical and sequential information.

In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 275–281, 2015. → pages 128

[118] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8): 888–905, 2000. → pages 34, 59, 61, 86, 90

[119] P. V. Singh, N. Sahoo, and T. Mukhopadhyay. Seeking variety: A dynamic model of employee blog reading behavior. In *Workshop on Information Systems and Economics*, 2010. → pages 25, 26, 27, 55

[120] M. Steinberger, M. Waldner, M. Streit, A. Lex, and D. Schmalstieg. Context-preserving visual links. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 17(12):2249–2258, 2011. → pages 38

[121] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2): 267–307, 2011. → pages 35, 62

[122] D. Toker, C. Conati, G. Carenini, and M. Haraty. Towards adaptive information visualization: on the influence of user characteristics. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 274–285. Springer, 2012. → pages 155

[123] G. D. Venolia and C. Neustaedter. Understanding sequence and reply relationships within email conversations: a mixed-model visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 361–368, 2003. → pages 6

[124] F. B. Viégas, S. Golder, and J. Donath. Visualizing email content: portraying relationships from conversational histories. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 979–988, 2006. → pages 7, 53, 146

[125] M. Wattenberg and D. Millen. Conversation thumbnails for large-scale discussions. In *Extended Abstract Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 742–743, 2003. → pages 6, 31

[126] M. Wattenberg and F. B. Viégas. The Word Tree, an interactive visual concordance. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 2008. → pages 6

[127] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. TIARA: a visual exploratory text analytic system. In *Proceedings ACM Conference on Knowledge Discovery and Data Mining*, pages 153–162, 2010. → pages 32, 53

[128] W. Willett, J. Heer, and M. Agrawala. Scented Widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 13(6): 1129–1136, 2007. → pages 57, 65

[129] Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu. OpinionFlow: Visual analysis of opinion diffusion on social media. *IEEE Transactions on Visualization and Computer Graphics (Proceedsings of VAST)*, 20(12):1763–1772, Dec 2014. → pages 54

[130] Y. Yang, S. Pan, Y. Song, J. Lu, and M. Topkara. User-directed non-disruptive topic model update for effective exploration of dynamic content. In *Proceedings of the ACM Conference on Intelligent User Interfaces (IUI)*, pages 158–168, 2015. → pages 81

[131] K.-P. Yee and M. Hearst. Content-centered discussion mapping. *Online Deliberation 2005/DIAC-2005*, 2005. → pages 31

[132] J. Zhao, C. Collins, F. Chevalier, and R. Balakrishnan. Interactive exploration of implicit and explicit relations in faceted datasets. *Transactions on Visualization and Computer Graphics*, 19(12):2080–2089, 2013. → pages 32, 33

[133] J. Zhao, N. Cao, Z. Wen, Y. Song, Y. Lin, and C. Collins. #FluxFlow: Visual analysis of anomalous information spreading on social media (proocedings of VAST). *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1773–1782, Dec 2014. → pages 54

[134] D. Zhou, S. A. Orshanskiy, H. Zha, and C. L. Giles. Co-ranking authors and documents in a heterogeneous network. In *Seventh IEEE International Conference on Data Mining*, pages 739–744, 2007. → pages 35, 59

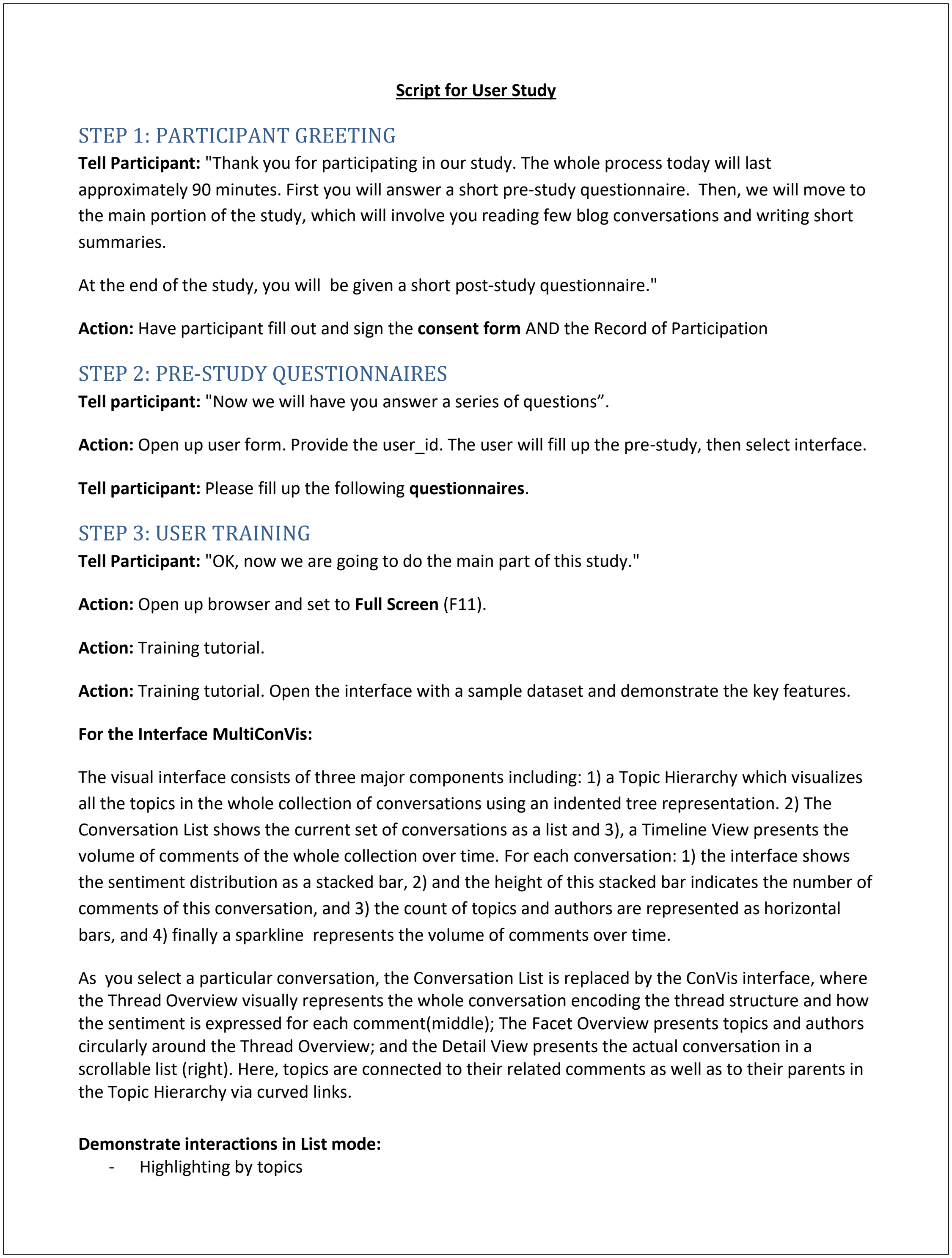
[135] A. R. Zinman. *Me, myself, and my hyperego: understanding people through the aggregation of their digital footprints*. PhD thesis, MIT, 2011. → pages 25, 27

Appendix A

Supplementary Materials for Chapter 3

This appendix contains supplemental materials for Chapter 3, namely the script used by the experimenter to run the study and questionnaires used during the study.

A.1 Script for User Study



Script for User Study

STEP 1: PARTICIPANT GREETING

Tell Participant: "Thank you for participating in our study. The whole process today will last approximately 90 minutes. First you will answer a short pre-study questionnaire. Then, we will move to the main portion of the study, which will involve you reading few blog conversations and writing short summaries.

At the end of the study, you will be given a short post-study questionnaire."

Action: Have participant fill out and sign the **consent form** AND the Record of Participation

STEP 2: PRE-STUDY QUESTIONNAIRES

Tell participant: "Now we will have you answer a series of questions".

Action: Open up user form. Provide the user_id. The user will fill up the pre-study, then select interface.

Tell participant: Please fill up the following **questionnaires**.

STEP 3: USER TRAINING

Tell Participant: "OK, now we are going to do the main part of this study."

Action: Open up browser and set to **Full Screen** (F11).

Action: Training tutorial.

Action: Training tutorial. Open the interface with a sample dataset and demonstrate the key features.

For the Interface MultiConVis:

The visual interface consists of three major components including: 1) a Topic Hierarchy which visualizes all the topics in the whole collection of conversations using an indented tree representation. 2) The Conversation List shows the current set of conversations as a list and 3), a Timeline View presents the volume of comments of the whole collection over time. For each conversation: 1) the interface shows the sentiment distribution as a stacked bar, 2) and the height of this stacked bar indicates the number of comments of this conversation, and 3) the count of topics and authors are represented as horizontal bars, and 4) finally a sparkline represents the volume of comments over time.

As you select a particular conversation, the Conversation List is replaced by the ConVis interface, where the Thread Overview visually represents the whole conversation encoding the thread structure and how the sentiment is expressed for each comment(middle); The Facet Overview presents topics and authors circularly around the Thread Overview; and the Detail View presents the actual conversation in a scrollable list (right). Here, topics are connected to their related comments as well as to their parents in the Topic Hierarchy via curved links.

Demonstrate interactions in List mode:

- Highlighting by topics

- Expand/ collapse topics
- Sorting conversations
- Click timeline button to show sentiment over time
- Filter by time

Demonstrate interactions in Conversation Mode:

- ***Hovering*** the mouse over a **facet** element
    - related comments and facets are highlighted
    - tooltips become visible
- ***Clicking*** over a **facet** element:
    - a thick border is drawn along that element
    - the interface scrolls down to related comments in detail view
    - topic words are highlighted
- ***Hovering*** over a **comment**
    - related topic and author are highlighted
- ***Clicking*** a **comment**
    - sentiment words are highlighted

For the Interface Macrumors:

- Demonstrate Interactions:

- Sort the list of conversations
- Search by keyword
- Switching between list mode and conversation mode.

STEP 4: SELECT TASK

Please read the following task.

Dataset: iPhone bending

The issue of ***'iPhone bending'*** *went viral on social media after the iPhone 6 was launched in September 2014. This incident triggered a huge amount of discussions in Macrumors, a blog site that regularly publishes Apple related news and allows participants to make comments.*

Now, you are going to explore a set of conversations where people are discussing about this issue. You can take notes during the exploration using the opened text editor. At the end of exploring and reading through the set of conversations, your task is to write a summary of what you think are the major discussion points and most insightful comments within the set of conversations. You have 20 minutes to complete the task.

Dataset: iPad release

iPad air 2 was launched in October 2014. This event triggered a huge amount of discussions in Macrumors, a blog site that regularly publishes Apple related news and allows participants to make comments.

Now, you are going to explore a set of conversations where people are discussing about this issue. You can take notes during the exploration using the opened text editor. At the end of exploring and reading through the set of conversations, your task is to write a summary of what you think are the major discussion points and most insightful comments within the set of conversations. You have 20 minutes to complete the task.

STEP 5: IN-STUDY QUESTIONNAIRE

After each task, the participant will fill up a set of in-study questionnaires

Do the above steps (6-7) two times (perform two tasks with two different datasets).

STEP 6: POST STUDY QUESTIONNAIRE

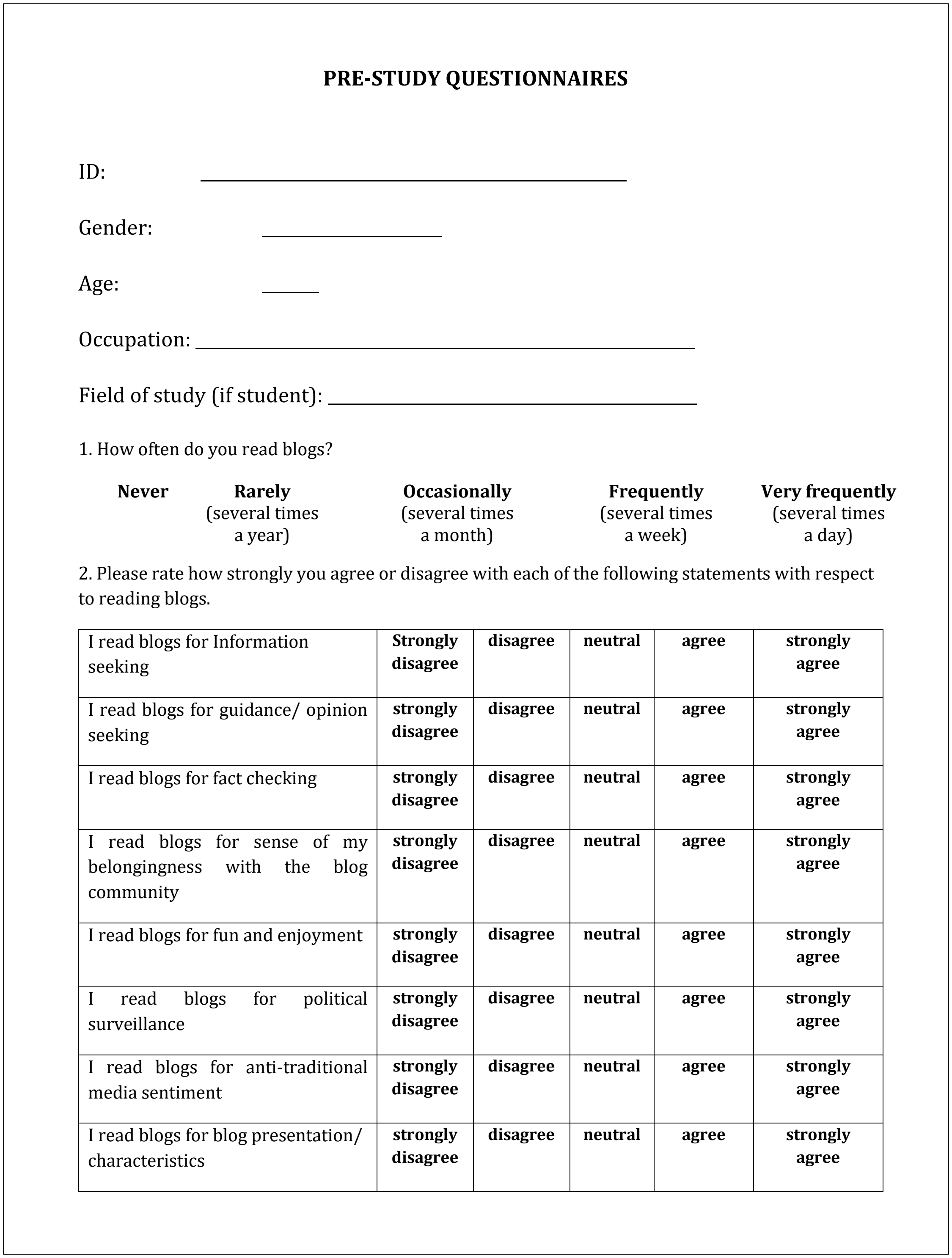
At the end of all the tasks, the participant will fill up a post-study questionnaires

STEP 7: DEBRIEFING

Tell Participant: “Thank you very much again for your participation. Would you have any other comments or questions?”

