A NATURAL LANGUAGE QUESTION ANSWERING SYSTEM FOR EXPLORING ONLINE CONVERSATIONS

NADIA ASHFAQ SIDDIQUI

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF ARTS

GRADUATE PROGRAM IN
INFORMATION SYSTEMS & TECHNOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO
NOVEMBER 2020

Abstract

The proliferation of social media has resulted in the exponential growth of online conversations. Due to the volume and complexity of conversations, it is often extremely difficult to gain insights from such conversations. This dissertation hypothesizes that synergetic integration of natural language processing with information visualization techniques can help users to better fulfill their information needs. More specifically, we developed a question-answering method that allows the user to ask questions about a conversation and then automatically answers the question by highlighting results in a visual interface. The visual interface, named ConVisQA, was developed by extending ConVis which visually summarizes a conversation by providing an overview of topics and sentiment information. We demonstrate the effectiveness of our approach through a user study with blog readers. The dissertation concludes with a user study comparing our interface with a traditional interface for blog reading as well as considerations for future work.

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, Dr. Enamul Hoque Prince, for his enormous support throughout my master's study. As a great supervisor, Dr. Enamul demonstrated to me how research should be done, provided insightful advice, and most importantly constantly challenged me to do better. He has been very supportive throughout the program of study. I would like to sincerely thank my thesis examination committee members, Prof. Stephen Chen,Prof. Gerd Grau and Prof. Sepideh Ebrahimi for their time and valuable feedback to improve this research work. I would also like to thank Md Tahmid Rahman Laskar for his occasional guidance and help in understanding various tools and development platforms.

Preface

This thesis is submitted to the Faculty of Graduate Studies in partial fulfillment of the requirements for a Master of Arts Degree in Information Systems and Technology. The entire work presented here is done by the author **Nadia Siddiqui** under the supervision of **Dr. Enamul Hoque Prince**. Some parts of this thesis has been published as:

- Nadia Siddiqui and Dr.Enamul Hoque Prince, ConVisQA: A Natural Language Interface for Visually Exploring Online Conversations. In *Proceedings of the 24th International Conference Information Visualisation (IV), July 2020.*

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Motivation

The surge of social media sites is generating conversational data at an accelerating pace that continues to grow every day. Often many people from around the globe contribute to a discussion on various online platforms such as Twitter, Facebook, Reddit, etc., which generates a very long communication thread with hundreds or thousands of comments from different people. Recent statistics report from [60] Alexa's Internet traffic rating service discloses that Twitter and Facebook are among the top 50 most visited sites in the world in blogging and micro blogging websites. A recent study from Pew Research Center shows that 22 % of American adult internet users [47], use Twitter at least every day that is generating over 500 million of tweets, daily in year 2019 [45]

This profusion of conversational data reveals a great opportunity for important various findings in exploring and analyzing such a huge volume of data. Unfortunately, given such large conversations, it becomes extremely difficult for users to fulfill their information needs [19]. This problem is commonly known as information overload, where users may feel overwhelmed by the massive amount of potentially relevant information, and as a result, they fail to complete their information seeking tasks [2].

To address this problem, this dissertation takes a visual text analytic approach, combining natural language processing methods for understating and summarizing discussions and information visualization techniques to present an overview of the conversational data to users. We consider that if the user can ask questions about a long conversation and get answers quickly in a visual form that may help the user to fulfill her information needs without needing to scroll through the huge discussion.

Question answering (QA) is a research area in the intersections of information retrieval (IR) and natural language processing (NLP), which is widely used in building such systems that are capable to automatically answer questions asked by users in natural language. Question answering is a difficult task categorized by information requirements articulated as natural language statements or questions [31]. In contrast, traditional information retrieval approaches use the whole documents such as web pages as a response to the information request, in question answering, a specific portion of the information is returned as an answer to the user request. The stakeholders of a natural language question-answering system are looking for a succinct, coherent, and accurate answer which could be a word, sentence, paragraph, image, or a whole text [40].