Action: Get Payment form signed

A.2 Questionnaires



PRE-STUDY QUESTIONNAIRES

ID: ______________________________

Gender: ______________

Age: ____

Occupation: ____________________________________

Field of study (if student): ___________________________

1. How often do you read blogs?

Never	Rarely	Occasionally	Frequently	Very frequently
	(several times a year)	(several times a month)	(several times a week)	(several times a day)

2. Please rate how strongly you agree or disagree with each of the following statements with respect to reading blogs.

I read blogs for Information seeking	**Strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for guidance/ opinion seeking	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for fact checking	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for sense of my belongingness with the blog community	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for fun and enjoyment	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for political surveillance	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for anti-traditional media sentiment	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for blog presentation/ characteristics	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**

3. What are the types of blogs you generally read?

□ Political

□ Sports

□ Business

□ Technology

□ Health

□ Personal

□ Others (Specify):___________________

4. How often do you comment on other people's blogs?

Never	**Rarely**	**Occasionally**	**Frequently**	**Very frequently**
	(several times a year)	(several times a month)	(several times a week)	(several times a day)

5. How often do you write your own blog (any type)?

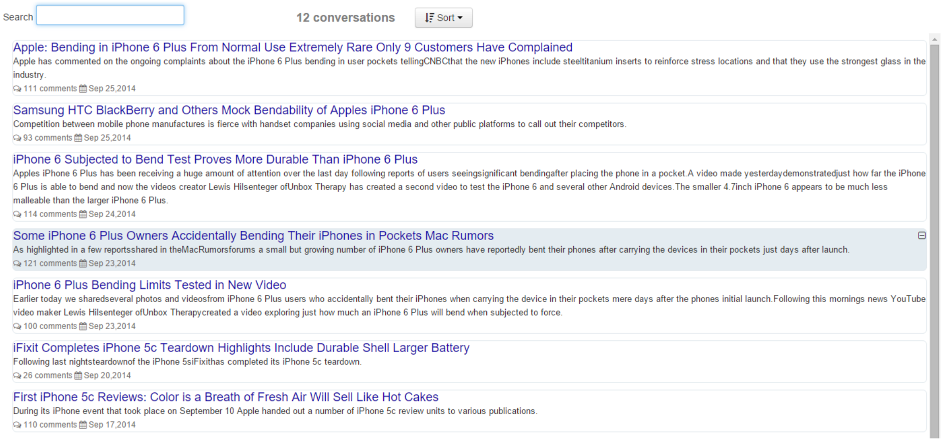
Never	**Rarely**	**Occasionally**	**Frequently**	**Very frequently**
	(several times a year)	(several times a month)	(several times a week)	(several times a day)

6. On an average how many blog conversations do you read in the same session?

1-2	**3-5**	**6-10**	**10-20**	**>20**

IN-STUDY QUESTIONNAIRES

Please rate how strongly you agree or disagree with each of the following statements with respect to the interface (A) you have used for exploring blog conversations.



Apple: Bending in iPhone 6 Plus From Normal Use Extremely Rare Only 9 Customers Have Complained
Apple has commented on the ongoing complaints about the iPhone 6 Plus bending in user pockets tellingCNBCthat the new iPhones include steeltitanium inserts to reinforce stress locations and that they use the strongest glass in the industry.
111 comments Sep 25,2014

Samsung HTC BlackBerry and Others Mock Bendability of Apples iPhone 6 Plus
Competition between mobile phone manufactures is fierce with handset companies using social media and other public platforms to call out their competitors.
93 comments Sep 25,2014

iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus
Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.
114 comments Sep 24,2014

Some iPhone 6 Plus Owners Accidentally Bending Their iPhones in Pockets Mac Rumors
As highlighted in a few reportsshared in theMacRumorsforums a small but growing number of iPhone 6 Plus owners have reportedly bent their phones after carrying the devices in their pockets just days after launch.
121 comments Sep 23,2014

iPhone 6 Plus Bending Limits Tested in New Video
Earlier today we sharedseveral photos and videosfrom iPhone 6 Plus users who accidentally bent their iPhones when carrying the device in their pockets mere days after the phones initial launch.Following this mornings news YouTube video maker Lewis Hilsenteger ofUnbox Therapycreated a video exploring just how much an iPhone 6 Plus will bend when subjected to force.
100 comments Sep 23,2014

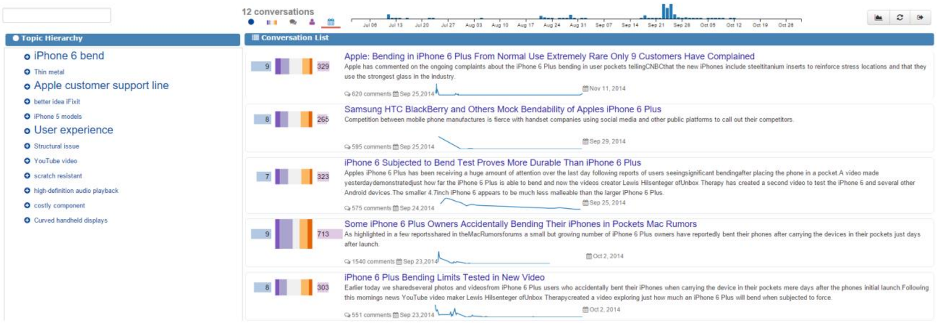
iFixit Completes iPhone 5c Teardown Highlights Include Durable Shell Larger Battery
Following last nightsteardownof the iPhone 5siFixithas completed its iPhone 5c teardown.
26 comments Sep 20,2014

First iPhone 5c Reviews: Color is a Breath of Fresh Air Will Sell Like Hot Cakes
During its iPhone event that took place on September 10 Apple handed out a number of iPhone 5c review units to various publications.
110 comments Sep 17,2014

I found this interface to be **useful** for browsing conversations	**Strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found this interface **easy to use**	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found this interface **enjoyable** to use	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **find the major points** that were discussed in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **find** more **insightful comments** in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **write a more informative summary** about the major points that were discussed in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**

IN-STUDY QUESTIONNAIRES

Please rate how strongly you agree or disagree with each of the following statements with respect to the interface (B) you have just used for exploring blog conversations.



9 329 Apple: Bending in iPhone 6 Plus From Normal Use Extremely Rare Only 9 Customers Have Complained
Apple has commented on the ongoing complaints about the iPhone 6 Plus bending in user pockets tellingCNBCthat the new iPhones include steeltitanium inserts to reinforce stress locations and that they use the strongest glass in the industry.
620 comments Sep 25,2014 Nov 11, 2014

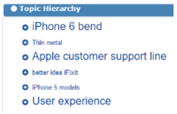
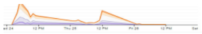
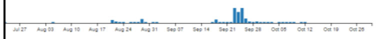
8 265 Samsung HTC BlackBerry and Others Mock Bendability of Apples iPhone 6 Plus
Competition between mobile phone manufactures is fierce with handset companies using social media and other public platforms to call out their competitors.
595 comments Sep 25,2014 Sep 29, 2014

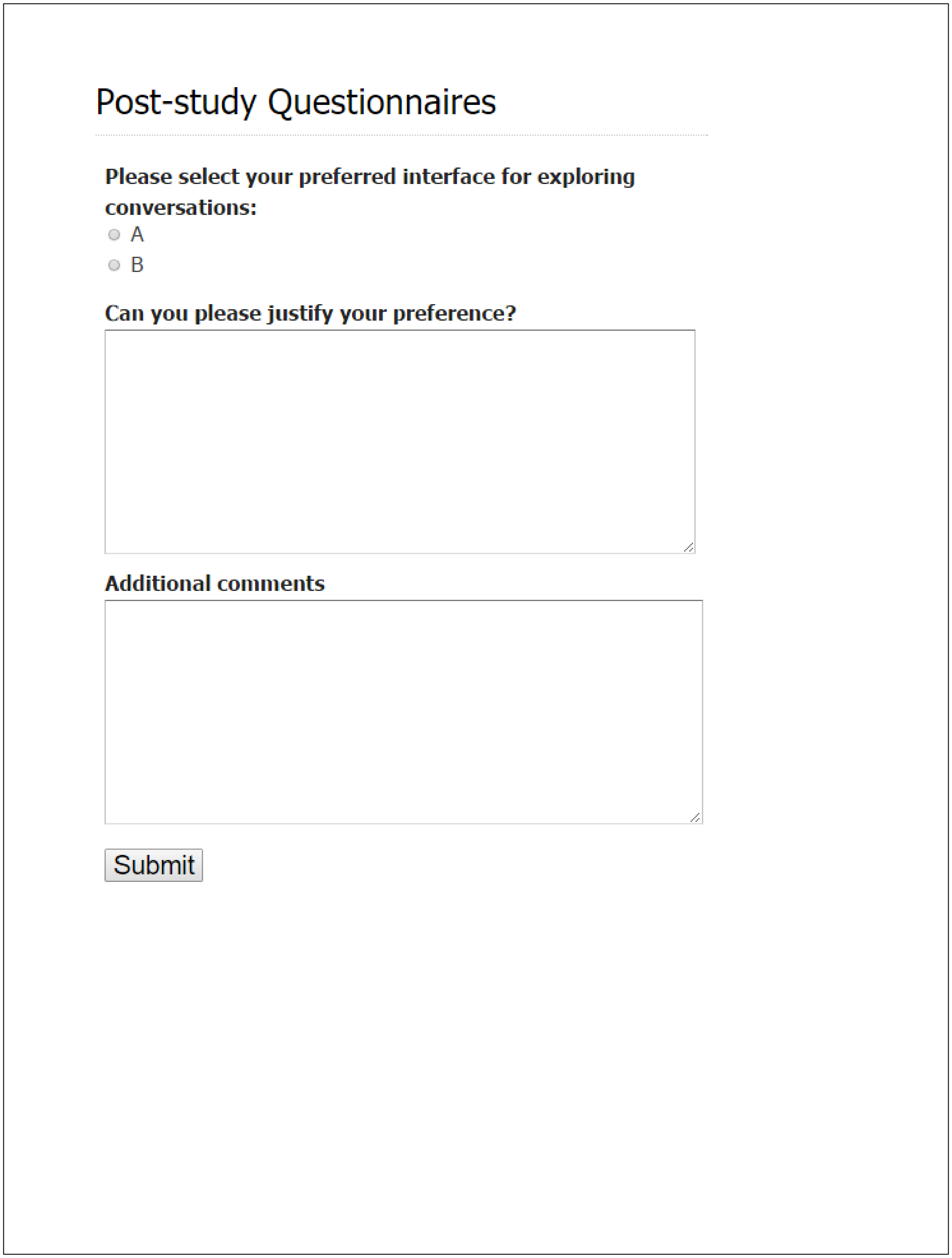
7 323 iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus
Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.
575 comments Sep 24,2014 Sep 25, 2014

9 713 Some iPhone 6 Plus Owners Accidentally Bending Their iPhones in Pockets Mac Rumors
As highlighted in a few reportsshared in theMacRumorsforums a small but growing number of iPhone 6 Plus owners have reportedly bent their phones after carrying the devices in their pockets just days after launch.
1540 comments Sep 23,2014 Oct 2, 2014

8 303 iPhone 6 Plus Bending Limits Tested in New Video
Earlier today we sharedseveral photos and videosfrom iPhone 6 Plus users who accidentally bent their iPhones when carrying the device in their pockets mere days after the phones initial launch.Following this mornings news YouTube video maker Lewis Hilsenteger ofUnbox Therapycreated a video exploring just how much an iPhone 6 Plus will bend when subjected to force.
551 comments Sep 23,2014 Oct 2, 2014

I found this interface to be **useful** for browsing conversations	**Strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found this interface **easy to use**	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found this interface **enjoyable** to use	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **find the major points** that were discussed in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **find** more **insightful comments** in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **write a better summary** about the major points that were discussed in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**

I found the **topic hierarchy** to be useful.  Topic Hierarchy iPhone 6 bend Thin metal Apple customer support line better idea iFixit iPhone 5 models User experience	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found **visual summary** of each conversation to be useful.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found the visual representation of **sentiment distribution over time** to be useful. 	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found the interactive feature for **filtering conversation by timeline** to be useful. 	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
The switching between **Conversation List** and **Conversation View** was easy to understand.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**



Post-study Questionnaires

Please select your preferred interface for exploring conversations:

○ A

○ B

Can you please justify your preference?

Additional comments

Appendix B

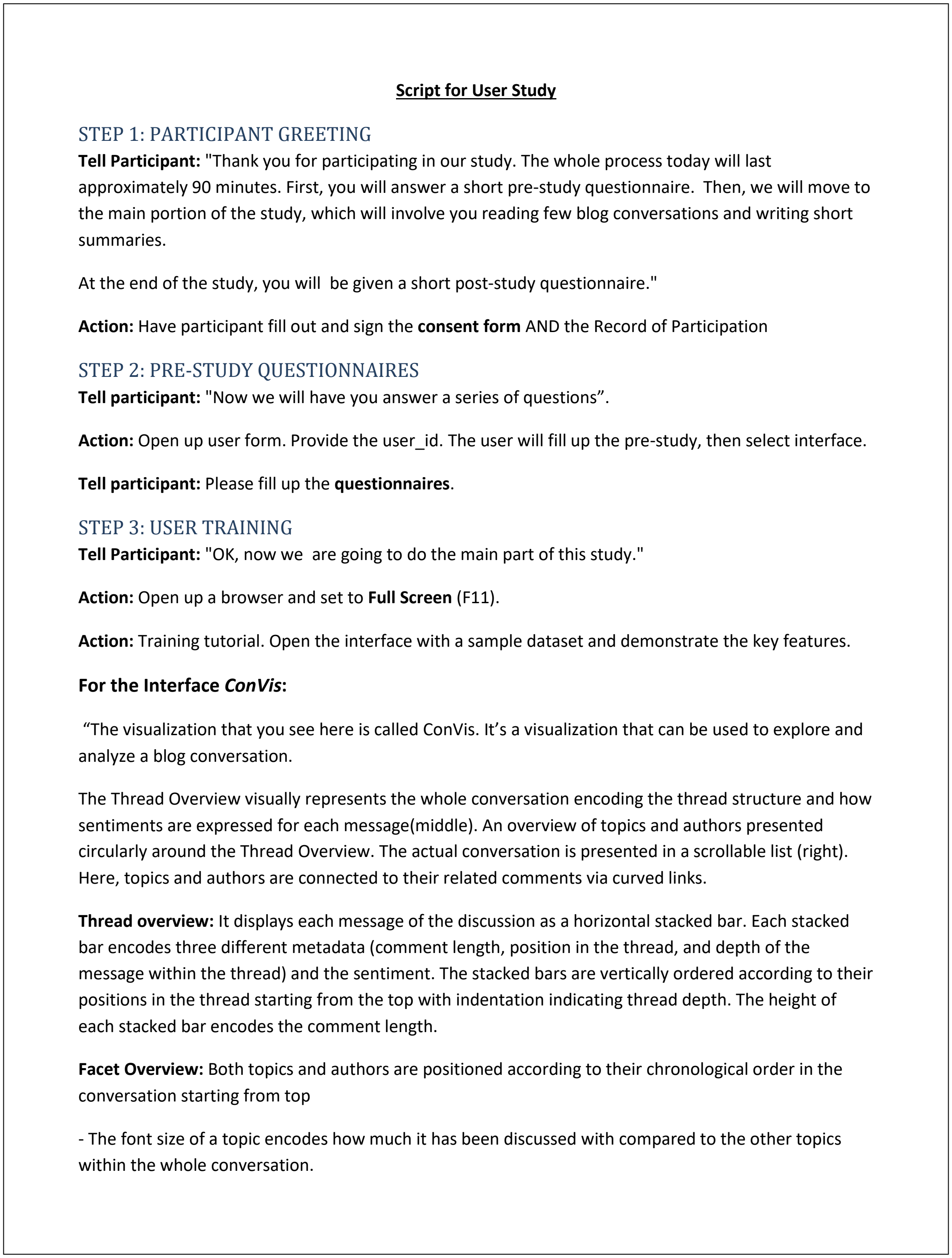
Supplemental Materials for Chapter 4

This appendix contains supplemental materials for Chapter 4, namely the script used by the experimenter to run the study, questionnaires that were used during the study, and the instructions for human raters who rated the user-generated summaries.

B.1 User Study 1

This section contains the documents for Study 1 as described in Section 4.6.1, where we compare between ConVisIT, ConVis, and a traditional interface.

B.1.1 Script for User Study



Script for User Study

STEP 1: PARTICIPANT GREETING

Tell Participant: "Thank you for participating in our study. The whole process today will last approximately 90 minutes. First, you will answer a short pre-study questionnaire. Then, we will move to the main portion of the study, which will involve you reading few blog conversations and writing short summaries.

At the end of the study, you will be given a short post-study questionnaire."

Action: Have participant fill out and sign the **consent form** AND the Record of Participation

STEP 2: PRE-STUDY QUESTIONNAIRES

Tell participant: "Now we will have you answer a series of questions”.

Action: Open up user form. Provide the user_id. The user will fill up the pre-study, then select interface.

Tell participant: Please fill up the **questionnaires**.

STEP 3: USER TRAINING

Tell Participant: "OK, now we are going to do the main part of this study."

Action: Open up a browser and set to **Full Screen** (F11).

Action: Training tutorial. Open the interface with a sample dataset and demonstrate the key features.

For the Interface ***ConVis***:

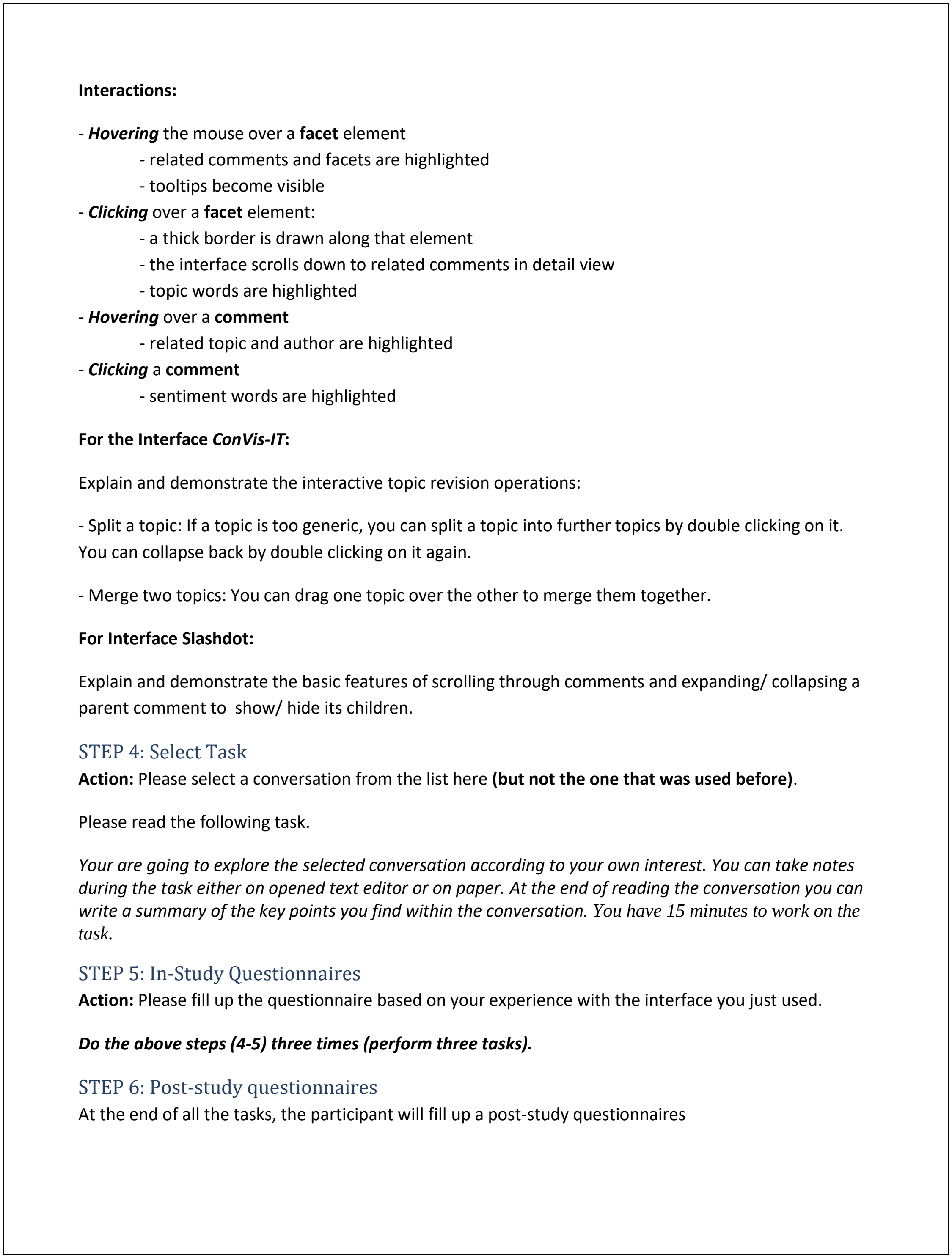
“The visualization that you see here is called ConVis. It’s a visualization that can be used to explore and analyze a blog conversation.

The Thread Overview visually represents the whole conversation encoding the thread structure and how sentiments are expressed for each message(middle). An overview of topics and authors presented circularly around the Thread Overview. The actual conversation is presented in a scrollable list (right). Here, topics and authors are connected to their related comments via curved links.

Thread overview: It displays each message of the discussion as a horizontal stacked bar. Each stacked bar encodes three different metadata (comment length, position in the thread, and depth of the message within the thread) and the sentiment. The stacked bars are vertically ordered according to their positions in the thread starting from the top with indentation indicating thread depth. The height of each stacked bar encodes the comment length.

Facet Overview: Both topics and authors are positioned according to their chronological order in the conversation starting from top

- The font size of a topic encodes how much it has been discussed with compared to the other topics within the whole conversation.



Interactions:

- ***Hovering*** the mouse over a **facet** element
    - related comments and facets are highlighted
    - tooltips become visible
- ***Clicking*** over a **facet** element:
    - a thick border is drawn along that element
    - the interface scrolls down to related comments in detail view
    - topic words are highlighted
- ***Hovering*** over a **comment**
    - related topic and author are highlighted
- ***Clicking*** a **comment**
    - sentiment words are highlighted

For the Interface *ConVis-IT*:

Explain and demonstrate the interactive topic revision operations:

- Split a topic: If a topic is too generic, you can split a topic into further topics by double clicking on it. You can collapse back by double clicking on it again.

- Merge two topics: You can drag one topic over the other to merge them together.

For Interface Slashdot:

Explain and demonstrate the basic features of scrolling through comments and expanding/ collapsing a parent comment to show/ hide its children.

STEP 4: Select Task

Action: Please select a conversation from the list here **(but not the one that was used before)**.

Please read the following task.

Your are going to explore the selected conversation according to your own interest. You can take notes during the task either on opened text editor or on paper. At the end of reading the conversation you can write a summary of the key points you find within the conversation. You have 15 minutes to work on the task.

STEP 5: In-Study Questionnaires

Action: Please fill up the questionnaire based on your experience with the interface you just used.

Do the above steps (4-5) three times (perform three tasks).

STEP 6: Post-study questionnaires

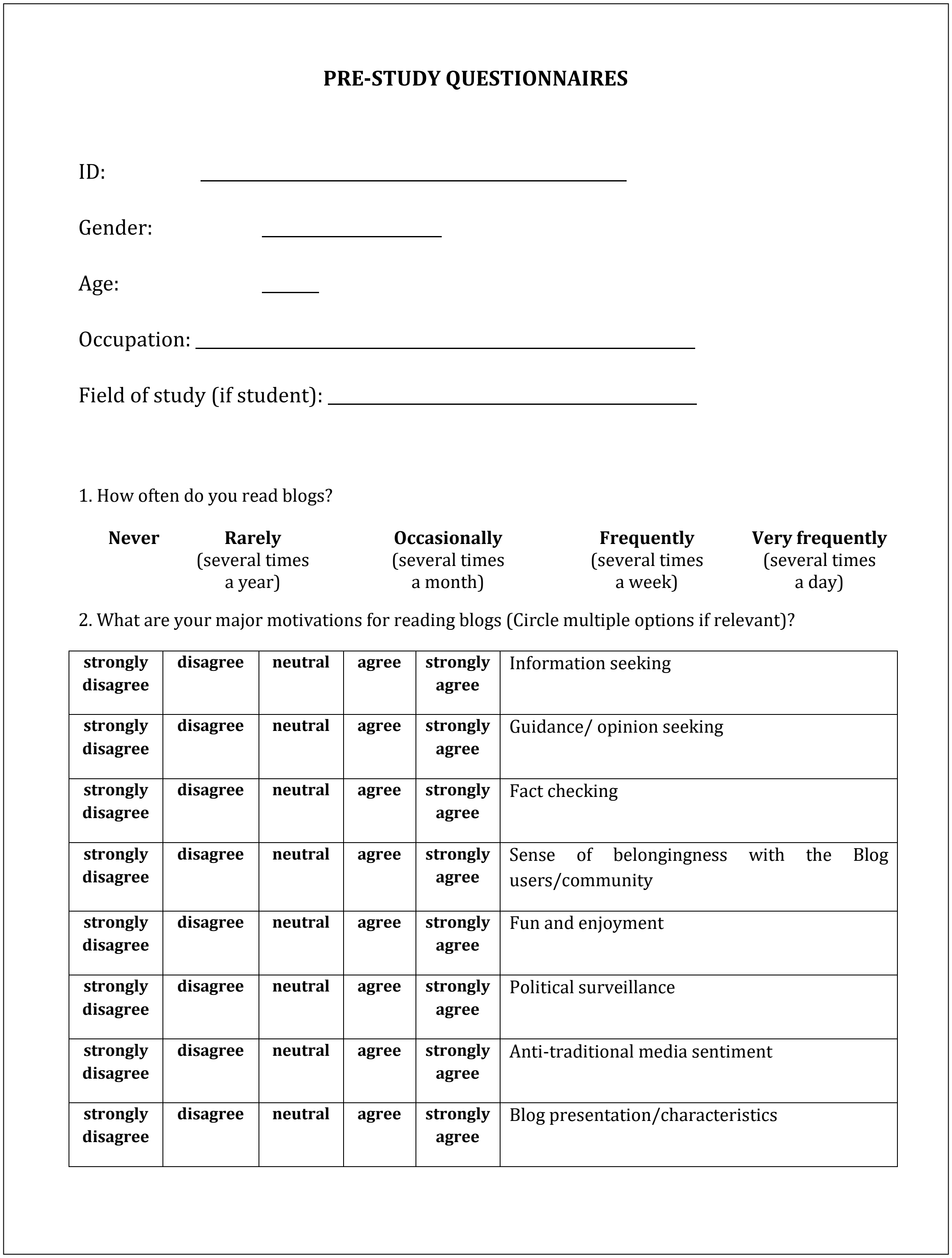
At the end of all the tasks, the participant will fill up a post-study questionnaires

STEP 7: DEBRIEFING

Tell Participant: “Thank you very much again for your participation. Would you have any other comments or questions?”

Action: Get Payment form signed

B.1.2 Questionnaires



PRE-STUDY QUESTIONNAIRES

ID: ____________________________

Gender: ____________

Age: ____

Occupation: ________________________________

Field of study (if student): ________________________

1. How often do you read blogs?

Never	**Rarely** (several times a year)	**Occasionally** (several times a month)	**Frequently** (several times a week)	**Very frequently** (several times a day)

2. What are your major motivations for reading blogs (Circle multiple options if relevant)?

strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Information seeking
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Guidance/ opinion seeking
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Fact checking
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Sense of belongingness with the Blog users/community
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Fun and enjoyment
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Political surveillance
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Anti-traditional media sentiment
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Blog presentation/characteristics

3. What are the types of blogs you generally read?

□ Political

□ Sports

□ Business

□ Technology

□ Health

□ Personal

□ Others (Specify):___________________

4. How often do you comment on other people's blog?

Never	**Rarely**	**Occasionally**	**Frequently**	**Very frequently**
	(several times	(several times	(several times	(several times
	a year)	a month)	a week)	a day)

5. How often do you write blog (any type)?

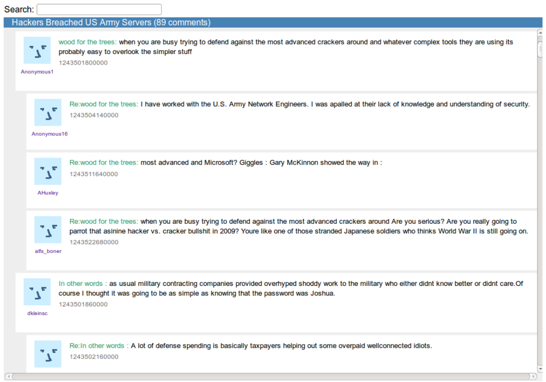
Never	**Rarely**	**Occasionally**	**Frequently**	**Very frequently**
	(several times	(several times	(several times	(several times
	a year)	a month)	a week)	a day)

6. How often do you use blogs to make a decision/ choice?

Never	**Rarely**	**Occasionally**	**Frequently**	**Very frequently**
	(several times	(several times	(several times	(several times
	a year)	a month)	a week)	a day)

IN-STUDY QUESTIONNAIRES

Please rate how strongly you agree or disagree with each of the following statements with respect to the interface (Slashdot) you have used for exploring blog conversations.