This dissertation posits that by integrating natural language processing (NLP), information retrieval, and information visualization (InfoVis) techniques in a synergistic way, we can better support users to fulfill their information needs in responding to their queries expressed in natural language. Language processing techniques are capable to extract required information e.g. topics, people's opinions in a discussion on a topic, or participant's stance about a topic. Hence these techniques can extract a piece of information that answers user's queries and combined with visualization techniques, the response of the user's query can be visualized in a user interface. In this thesis, we propose a system to support answering questions in natural language for precise information seeking by combining InfoVis with NLP techniques

1.2 The Problem

Internet users tend to participate and post comments in social media websites about online contents such as news, videos, reviews, blogs, and debates. The very common format to exhibit comments in an online discussion is a paginated list, organized by timestamp or by user's evaluations and ratings. It is generally known that such lists are not capable to scale and can lead to cyberpolarization which aids to support extreme or controversial opinions [11]. When a reader wants to explore such a large conversation, traditional social media sites provide very limited support. These discussions only display the original posts and subsequent replies in hierarchical order and paginated content, usually in descending order. Thus, the reader needs to go through a long conversation with hierarchically organized comments until their information requirements are satisfied [23]. It is very difficult for users to explore and go through all the previous comments in an online discussion to locate the information they need making it almost impossible to gain important insights form the discussion [22]. Glancing through a blog collection is presently not well supported. Users cannot gain an overview of a blog easily, nor do they receive satisfactory support for finding potentially motivating or controversial posts and comments in the blog [50].

To illustrate the problem, let us consider the issue of the current pandemic situation coronavirus Figure 1.1 is a sample of one-word search query "Coronavirus" on a blog website "Reddit " and the result contained more than 50 responses as main blog posts. This blog post had 3.8 thousand comments and there is an indentation at different levels. If a user wants to read this blog conversation, there are chances that the user will not be able to explore this whole conversation and will stop reading through without having his/her information needs being fulfilled. There is very limited support to explore these conversations in terms of question-answering in form of natural language.

The Question Answer (QA) systems are established to retrieve accurate and concise answers to human queries posted in natural language. The stakeholders of a natural language QA system look for a brief and accurate answer to their questions. The response to their question may denote a term, sentence, a subsection, a picture, audio recording, a book, or a complete text document. "The main purpose of a QA system is to find out "WHO did WHAT to WHOM, WHERE, WHEN, HOW and WHY?" [35]. The increased popularity and usage of social media has generated a demand for various services to assist the internet in locating desired information quickly. Question answering (QA) systems are one of these services that have gained the researcher's attention. Various techniques to provide efficient and accurate retrieval of answers to questions have been introduced and evaluated for different scenarios. For example, a user may ask a question such as "Which