Hackers Breached US Army Servers (89 comments)

wood for the trees: when you are busy trying to defend against the most advanced crackers around and whatever complex tools they are using its probably easy to overlook the simpler stuff
1243501800000
Anonymous1

Re:wood for the trees: I have worked with the U.S. Army Network Engineers. I was apalled at their lack of knowledge and understanding of security.
1243504140000
Anonymous16

Re:wood for the trees: most advanced and Microsoft? Giggles : Gary McKinnon showed the way in :
1243511640000
AHuxley

Re:wood for the trees: when you are busy trying to defend against the most advanced crackers around Are you serious? Are you really going to parrot that asinine hacker vs. cracker bullshit in 2009? Youre like one of those stranded Japanese soldiers who thinks World War II is still going on.
1243522680000
alfs_boner

In other words : as usual military contracting companies provided overhyped shoddy work to the military who either didnt know better or didnt care.Of course I thought it was going to be as simple as knowing that the password was Joshua.
1243501860000
dkleinsc

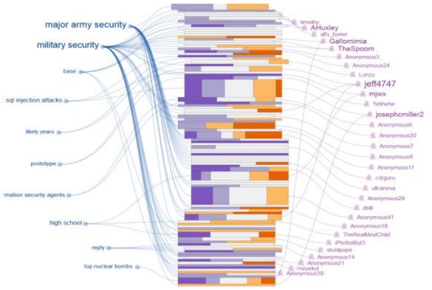
Re:In other words : A lot of defense spending is basically taxpayers helping out some overpaid wellconnected idiots.
1243502160000

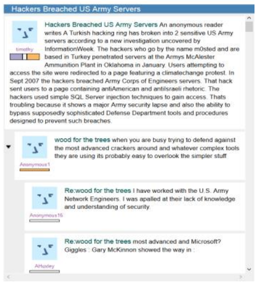
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found this interface to be **useful** for browsing conversations
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found this interface **easy to use**
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found this interface **enjoyable** to use
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	This interface enabled me to find more **insightful** comments

strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found the indented-list representation of the conversation to be useful
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Scrolling through the long conversation is useful for finding more insightful comments.
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	Showing the detailed comments only (without any overview) is useful for browsing conversations

IN-STUDY QUESTIONNAIRES

Please rate how strongly you agree or disagree with each of the following statements with respect to the interface (ConVis) you have used for exploring blog conversations.




Hackers Breached US Army Servers

timothy

Hackers Breached US Army Servers An anonymous reader writes A Turkish hacking ring has broken into 2 sensitive US Army servers according to a new investigation uncovered by InformationWeek. The hackers who go by the name m0sted and are based in Turkey penetrated servers at the Armys McAlester Ammunition Plant in Oklahoma in January. Users attempting to access the site were redirected to a page featuring a climatechange protest. In Sept 2007 the hackers breached Army Corps of Engineers servers. That hack sent users to a page containing antiAmerican and antiIsraeli rhetoric. The hackers used simple SQL Server injection techniques to gain access. Thats troubling because it shows a major Army security lapse and also the ability to bypass supposedly sophisticated Defense Department tools and procedures designed to prevent such breaches.

Anonymous1

wood for the trees when you are busy trying to defend against the most advanced crackers around and whatever complex tools they are using its probably easy to overlook the simpler stuff

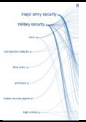
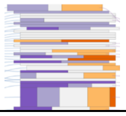
Anonymous16

Re:wood for the trees I have worked with the U.S. Army Network Engineers. I was apalled at their lack of knowledge and understanding of security.

AHuxley

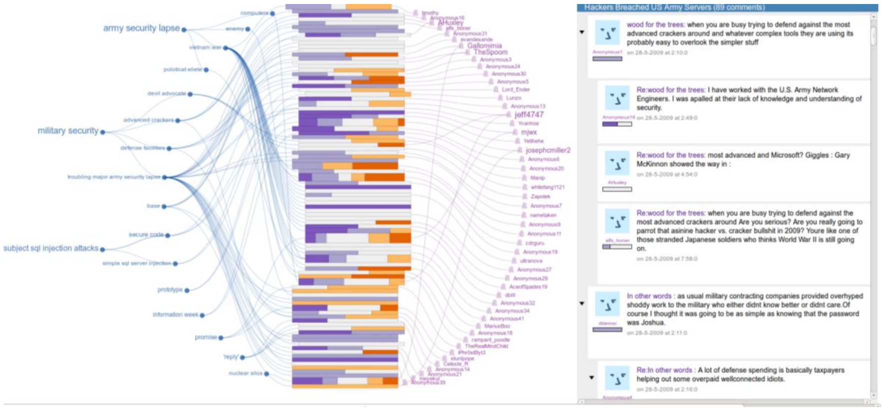
Re:wood for the trees most advanced and Microsoft? Giggles : Gary McKinnon showed the way in :

strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found this interface to be **useful** for browsing conversations
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found this interface **easy to use**
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found this interface **enjoyable** to use
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	This interface enabled me to find more **insightful** comments

strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**		I found the visual representation of the discussion topic/ author to be useful
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**		I found visual representation of the thread overview to be useful
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**		I found the highlighting the relations between topic and author to be useful
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**		I found the selection of comments based on topic/ author to be useful for navigating long conversation

IN-STUDY QUESTIONNAIRES

Please rate how strongly you agree or disagree with each of the following statements with respect to the interface (ConVisIT) you have used for exploring blog conversations.



Hackers Breached US Army Servers (89 comments)

wood for the trees: when you are busy trying to defend against the most advanced crackers around and whatever complex tools they are using its probably easy to overlook the simpler stuff
on 28-5-2009 at 2:10:0

Re:wood for the trees: I have worked with the U.S. Army Network Engineers. I was apalled at their lack of knowledge and understanding of security.
on 28-5-2009 at 2:49:0

Re:wood for the trees: most advanced and Microsoft? Giggles : Gary McKinnon showed the way in :
on 28-5-2009 at 4:54:0

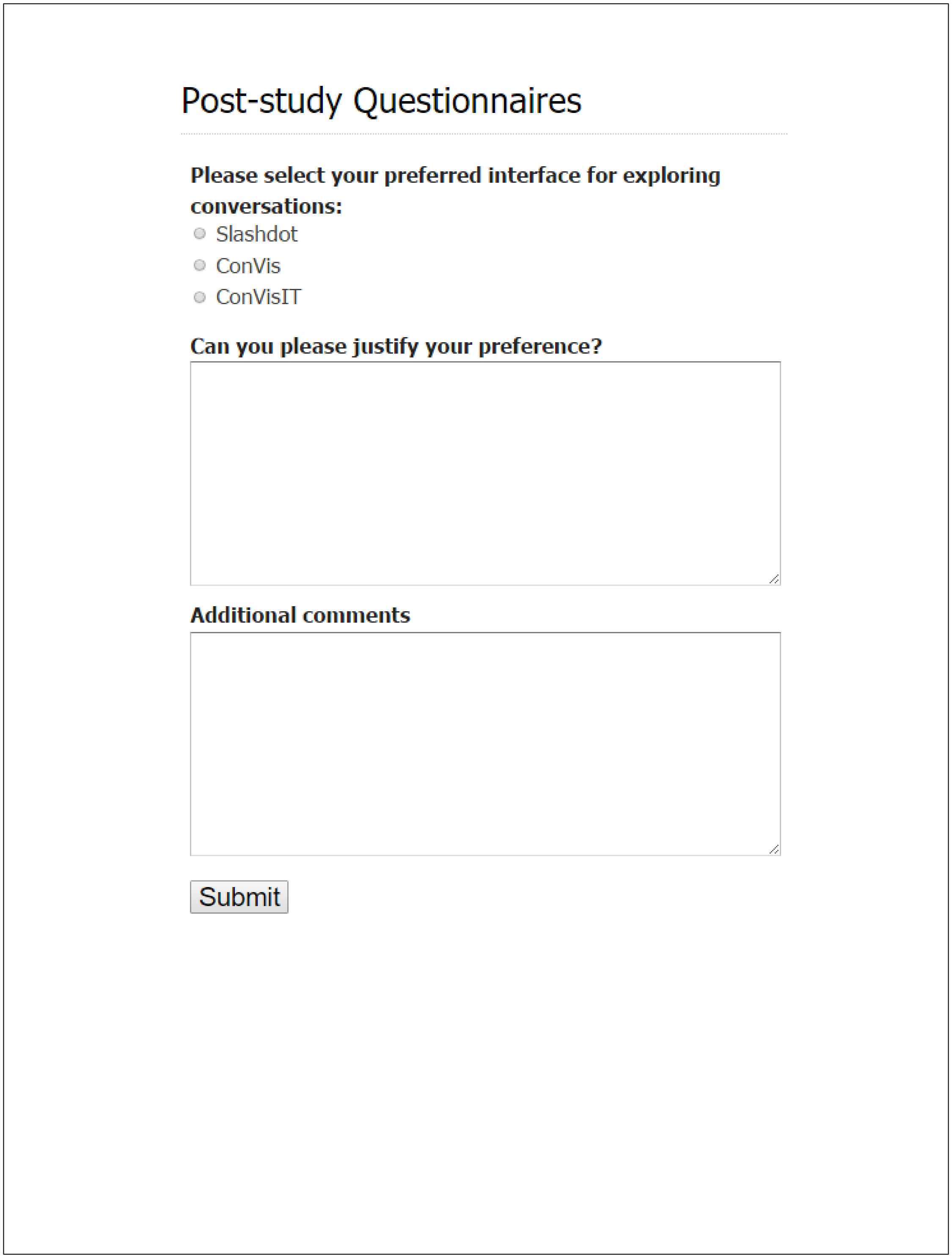
Re:wood for the trees: when you are busy trying to defend against the most advanced crackers around Are you serious? Are you really going to parrot that asinine hacker vs. cracker bullshit in 2009? Youre like one of those stranded Japanese soldiers who thinks World War II is still going on.
on 28-5-2009 at 7:58:0

In other words : as usual military contracting companies provided overhyped shoddy work to the military who either didnt know better or didnt care.Of course I thought it was going to be as simple as knowing that the password was Joshua.
on 28-5-2009 at 2:11:0

Re:In other words : A lot of defense spending is basically taxpayers helping out some overpaid wellconnected idiots.
on 28-5-2009 at 2:16:0

strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found this interface to be useful for browsing conversations
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found this interface **easy to use**
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found this interface **enjoyable** to use
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	This interface enabled me to find more **insightful comments**

strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found the feature of **splitting** a topic further into sub-topics to be useful.
strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**	I found the feature of **merging** topics together to be useful.



Post-study Questionnaires

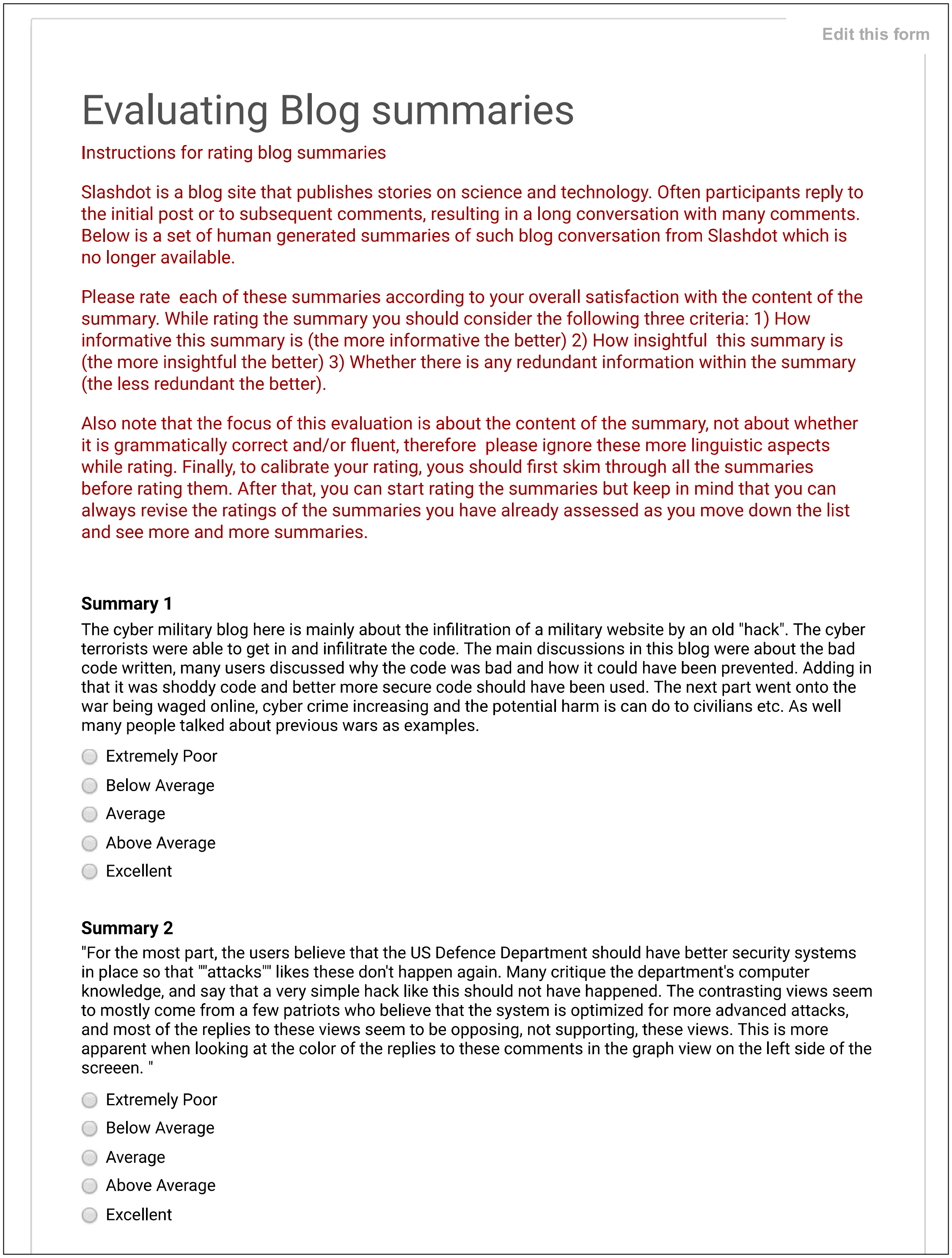
Please select your preferred interface for exploring conversations:

○ Slashdot
○ ConVis
○ ConVisIT

Can you please justify your preference?

Additional comments

B.1.3 Evaluating User-Generated Summary by Human Raters



Evaluating Blog summaries

Instructions for rating blog summaries

Slashdot is a blog site that publishes stories on science and technology. Often participants reply to the initial post or to subsequent comments, resulting in a long conversation with many comments. Below is a set of human generated summaries of such blog conversation from Slashdot which is no longer available.

Please rate each of these summaries according to your overall satisfaction with the content of the summary. While rating the summary you should consider the following three criteria: 1) How informative this summary is (the more informative the better) 2) How insightful this summary is (the more insightful the better) 3) Whether there is any redundant information within the summary (the less redundant the better).