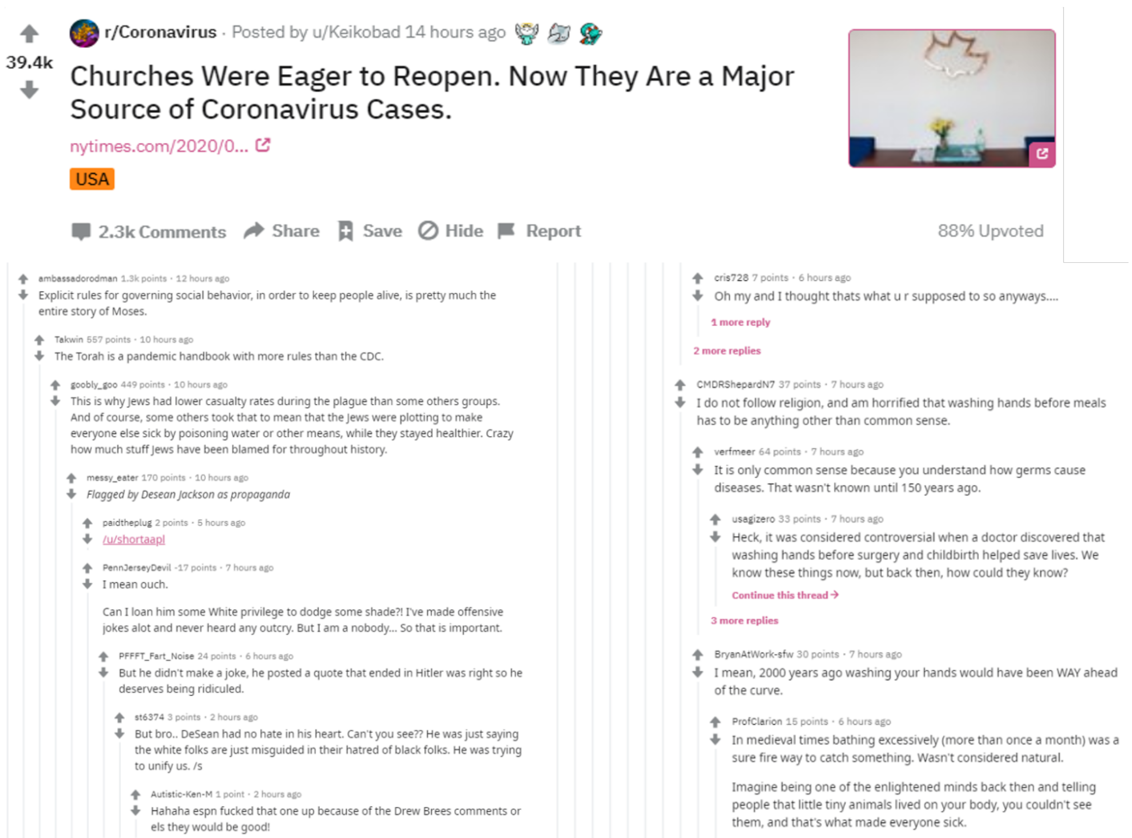


Figure 1.1: A list of conversations returned a discussion with 2.3k comments by Reddit to user's one-word query "Coronavirus" (Accessed June 2020)

country had the highest coronavirus cases?" a search engine developed on keyword-based search such as Google may display a massive list of documents and web pages from the various sources whereas a question answering system will try to answer the question with the name of a country and number of cases [49]. To the best of our knowledge, there is no such system that answers user's questions, or meta-questions about blog conversations, posed in natural language by visualizing the answer in a user interface. Our research aims to address this gap by using topic modeling, sentiment analysis, and information retrieval methods to answer user's questions and then visualizes the answer within a user interface using InfoVis techniques.

1.3 Approach

The primary goal of this thesis is to develop Question answering (QA) with the capability to automatically answer questions about blog conversations asked by human beings in natural language and then visualizing the answer. Our hypothesis is that by combining Natural Language Processing (NLP) and Information Visualization (InfoVis) techniques we can support answering user's questions using various facets such as sentiment, topics, and opinions or authors in different kinds of multi-party conversations (e.g. debate, blog conversations).

But how can we integrate NLP and InfoVis techniques effectively to enable efficient information retrieval for a question answering system? More specifically, we pose the following research questions to evaluate Usability and efficacy of interface:

1. What kind of questions people may ask while they explore a conversation?

2. What metadata and text analysis approaches are useful to answer these questions?

3. How a question answering interface for exploring multiparty conversations may improve user performance and subjective ratings?

4. Can such question answering interface scale up for a large conversational dataset with thousands of comments?

Keeping the above questions in mind, we introduce a natural language interface that supports users to quickly locate and browse through the comments that are relevant to their information needs. Our system takes a question about the conversation from the user as input and then automatically finds the answers using natural language processing techniques. It then presents the results by highlighting in a visual interface, enabling the user to quickly navigate through the comments that match her information needs. Our interactive question answering system named ConVisQA allows the user to express their specific information needs from an online conversation through natural language. In response, the system uses natural language processing (NLP) techniques to automatically parse the questions, locate the answer throughout the conversation and displays highlighted information to