Also note that the focus of this evaluation is about the content of the summary, not about whether it is grammatically correct and/or fluent, therefore please ignore these more linguistic aspects while rating. Finally, to calibrate your rating, yous should first skim through all the summaries before rating them. After that, you can start rating the summaries but keep in mind that you can always revise the ratings of the summaries you have already assessed as you move down the list and see more and more summaries.

Summary 1

The cyber military blog here is mainly about the infilitration of a military website by an old "hack". The cyber terrorists were able to get in and infilitrate the code. The main discussions in this blog were about the bad code written, many users discussed why the code was bad and how it could have been prevented. Adding in that it was shoddy code and better more secure code should have been used. The next part went onto the war being waged online, cyber crime increasing and the potential harm is can do to civilians etc. As well many people talked about previous wars as examples.

○ Extremely Poor

○ Below Average

○ Average

○ Above Average

○ Excellent

Summary 2

"For the most part, the users believe that the US Defence Department should have better security systems in place so that ""attacks"" likes these don't happen again. Many critique the department's computer knowledge, and say that a very simple hack like this should not have happened. The contrasting views seem to mostly come from a few patriots who believe that the system is optimized for more advanced attacks, and most of the replies to these views seem to be opposing, not supporting, these views. This is more apparent when looking at the color of the replies to these comments in the graph view on the left side of the screeen. "

○ Extremely Poor

○ Below Average

○ Average

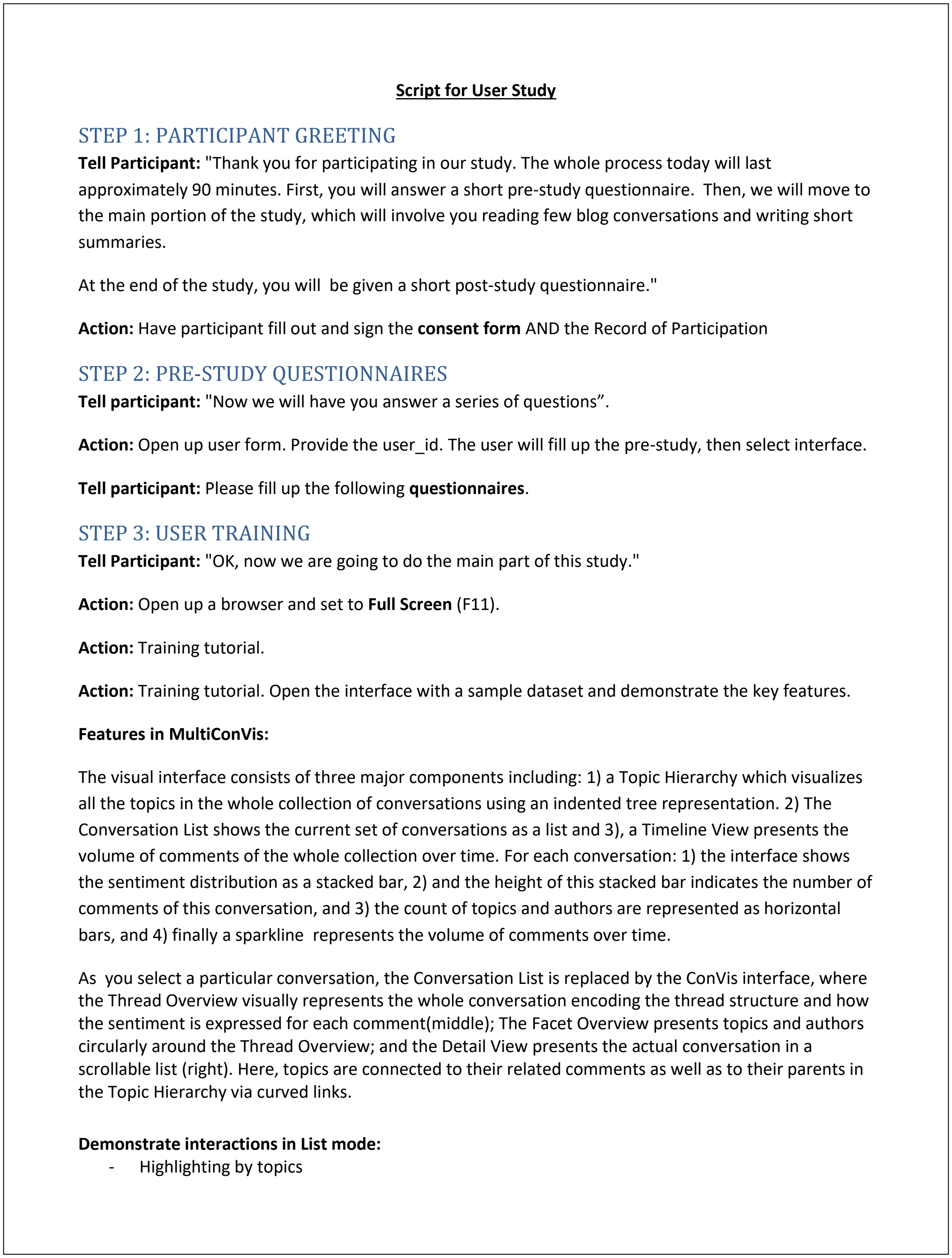
○ Above Average

○ Excellent

B.2 User Study 2

This section contains the documents for Study 2 as described in Section 4.6.2, where we compare between ConVisIT, ConVis, and a traditional interface.

B.2.1 Script for User Study



Script for User Study

STEP 1: PARTICIPANT GREETING

Tell Participant: "Thank you for participating in our study. The whole process today will last approximately 90 minutes. First, you will answer a short pre-study questionnaire. Then, we will move to the main portion of the study, which will involve you reading few blog conversations and writing short summaries.

At the end of the study, you will be given a short post-study questionnaire."

Action: Have participant fill out and sign the **consent form** AND the Record of Participation

STEP 2: PRE-STUDY QUESTIONNAIRES

Tell participant: "Now we will have you answer a series of questions".

Action: Open up user form. Provide the user_id. The user will fill up the pre-study, then select interface.

Tell participant: Please fill up the following **questionnaires**.

STEP 3: USER TRAINING

Tell Participant: "OK, now we are going to do the main part of this study."

Action: Open up a browser and set to **Full Screen** (F11).

Action: Training tutorial.

Action: Training tutorial. Open the interface with a sample dataset and demonstrate the key features.

Features in MultiConVis:

The visual interface consists of three major components including: 1) a Topic Hierarchy which visualizes all the topics in the whole collection of conversations using an indented tree representation. 2) The Conversation List shows the current set of conversations as a list and 3), a Timeline View presents the volume of comments of the whole collection over time. For each conversation: 1) the interface shows the sentiment distribution as a stacked bar, 2) and the height of this stacked bar indicates the number of comments of this conversation, and 3) the count of topics and authors are represented as horizontal bars, and 4) finally a sparkline represents the volume of comments over time.

As you select a particular conversation, the Conversation List is replaced by the ConVis interface, where the Thread Overview visually represents the whole conversation encoding the thread structure and how the sentiment is expressed for each comment(middle); The Facet Overview presents topics and authors circularly around the Thread Overview; and the Detail View presents the actual conversation in a scrollable list (right). Here, topics are connected to their related comments as well as to their parents in the Topic Hierarchy via curved links.

Demonstrate interactions in List mode:

- Highlighting by topics

- Expand/ collapse topics
- Sorting conversations
- Click timeline button to show sentiment over time
- Filter by time

Demonstrate interactions in Conversation Mode:

- ***Hovering*** the mouse over a **facet** element
    - related comments and facets are highlighted
    - tooltips become visible
- ***Clicking*** over a **facet** element:
    - a thick border is drawn along that element
    - the interface scrolls down to related comments in detail view
    - topic words are highlighted
- ***Hovering*** over a **comment**
    - related topic and author are highlighted
- ***Clicking*** a **comment**
    - sentiment words are highlighted
- ***Adding summary to topics:***
    - You can summarize the keypoints that were discussed about a topic, by clicking on the summary icon of that topic node.

Additional features in MultiConVisIT:

Explain topic hierarchy revision features:

You can revise the topic hierarchy presented here according to your own needs. For example, if you think that the current topic is too broad, you can ask the system to show fewer more generic children nodes, by either double clicking or selecting the menu by right clicking on it. You can also remove this additional level of children by double clicking on the parent topic.

You can also change the topic assignment by dragging one topic over another topic. There are two ways to do this: 1) you can merge a topic as a sibling to another topic, 2) you can place it as a child to another topic.

If you feel that a topic is less relevant, or doesn't make any sense, you can drag that topic to the recycle bin.

You can also rename a topic if the current topic does not represent its corresponding textual comments. This is critical for creating a more informative summary because the name of the topic needs to match the major discussion points of the summaries.

Finally, at any time you can undo the last topic revision operation you have made, by clicking on the undo button.

STEP 4: Select Task

Please read the following task.

For Dataset iPhone bending:

The issue of ***'iPhone bending'*** *went viral on social media after the iPhone 6 was launched in September 2014. Soon after the product was released, some people claimed that this new phone can easily bend in the pocket while sitting on it. This incident triggered a huge amount of discussions in Macrumors, a blog site that regularly publishes Apple related news and allows participants to make comments.*

You are working for Apple as a business analyst. Your task is to find the major discussion points about the iPhone bending issue and summarize each of them under the most appropriate corresponding topic. The final outcome will be a summary of conversations organized according to a topic hierarchy that you will have to show and discuss with your colleagues. So you want to make sure that the topic hierarchy and the summary of major discussion points are as informative and as clear as possible.

You have 30 minutes to work on the task.

For Dataset iPad release:

The iPad air 2 was launched in October 2014. This event triggered a huge amount of discussions in Macrumors, a blog site that regularly publishes Apple related news and allows participants to make comments.

You are working for Apple as a business analyst. Your task is to find the major discussion points about the iPad release issue and summarize each of them under the most appropriate corresponding topic. The final outcome will be a summary of conversations organized according to a topic hierarchy that you will have to show and discuss with your colleagues. So you want to make sure that the topic hierarchy and the summary of major discussion points are as informative and as clear as possible.

You have 30 minutes to work on the task.

STEP 5: IN-STUDY QUESTIONNAIRE

After each task, the participant will fill up a set of in-study questionnaires

Do the above steps (6-7) two times (perform two tasks with two different datasets).

STEP 6: POST STUDY QUESTIONNAIRE

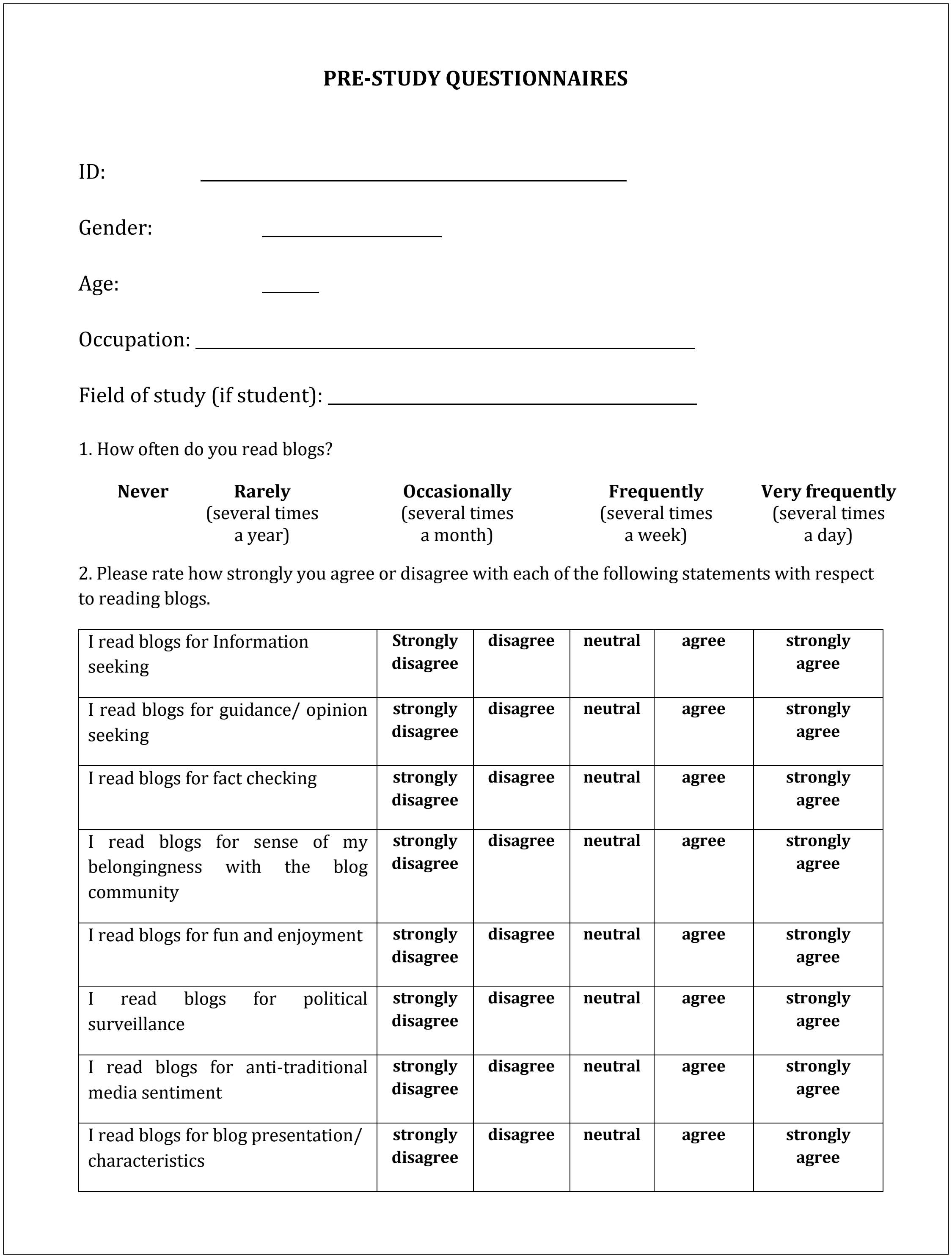
At the end of all the tasks, the participant will fill up post-study questionnaires

STEP 7: DEBRIEFING

Tell Participant: "Thank you very much again for your participation. Would you have any other comments or questions?"