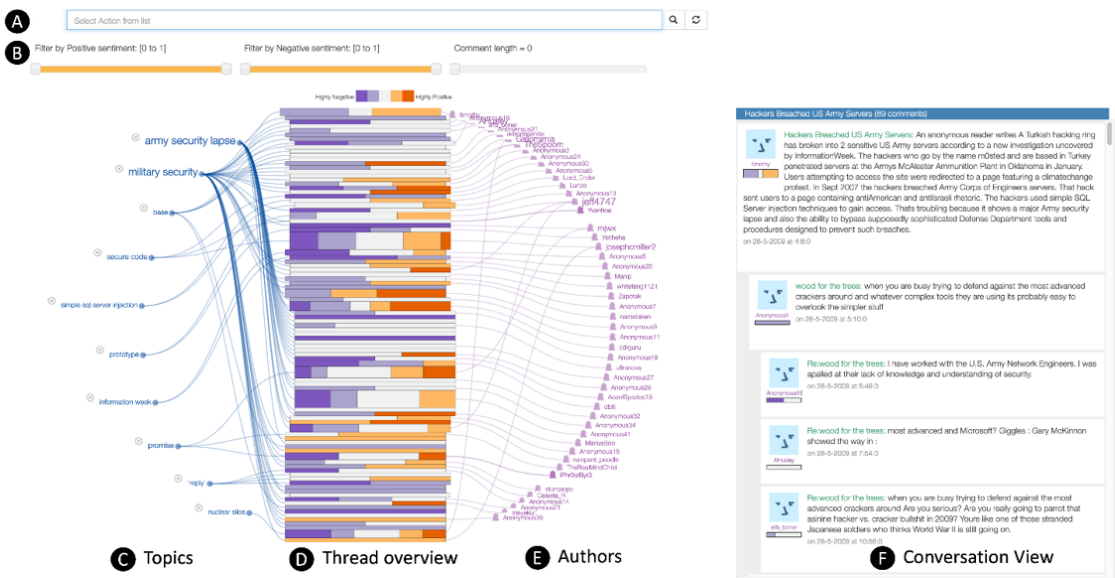


Figure 1.2: ConVisQA user interface allows users to explore a blog conversation for Question Answering

the user. Then the system allows users to further explore a conversation based on certain facets (e.g. sentiment, topics of discussions, etc.) and highlights them to the user. Also, the system can filter based on different facets (e.g. filter comments based on topics, sentiment scores, comment length). This system integrates advanced natural language processing and visual analytical techniques for retrieving specific information from large conversations, enabling a user to quickly locate the relevant information to a given question so that they can fulfill their specific needs.

1.4 Thesis Organization

We will start Chapter 2 with a literature review on existing visualization techniques for online conversations. We will also review recent techniques for natural language interaction techniques for visualizations. Then in Chapter 3, we will describe our proposed approaches including offline processing and online processing for the question-answering task. Next, in Chapter 4 we will explain the design of the ConVisQA interface. In Chapter 5, we will present the user evaluation of our approach. Finally, we will conclude our thesis with an overview of future work in Chapter 6.

2 Literature Review

Our work is in the intersection of question answering, visual analytics and NLP techniques for multi-threaded conversational data. In this Chapter, we will first discuss different types of visualizations for supporting exploration of a collection of conversations. These visual interfaces can be categorized based on the information they extract and visualize: (a) metadata of the conversations, such as timestamps, tags, and authors, (b) text analysis, e.g. topic modeling and opinion extraction. We will also review the literature on natural language interactions with data.

2.1 Visualizations for Exploring Conversations

Early efforts on visualizing online conversations has mostly focused on visualizing the thread construction of a conversation using tree visualization techniques, such as using a mixed model visualization to show both linear sequence and reply associations [51], thumbnail metaphor using a sequence of rectangles [7] [55], and radial tree layout [38]. Opinion Space represents interactive map of a conversation.

Each point represents a user and comment and is based on a metric relationship between users based on approximately similar opinions[11]. A few systems rely on clustering techniques to visualize the conversations as clusters of words grouped by topics [3]. However, these visualizations do not study and visualize the actual text or comments of the conversations.

Others have analyzed the textual content of conversations and primarily focused on topic summarization [41] [49], or visualizing the content progression over time [18]. For example, ConVis visualizes blog conversations using topics, authors, and sentiment. It also provides several interaction features such as highlighting based on multiple facets to support the user in exploring and navigating the conversation. ConToVi [10] visualizes dynamics between different topics and speakers in conversations like political debates using animations with radial visualization. It also presents speaker's behavior using categories like sentiment, courtesy, and expression. Another attempt to further enhance the communication between humans and computer is a Customer Service Automatic Answering System, that extracts Question Answer sets from online documents and stores them in the knowledge base [15]. This system evaluates the customer's question, then learns the meaning of the customers' question accurately and retrieves the knowledge base. In response, it returns a high-quality image as answer to the user. While the above techniques have shown visualizations for exploring conversations promises in providing an overview of the conversations and interactive features for navigating through the comments, the user may find it still difficult and time consuming to locate the comments of interests related to a specific question (e.g. "Which comments are saying negative about pandemic"?) using such visual interfaces.

2.2 Visualizing Topic Models

Understanding and navigating large collections of documents has become an important activity in many spheres. However, many document collections are not coherently organized and organizing them by hand is impractical. There is need to automate ways to discover and visualize the structure of a collection to facilitate the exploration of its contents conveniently.

To support precise and meaningful search in large corpora or documents, topic modeling have been found very effective in determining the core discussion topics and arguments that pervade a large and unstructured blog conversation [9]. A fully web-based coordinated-view system was introduced to view Topic Streams from twitter data and the topics were visualized as a stream in a temporally adjustable stacked graph visualizing how the topics evolve over time [52]. Themail visualizes how topics in an archived email conversation evolve over time. The key is to arrange the keywords selected based on term-frequency inverse document-frequency (TF-IDF) along a horizontal time axis [52]. TextFlow [5] is another visual analysis tool that helps users analyze how and why the associated topics evolve gradually. Users can discover the progression patterns at different levels of details of individual topics over time which is represented as splitting and integrating relationship using incremental Hierarchical Dirichlet process. TIARA [56] characterizes the chronological evolution of topics from an email assortment by applying the ThemeRiver visualization [16], where every layer in the stacked graph represents a topic and the keywords of each topic are dispersed gradually. From the height of each topic and its content distributed over time, the user can observe the topic progression. Topicpanorama visualizes topics by performing a meaningful alliance of their keywords [53].

2.3 Opinion Visualization

There is a rising attention in analysing the opinions or feelings expressed in conversations, generally processing microblogs website seeking this expression [8] [48]. TwitInfo [34] is designed to support visualizing enhanced, accurate and acquisitive sentiment information of micro-blogging websites in form of huge collection of tweets. On the other hand, OpinionFlow is more concentrated on visualizing the dispersion of opinions about a particular theme or topic (e.g., 'pandemic or any other current affair') among the micro bloggers with a combination of a density map and a Sankey diagram [57] Sometimes the information about opinion are combined with other important facets of information dispersal such as chronological information, and the connections among conversation threads and authors [59].

2.4 Sentiment Analysis

Sentiment analysis approaches can be generally categorized into two types: learning-based and lexical-based [14] [58]. Learning-based method uses recognized properties derived from labelled training data to make predictions about unlabeled new data. In text data, it derives the relationship between the features of the text segment.

There are various studies demonstrating that lexicon-based sentiment analysis is more accurate than learning based approaches [26]. On the other hand, lexical-based methods typically search a document or text for sentiment or feeling pointers specified in the existing lexicons used [14][12][32]. The effects of the pointers are then accumulated in order to get the foremost polarity of the text. Compared to learning-based methods, lexical-based methods are easier to be applied across different data sets and the desired results can be achieved without training of dataset[54].

2.5 Natural language Interactions with Data

Natural language interfaces for data visualization have received considerable attention recently [13]. Typically, these interfaces respond to user queries by either creating a new visualization (DataTone [13]) and/or by highlighting answers within an existing visualization (Eviza [43]). Some systems enable follow-up data queries from users with limited support for pragmatics (e.g.Evizeon [21], Orko [44]). Some of them provide query auto-completion features either by supporting syntactic query formulation[43] or by supporting information recall and data preview [42].