Action: Get Payment form signed

B.2.2 Questionnaires



PRE-STUDY QUESTIONNAIRES

ID: ____

Gender: _________________

Age: _____

Occupation: ____

Field of study (if student): ___________________________________

1. How often do you read blogs?

Never	**Rarely** (several times a year)	**Occasionally** (several times a month)	**Frequently** (several times a week)	**Very frequently** (several times a day)

2. Please rate how strongly you agree or disagree with each of the following statements with respect to reading blogs.

I read blogs for Information seeking	**Strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for guidance/ opinion seeking	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for fact checking	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for sense of my belongingness with the blog community	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for fun and enjoyment	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for political surveillance	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for anti-traditional media sentiment	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I read blogs for blog presentation/ characteristics	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**

3. What are the types of blogs you generally read?

□ Political

□ Sports

□ Business

□ Technology

□ Health

□ Personal

□ Others (Specify):____________________

4. How often do you comment on other people's blogs?

Never	**Rarely**	**Occasionally**	**Frequently**	**Very frequently**
	(several times	(several times	(several times	(several times
	a year)	a month)	a week)	a day)

5. How often do you write your own blog (any type)?

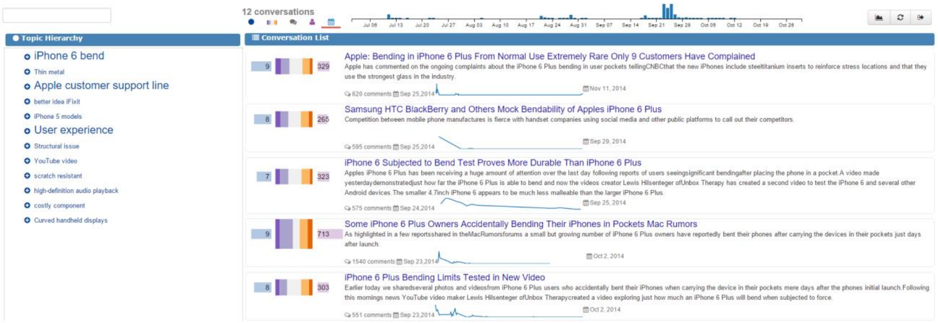
Never	**Rarely**	**Occasionally**	**Frequently**	**Very frequently**
	(several times	(several times	(several times	(several times
	a year)	a month)	a week)	a day)

6. On an average how many blog conversations do you read in the same session?

1-2	**3-5**	**6-10**	**10-20**	**>20**

IN-STUDY QUESTIONNAIRES

Please rate how strongly you agree or disagree with each of the following statements with respect to the interface (A) you have just used for exploring blog conversations.



- iPhone 6 bend
- Thin metal
- Apple customer support line
- better idea iFixit
- iPhone 5 models
- User experience
- Structural issue
- YouTube video
- scratch resistant
- high-definition audio playback
- costly component
- Curved handheld displays

9 329 Apple: Bending in iPhone 6 Plus From Normal Use Extremely Rare Only 9 Customers Have Complained
Apple has commented on the ongoing complaints about the iPhone 6 Plus bending in user pockets tellingCNBCthat the new iPhones include steeltitanium inserts to reinforce stress locations and that they use the strongest glass in the industry.
620 comments Sep 25,2014 Nov 11, 2014

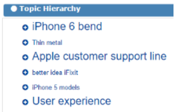
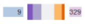
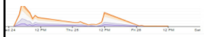
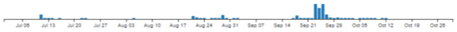
8 265 Samsung HTC BlackBerry and Others Mock Bendability of Apples iPhone 6 Plus
Competition between mobile phone manufactures is fierce with handset companies using social media and other public platforms to call out their competitors.
595 comments Sep 25,2014 Sep 29, 2014

7 323 iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus
Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.
575 comments Sep 24,2014 Sep 25, 2014

9 713 Some iPhone 6 Plus Owners Accidentally Bending Their iPhones in Pockets Mac Rumors
As highlighted in a few reportsshared in theMacRumorsforums a small but growing number of iPhone 6 Plus owners have reportedly bent their phones after carrying the devices in their pockets just days after launch.
1540 comments Sep 23,2014 Oct 2, 2014

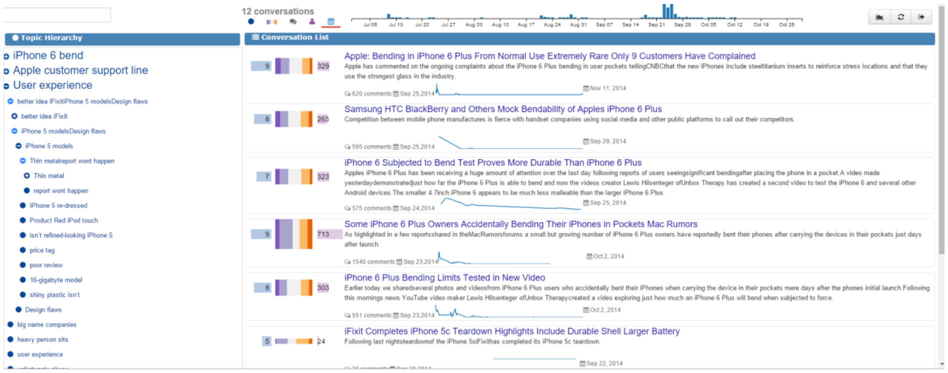
8 303 iPhone 6 Plus Bending Limits Tested in New Video
Earlier today we sharedseveral photos and videosfrom iPhone 6 Plus users who accidentally bent their iPhones when carrying the device in their pockets mere days after the phones initial launch.Following this mornings news YouTube video maker Lewis Hilsenteger ofUnbox Therapycreated a video exploring just how much an iPhone 6 Plus will bend when subjected to force.
551 comments Sep 23,2014 Oct 2, 2014

I found this interface to be **useful** for browsing conversations	**Strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found this interface **easy to use**	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found this interface **enjoyable** to use	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **find the major points** that were discussed in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **find** more **insightful comments** in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **write a better summary** about the major points that were discussed in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**

I found the **topic hierarchy** to be useful.  Topic Hierarchy iPhone 6 bend Thin metal Apple customer support line better idea IFixit iPhone 5 models User experience	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found **visual summary** of each conversation to be useful. 	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found the visual representation of **sentiment distribution over time** to be useful.  	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found the interactive feature for **filtering conversation by timeline** to be useful.  	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
The switching between **Conversation List** and **Conversation View** was easy to understand.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**

IN-STUDY QUESTIONNAIRES

Please rate how strongly you agree or disagree with each of the following statements with respect to the interface (B) you have just used for exploring blog conversations.



- iPhone 6 bend
- Apple customer support line
- User experience
  - better idea iFixitiPhone 5 modelsDesign flaws
    - better idea iFixit
    - iPhone 5 modelsDesign flaws
      - iPhone 5 models
        - Thin metalreport wont happen
          - Thin metal
          - report wont happen
        - iPhone 5 re-dressed
        - Product Red iPod touch
        - isn't refined-looking iPhone 5
        - price tag
        - poor review
        - 16-gigabyte model
        - shiny plastic isn't
      - Design flaws
  - big name companies
  - heavy person sits
  - user experience

9 | 329
Apple: Bending in iPhone 6 Plus From Normal Use Extremely Rare Only 9 Customers Have Complained
Apple has commented on the ongoing complaints about the iPhone 6 Plus bending in user pockets tellingCNBCthat the new iPhones include steeltitanium inserts to reinforce stress locations and that they use the strongest glass in the industry.
620 comments Sep 25,2014 Nov 11, 2014

8 | 265
Samsung HTC BlackBerry and Others Mock Bendability of Apples iPhone 6 Plus
Competition between mobile phone manufactures is fierce with handset companies using social media and other public platforms to call out their competitors.
595 comments Sep 25,2014 Sep 29, 2014

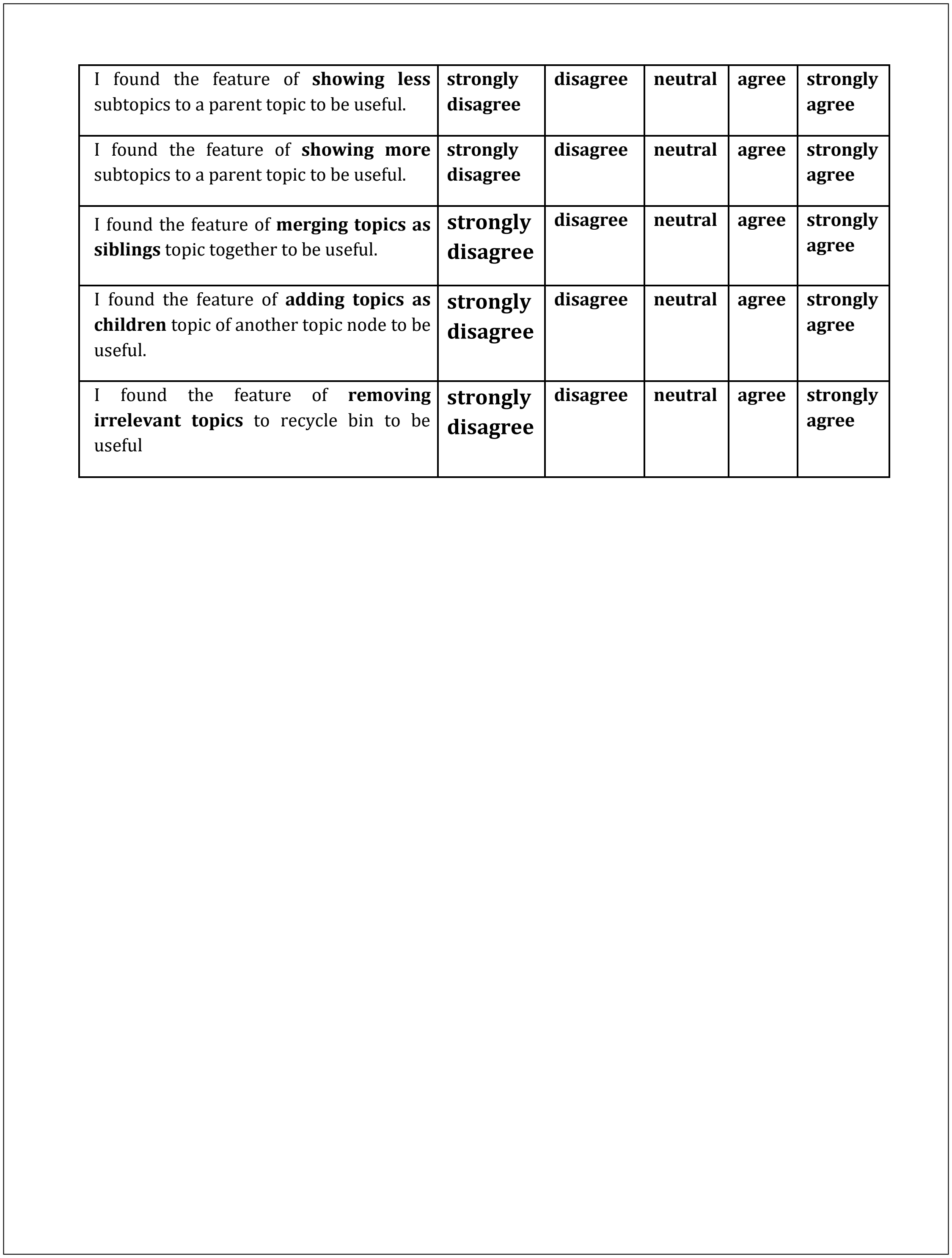
7 | 323
iPhone 6 Subjected to Bend Test Proves More Durable Than iPhone 6 Plus
Apples iPhone 6 Plus has been receiving a huge amount of attention over the last day following reports of users seeingsignificant bendingafter placing the phone in a pocket.A video made yesterdaydemonstratedjust how far the iPhone 6 Plus is able to bend and now the videos creator Lewis Hilsenteger ofUnbox Therapy has created a second video to test the iPhone 6 and several other Android devices.The smaller 4.7inch iPhone 6 appears to be much less malleable than the larger iPhone 6 Plus.
575 comments Sep 24,2014 Sep 25, 2014

9 | 713
Some iPhone 6 Plus Owners Accidentally Bending Their iPhones in Pockets Mac Rumors
As highlighted in a few reportsshared in theMacRumorsforums a small but growing number of iPhone 6 Plus owners have reportedly bent their phones after carrying the devices in their pockets just days after launch.
1540 comments Sep 23,2014 Oct 2, 2014

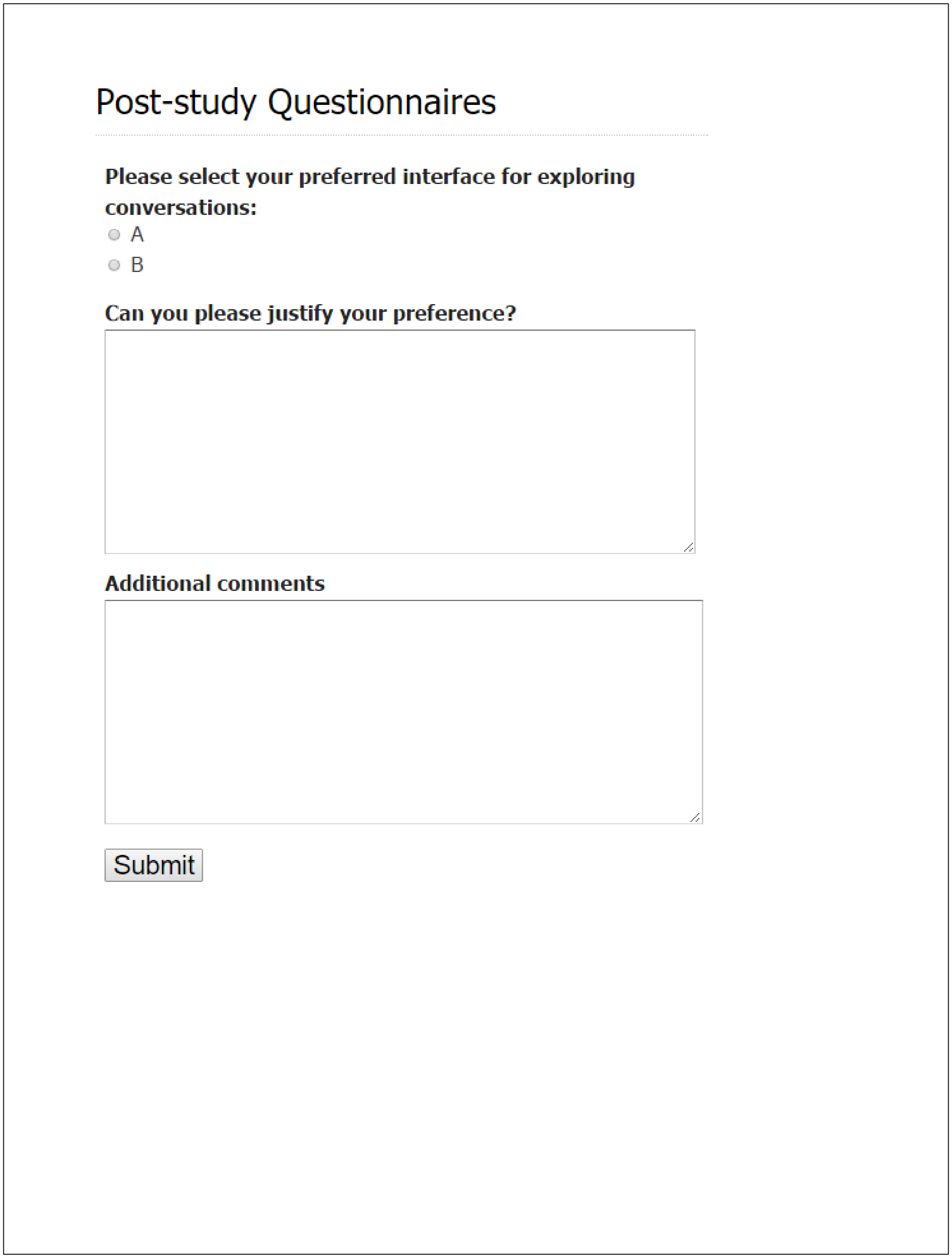
8 | 303
iPhone 6 Plus Bending Limits Tested in New Video
Earlier today we sharedseveral photos and videosfrom iPhone 6 Plus users who accidentally bent their iPhones when carrying the device in their pockets mere days after the phones initial launch.Following this mornings news YouTube video maker Lewis Hilsenteger ofUnbox Therapycreated a video exploring just how much an iPhone 6 Plus will bend when subjected to force.
551 comments Sep 23,2014 Oct 2, 2014