Generally, these systems recognized the importance of providing feedback on how the system interprets queries and enabling users to correct misunderstandings through interface widgets. However, most of these works largely depend on heuristics for parsing which are incapable of handling questions that are compositional or otherwise incomplete and ambiguous.

There has been a recent surge in research on conversational interfaces [39], one of the avenues of such research is automatic question answering with data. For instance, some works focus on answering questions with semi-structured table using semantic parsing techniques [27], [4]. However, the above works have mainly focused on interacting with tabular data whereas the conversational data used in our system is textual.

3 Proposed Methodology

In this chapter, We first present our user requirements analysis which informed our system design. We then provide an overview of our ConVisQA system which supports users to ask questions about conversations and get answers in visual forms.

3.1 User Requirements Analysis

In order to guide our system design, we performed an initial formative study with 3 users (2 females and one male, age range 21-35 yrs) who regularly read blogs. During the study, the participants were asked to read some given blog conversations according to their own needs and interest and then write a set of questions that come to their mind. We also requested them to suggest any improvements in the system design. In addition, we rely on previous literature review of why and how people read blogs [18] to get a sense of what kind of information needs they have in mind.

Through the formative study and the literature review to compile a list of questions that people may ask naturally, given a blog conversation they are reading. We used this list to inform the design of our prototype, including the most common types of questions they ask as well as what kind of keywords they use and the grammatical structures their questions usually follow. Table 3.1 shows a set of example questions that people typically ask while exploring conversations. We also identify what kind of data variables are involved in each of the questions and the analytical functions that are necessary to answer these questions.

During the initial formative study, P1 suggested various questions about authors who post comments in the blog conversation e.g which authors always post controversial comments? P2 suggested having sentiment filters for positive and negative comments and it will be better to cluster comments based on similar content or a key phrase. P3 reported, “that there should be a guide or tool-tip on each facet such as topic or author to guide the users what to do. P3 found it difficult to understand without demonstration of system functionality”.

3.2 System Overview

In this research, we build our system named ConVisQA on top of ConVis which is a visual text analytic system for online conversations. Generally, the architecture of a Question Answering System is composed of three major components, such as syntactic question processing, locating relevant passage or document and then pro-

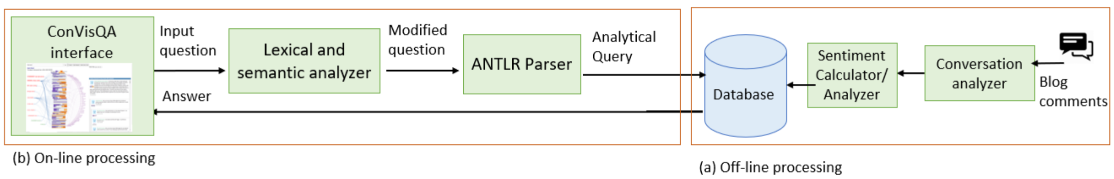


Figure 3.1: ConVisQA System Overview system for supporting question answering

cessing the answer [17] [25] (i) question/ query processing; (ii) Document retrieval and (iii) answer extraction. ConVisQA integrates these components in two phases as shown in Figure 3.1: (1) offline processing (Figure 3.1a) and (2) online processing (Figure 3.1b).

3.3 Offline Processing

In the offline processing (Figure 1a) we pre-process the set of conversations collected from different blog sites (e.g. DailyKos, Business Insider or Slashdot). by cleaning the data to retain only the conversational data in the crawled pages, followed by extracting the conversational structure, i.e., reply-relationships and quotation. We use a state-of-the-art tagger [33] to tokenize text and annotate the tokens with their part-of-speech tags [6]. After that, the conversation analyzer module performs topic modeling and sentiment analysis over the whole set of conversations and store the results in a database. We used a topic modeling approach that first segregates all sentences into groups as a set of topical clusters/segments by utilizing the conversational structure [24]. Then, representative key phrases are assigned to each of these segments (labeling).