5 | 24
iFixit Completes iPhone 5c Teardown Highlights Include Durable Shell Larger Battery
Following last nightsteardownof the iPhone 5siFixithas completed its iPhone 5c teardown.
Sep 22, 2014

I found this interface to be **useful** for browsing conversations	**Strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found this interface **easy to use**	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found this interface **enjoyable** to use	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **find the major points** that were discussed in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **find** more **insightful comments** in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
This interface enabled me to **write a better summary** about the major points that were discussed in the set of conversations.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**



I found the feature of **showing less** subtopics to a parent topic to be useful.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found the feature of **showing more** subtopics to a parent topic to be useful.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found the feature of **merging topics as siblings** topic together to be useful.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found the feature of **adding topics as children** topic of another topic node to be useful.	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**
I found the feature of **removing irrelevant topics** to recycle bin to be useful	**strongly disagree**	**disagree**	**neutral**	**agree**	**strongly agree**



Post-study Questionnaires

Please select your preferred interface for exploring conversations:

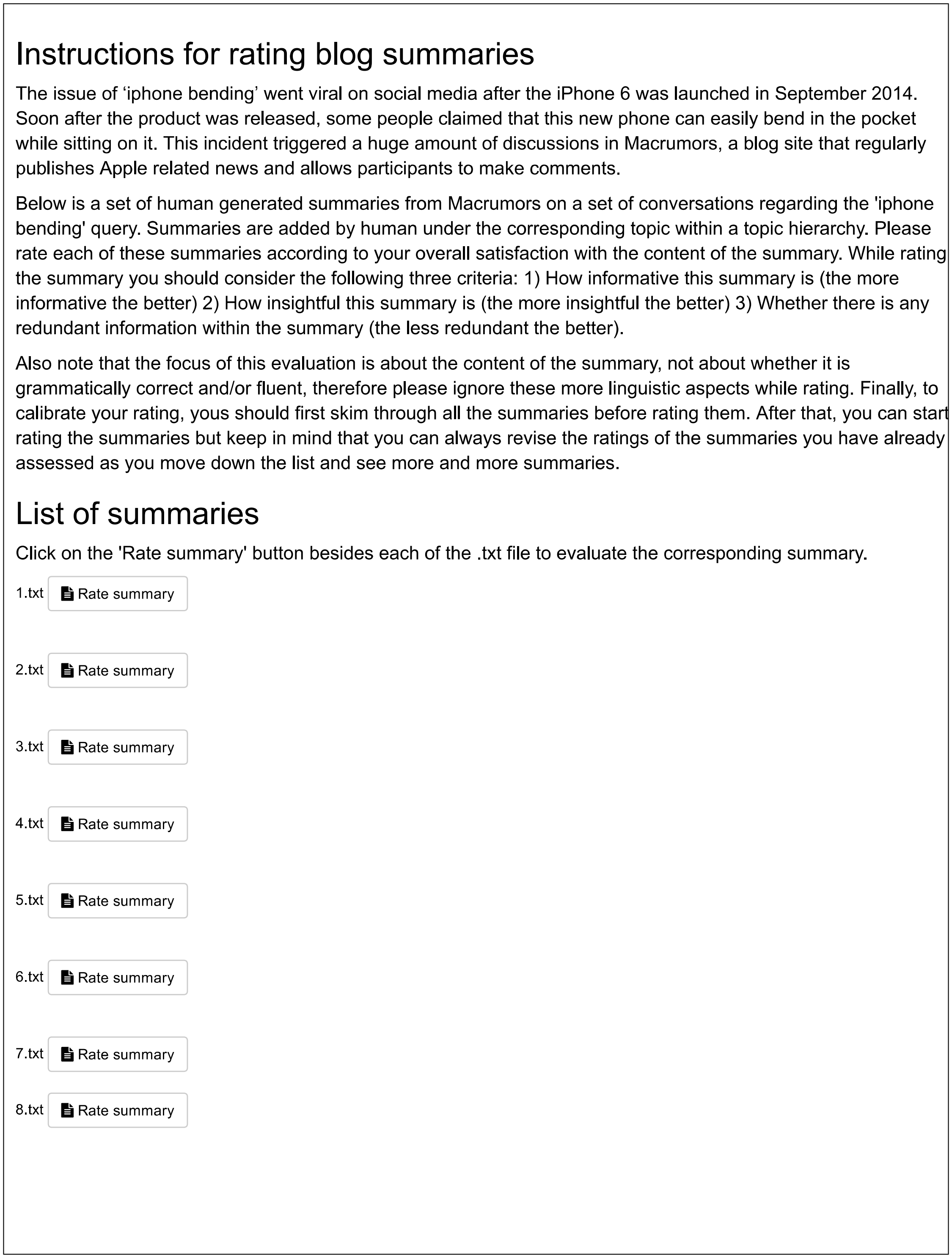
○ A

○ B

Can you please justify your preference?

Additional comments

B.2.3 Evaluating User-Generated Summary by Human Raters



Instructions for rating blog summaries

The issue of ‘iphone bending’ went viral on social media after the iPhone 6 was launched in September 2014. Soon after the product was released, some people claimed that this new phone can easily bend in the pocket while sitting on it. This incident triggered a huge amount of discussions in Macrumors, a blog site that regularly publishes Apple related news and allows participants to make comments.

Below is a set of human generated summaries from Macrumors on a set of conversations regarding the 'iphone bending' query. Summaries are added by human under the corresponding topic within a topic hierarchy. Please rate each of these summaries according to your overall satisfaction with the content of the summary. While rating the summary you should consider the following three criteria: 1) How informative this summary is (the more informative the better) 2) How insightful this summary is (the more insightful the better) 3) Whether there is any redundant information within the summary (the less redundant the better).

Also note that the focus of this evaluation is about the content of the summary, not about whether it is grammatically correct and/or fluent, therefore please ignore these more linguistic aspects while rating. Finally, to calibrate your rating, yous should first skim through all the summaries before rating them. After that, you can start rating the summaries but keep in mind that you can always revise the ratings of the summaries you have already assessed as you move down the list and see more and more summaries.

List of summaries

Click on the 'Rate summary' button besides each of the .txt file to evaluate the corresponding summary.

1.txt Rate summary

2.txt Rate summary

3.txt Rate summary

4.txt Rate summary

5.txt Rate summary

6.txt Rate summary

7.txt Rate summary

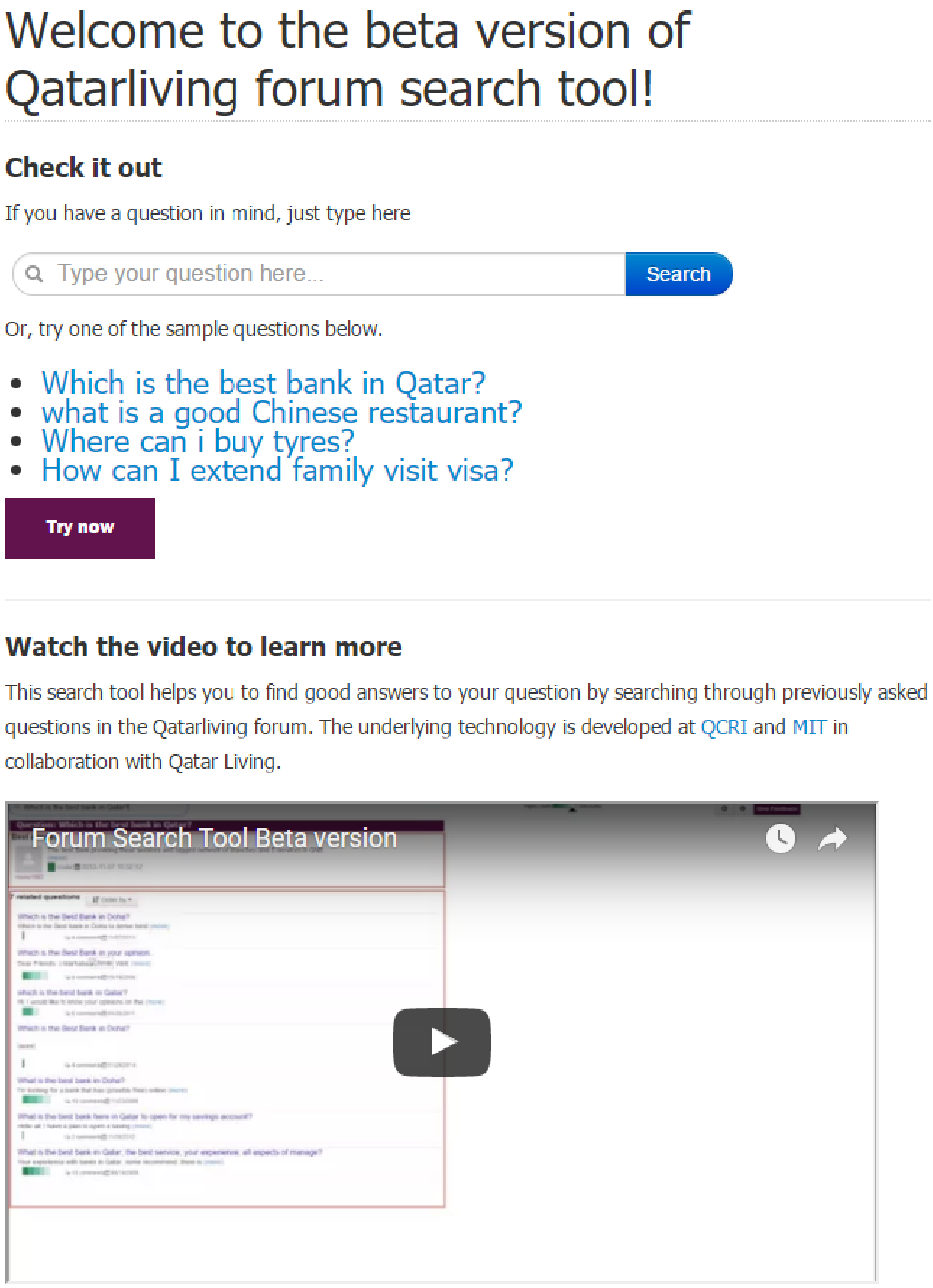
8.txt Rate summary

Appendix C

Supplemental Materials for Chapter 5

This appendix contains supplemental materials for Chapter 5, namely the questionnaires used during the study.

C.1 User Study



Welcome to the beta version of Qatarliving forum search tool!

Check it out

If you have a question in mind, just type here

Type your question here...

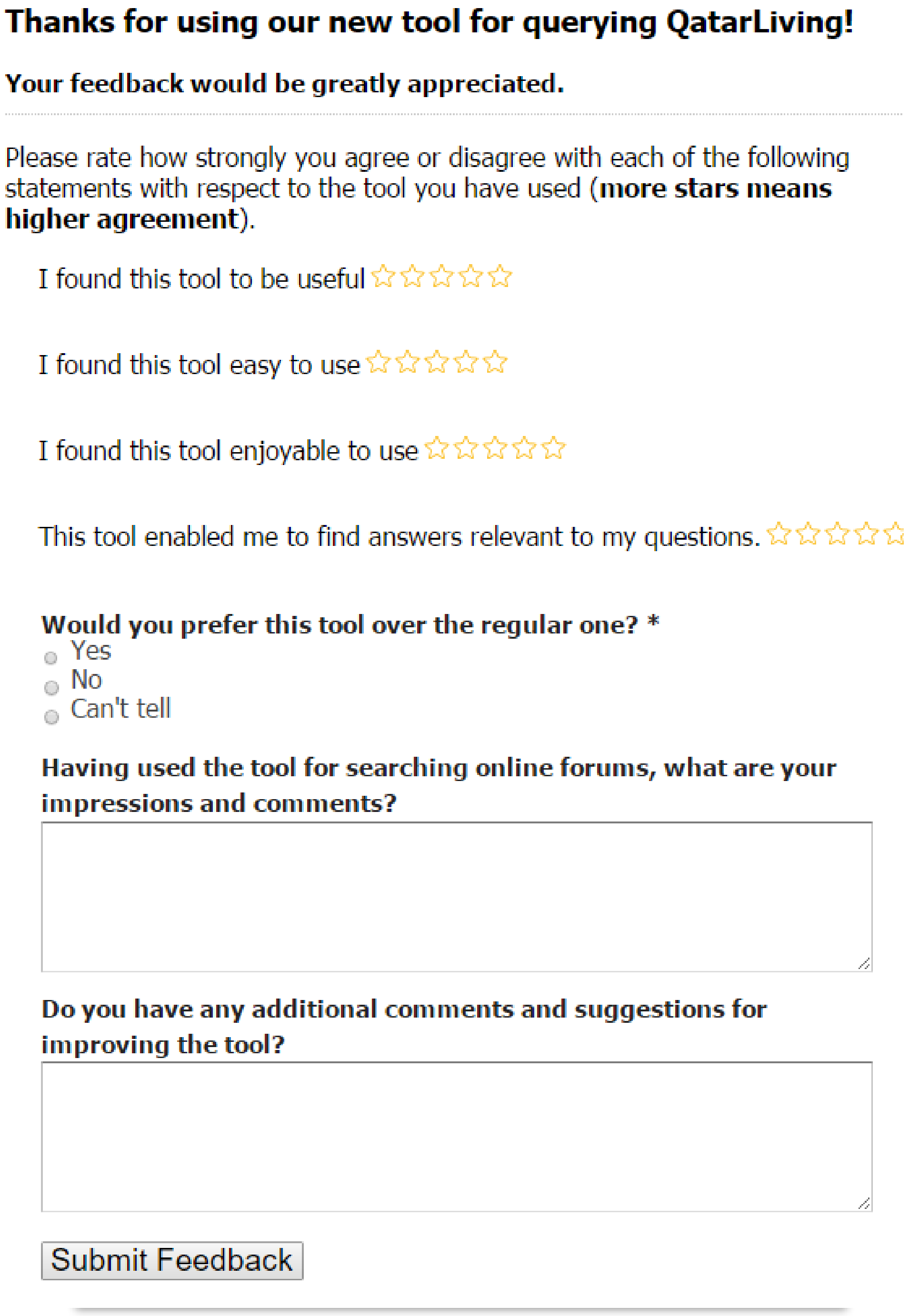
Or, try one of the sample questions below.

- Which is the best bank in Qatar?
- what is a good Chinese restaurant?
- Where can i buy tyres?
- How can I extend family visit visa?

Watch the video to learn more

This search tool helps you to find good answers to your question by searching through previously asked questions in the Qatarliving forum. The underlying technology is developed at QCRI and MIT in collaboration with Qatar Living.

Figure C.1: The introduction page.



Thanks for using our new tool for querying QatarLiving!

Your feedback would be greatly appreciated.

Please rate how strongly you agree or disagree with each of the following statements with respect to the tool you have used (**more stars means higher agreement**).

I found this tool to be useful ☆☆☆☆☆

I found this tool easy to use ☆☆☆☆☆

I found this tool enjoyable to use ☆☆☆☆☆

This tool enabled me to find answers relevant to my questions. ☆☆☆☆☆

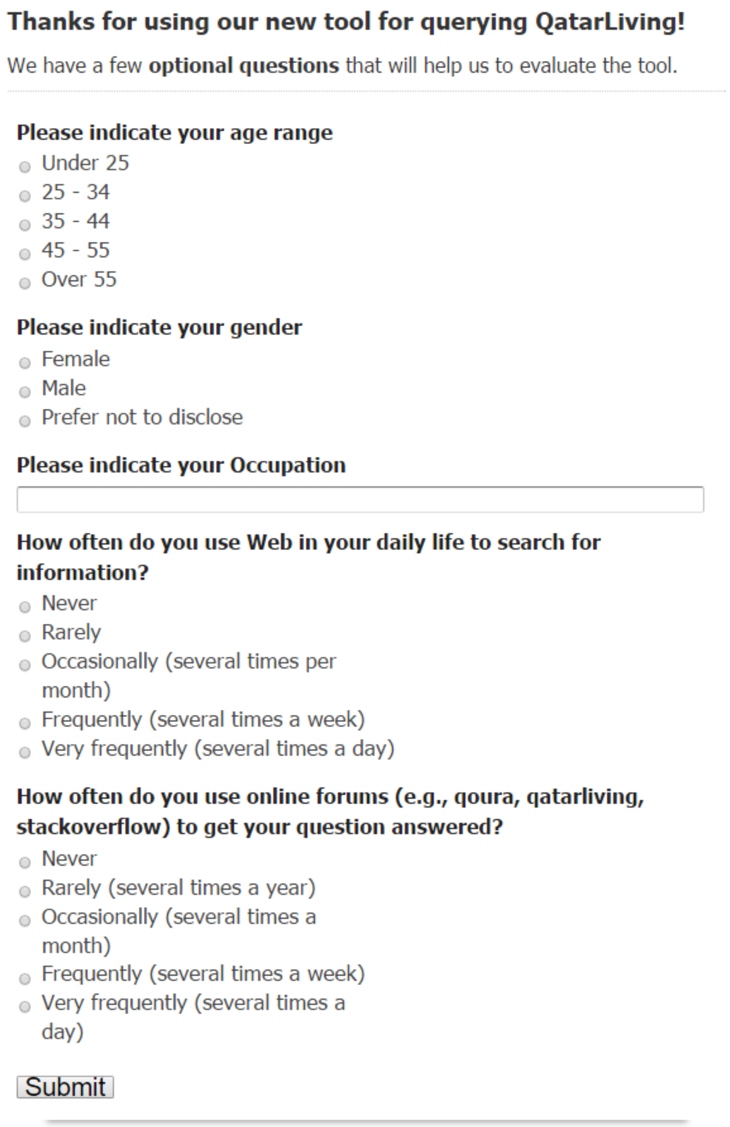
Would you prefer this tool over the regular one? *

○ Yes
○ No
○ Can't tell

Having used the tool for searching online forums, what are your impressions and comments?

Do you have any additional comments and suggestions for improving the tool?

Figure C.2: The post-study questionnaire regarding the user's subjective experience.



Thanks for using our new tool for querying QatarLiving!

We have a few **optional questions** that will help us to evaluate the tool.

Please indicate your age range

○ Under 25
○ 25 - 34
○ 35 - 44
○ 45 - 55
○ Over 55

Please indicate your gender

○ Female
○ Male
○ Prefer not to disclose

Please indicate your Occupation

How often do you use Web in your daily life to search for information?

○ Never
○ Rarely
○ Occasionally (several times per month)
○ Frequently (several times a week)
○ Very frequently (several times a day)

How often do you use online forums (e.g., qoura, qatarliving, stackoverflow) to get your question answered?

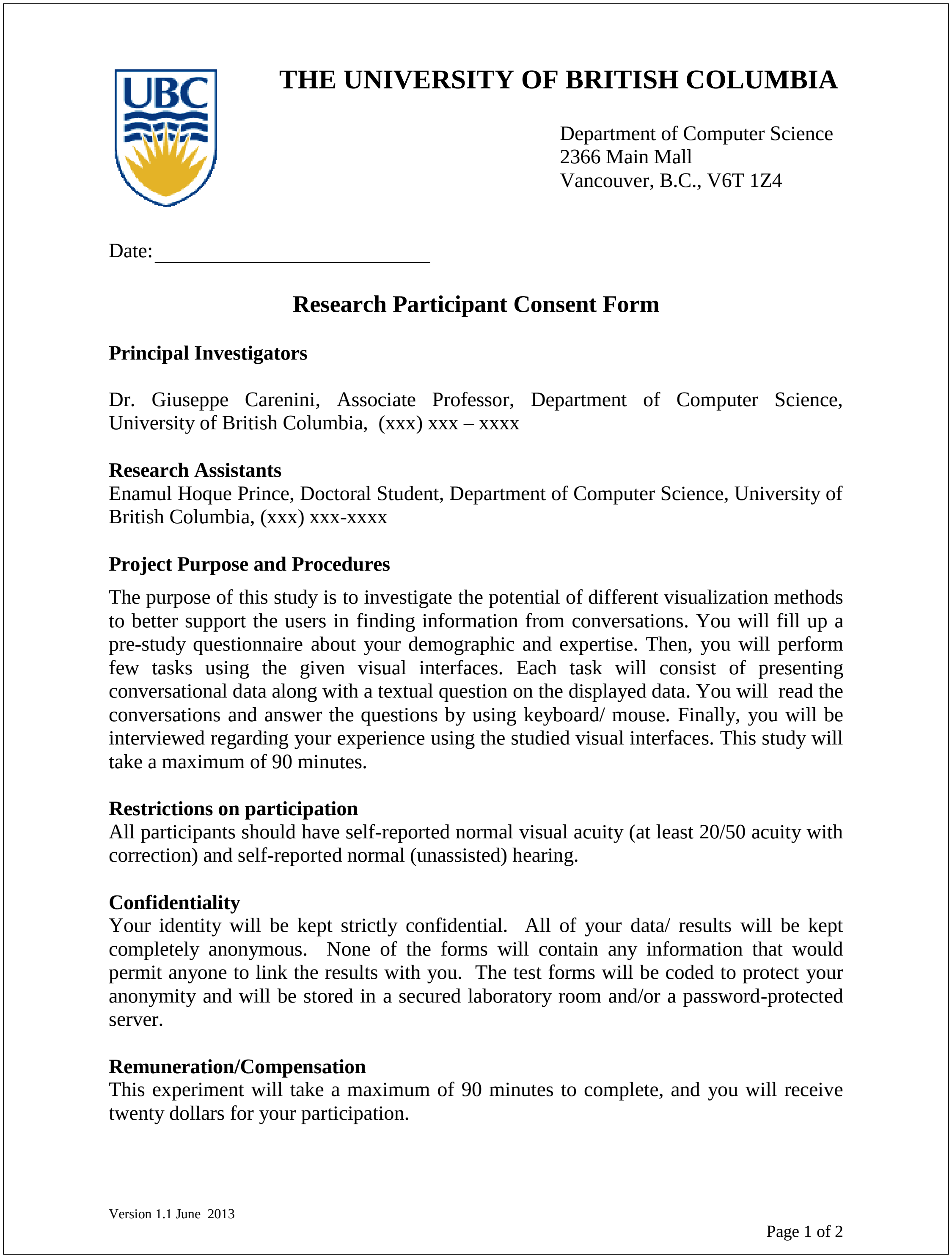
○ Never
○ Rarely (several times a year)
○ Occasionally (several times a month)
○ Frequently (several times a week)
○ Very frequently (several times a day)

Figure C.3: The post-study questionnaire regarding the user's background and prior Web experience.

Appendix D

Participant Consent Forms

The following consent form was used for the user study in Chapter 4. A consent form with identical wording was used in Chapter 3, with the exceptions about the amount of of payment ($15 instead of $20) and the duration of the study (75 minutes instead of 90 minutes).



THE UNIVERSITY OF BRITISH COLUMBIA

Department of Computer Science
2366 Main Mall
Vancouver, B.C., V6T 1Z4

Date:___________________________

Research Participant Consent Form

Principal Investigators

Dr. Giuseppe Carenini, Associate Professor, Department of Computer Science, University of British Columbia, (xxx) xxx – xxxx

Research Assistants
Enamul Hoque Prince, Doctoral Student, Department of Computer Science, University of British Columbia, (xxx) xxx-xxxx

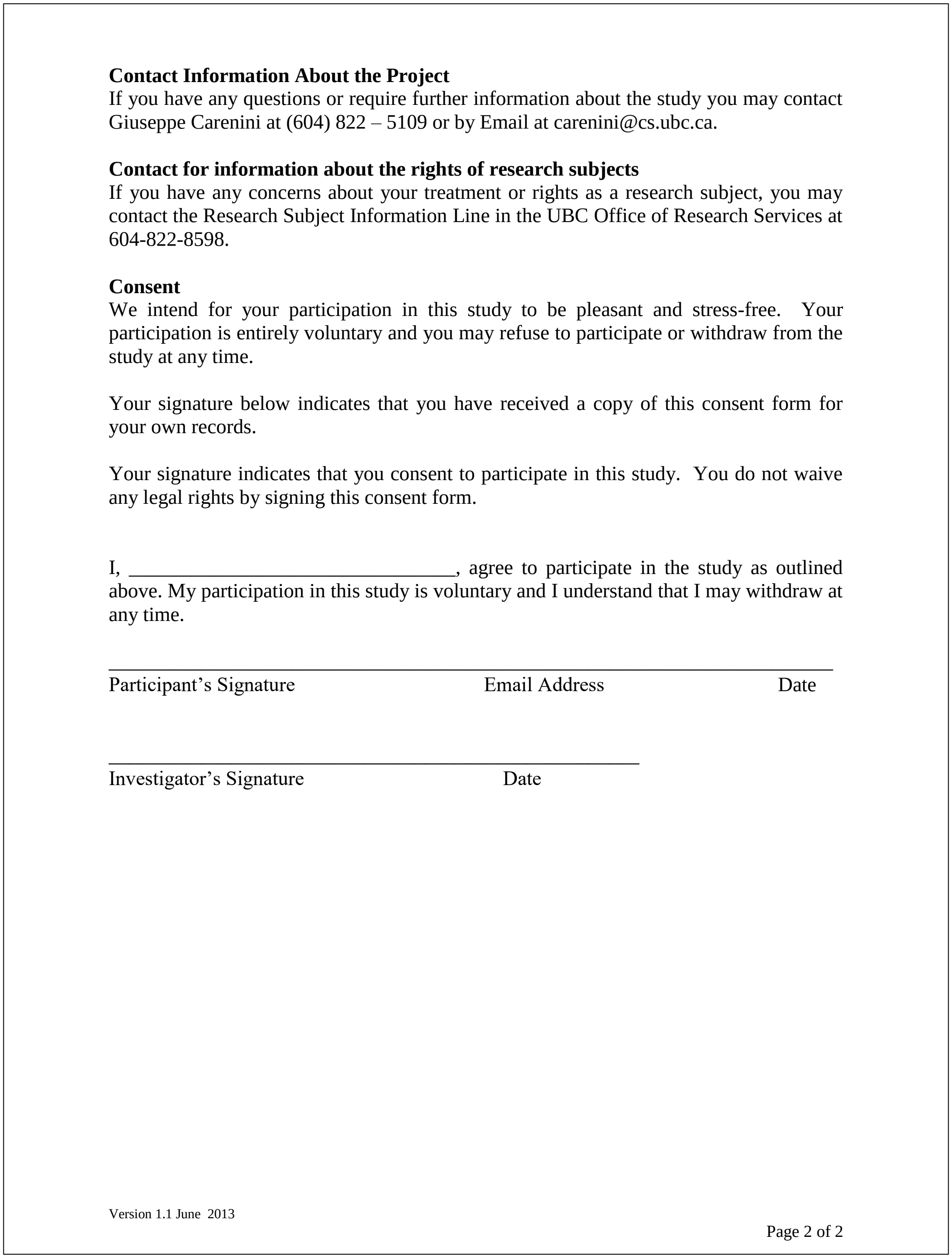
Project Purpose and Procedures

The purpose of this study is to investigate the potential of different visualization methods to better support the users in finding information from conversations. You will fill up a pre-study questionnaire about your demographic and expertise. Then, you will perform few tasks using the given visual interfaces. Each task will consist of presenting conversational data along with a textual question on the displayed data. You will read the conversations and answer the questions by using keyboard/ mouse. Finally, you will be interviewed regarding your experience using the studied visual interfaces. This study will take a maximum of 90 minutes.

Restrictions on participation
All participants should have self-reported normal visual acuity (at least 20/50 acuity with correction) and self-reported normal (unassisted) hearing.

Confidentiality
Your identity will be kept strictly confidential. All of your data/ results will be kept completely anonymous. None of the forms will contain any information that would permit anyone to link the results with you. The test forms will be coded to protect your anonymity and will be stored in a secured laboratory room and/or a password-protected server.

Remuneration/Compensation
This experiment will take a maximum of 90 minutes to complete, and you will receive twenty dollars for your participation.



Contact Information About the Project
If you have any questions or require further information about the study you may contact Giuseppe Carenini at (604) 822 – 5109 or by Email at carenini@cs.ubc.ca.

Contact for information about the rights of research subjects
If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604-822-8598.

Consent
We intend for your participation in this study to be pleasant and stress-free. Your participation is entirely voluntary and you may refuse to participate or withdraw from the study at any time.

Your signature below indicates that you have received a copy of this consent form for your own records.

Your signature indicates that you consent to participate in this study. You do not waive any legal rights by signing this consent form.

I, ________________________________, agree to participate in the study as outlined above. My participation in this study is voluntary and I understand that I may withdraw at any time.

____
Participant's Signature   Email Address   Date

____
Investigator's Signature   Date

Version 1.1 June 2013

Page 2 of 2