For sentiment analysis, we applied the Semantic Orientation CALculator (SO-CAL) [46], which is a lexicon-based approach for determining whether a text expresses a positive vs. negative opinion. SO-CAL computes polarity as numeric values. 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. Finally, the results are stored in a database for efficient retrieval within the interface.

3.4 Online Processing

When the user types a question (e.g., "Which one is the most controversial topic of discussion?"), the system performs the following three steps on the fly:

1. Lexical and semantical analyzer pre-process the questions and re-write the original input question so that the parser can recognize the question.

2. The ANTLR parser takes the modified question and uses a set of pre-defined grammar rules to automatically recognize the entities and analytical functions involved in the to build a query for retrieving results from the database.

3. The ConVisQA Interface obtains the results from the database and presents the results to the user by highlighting the results in the visualization as well as by presenting in text. Below is a brief description of each step to be performed in the implementation of ConVisQA.

3.4.1 The Lexical and Semantical Analyzer

This module first tokenizes the question, remove the stop words, and apply part of speech tagging using Stanford CoreNLP (see Figure 3.2 [33]. It then finds the words that are mentioned in the input questions but do not exist in our pre-defined lexicon. It then applies the word2vec model to check if the user has used a term that is not in our lexicon, but it is most likely that the term is already present in our lexicon. In such a case, we replace the user's term with the term in our lexicon so that the parser can recognize it.

3.4.2 The ANTLR Parser

We use an ANTLR parser that employs a top-down parsing strategy named LL (*) [37]. We choose ANTLR parser because it allows for better flexibility in specifying the grammar rules and has been successfully applied in natural language interface for data visualizations recently [21]. This parser reads the input from left to right, performing the leftmost derivation of the input query. This parser takes a context-

```
text = word_tokenize("who had posted the most negative comments about topic X?")
nltk.pos_tag(text)
```

```
[('who', 'WP'),
 ('had', 'VBD'),
 ('posted', 'VBN'),
 ('the', 'DT'),
 ('most', 'RBS'),
 ('negative', 'JJ'),
 ('comments', 'NNS'),
 ('about', 'IN'),
 ('topic', 'NN'),
 ('X', 'NN'),
 ('?', '.')]
```

Figure 3.2: Part of Speech Tagging on User Input

free grammar with a set of production rules. We design the hand-crafted grammar rules based on prior analysis of how people may ask different analytical questions while they explore conversations.

In particular, we first identified and defined a set of analytical functions (e.g., sort, filter, find extrema, compare) as well as a set of data variables that are involved in a question (e.g. topic, author, comment, sentiment, posting time). Table 1 shows some example questions along with the corresponding analytical functions and data variables that are involved with them.

In the next step, we use the data variables and analytical functions to build grammar rules that the ANTLR parser can process. For instance, consider the following production rule:

$$G^0 \rightarrow .* < extreme >< sentiment > Comment\ about < topic >? \tag{3.1}$$

Sr.	***Question***	***Data Variables***	***Analytical Functions***
1	What is the most/least controversial topic?	Topics, comments	Find extrema
2.	What is the most/ least controversial comment?	Author, Comment	Filter, Find extrema
3.	What is the most/least controversial comment on topic X?	Topic, Comment	Filter, Find extrema
4.	Who is the most/least controversial author?	Author, Comment	Filter, Find extrema
5.	What is the most/least controversial comment?	Comment Filter	Find extrema
6.	What is the most/least controversial comment of an author?	Comment Author	Filter, Find extrema
7.	Who had posted the most negative/positive comments about topic X?	Author, Topics, Comments	Filter, Find extrema
8.	Who was the most dominant participant of the conversation?	Author, Comments	Find extrema
9.	Can you get rid of topic X?	Topic, comments	Filter
10.	Which topics are generating more discussions?	Topic, comments	Sort
11.	Which comments are supporting the claim made in the post?	author, comments	NA
12.	Which authors always posts controversial comments?	author, comments	NA

Table 3.1: A set of example questions along with data variable and analytical functions that are involved in the questions.

In this rule, $< extreme >$ is a non-terminal symbol that can represent analytical tokens like most, least, etc. $< extreme >$ can take values like 'negative', 'positive', and neutral. Finally, $< extreme >$ can take any topic name in the conversation. Given the question "Which one is the most negative comment about topic X", the parser generates a query that finds all the comments about the topic X from the database and then selects the comment that has the most negative score.

4 ConVisQA Interface

In this chapter, we first briefly explain the interface ConVis and its features,then we describe the common visual encodings shared by both ConVis and ConVisQA. Then we demonstrate how the extended interface ConVisQA helps the user in exploring conversations by using natural language question answering and additional features for improving scalability. This chapter explains the differences and similarities in both interfaces.

4.1 Design Rationale

ConVisQA is developed on top of ConVis [18], a visual interface for exploring online conversations which was originally designed based on requirements of the blog-reading tasks. We now explain the major components of ConVis Interface as shown in Figure 4.1, followed by our rationale for extensions made in our ConVisQA interface.

The ConVis interface is primarily designed as an overview + details approach, consisting of following components.

- Thread Overview: the Thread Overview hierarchically represents a visual summary of the whole conversation, and allows the user to navigate through the comments 4.1, middle).

- Sentiment Overview: The horizontal coloured stacked bar in the middle displays each comment. 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 a comment.The stacked bars are vertically ordered according to their positions in the thread starting from top with indentation indicating thread depth, allowing the user to see the whole thread structure at a glance.

- Facet Overview: Topics and authors are presented in a circular layout around the Thread Overview (see 4.1). 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

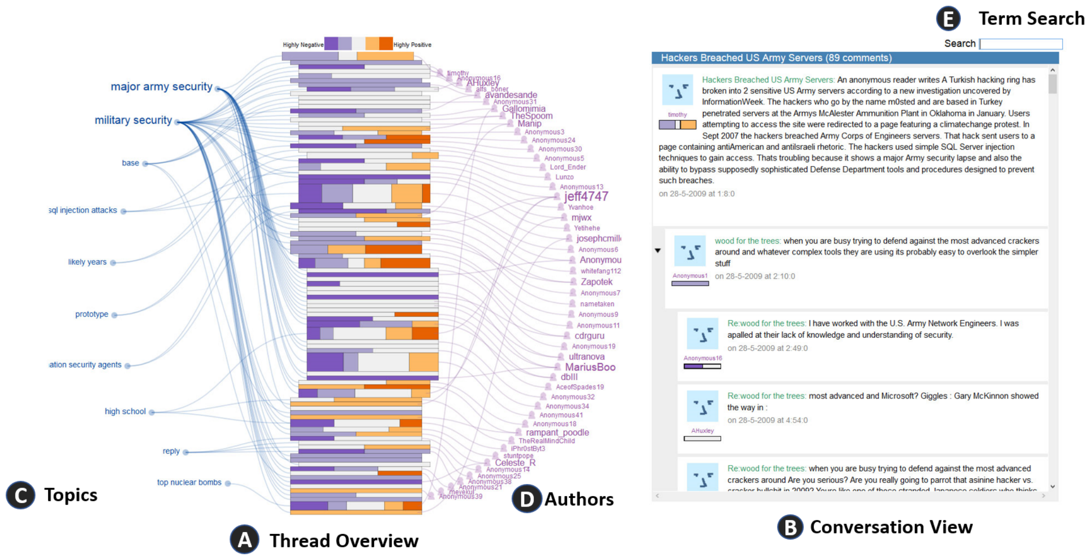


Figure 4.1: A snapshot of ConVis for exploring blog conversation

conversation. Likewise, the font size of an author encodes how many times a participant has posted in a conversation.

- The Conversation View: The Conversation View displays the actual text of the comments as a scrollable list (see 4.1, right). The interface also provides interaction feature such as highlighting user-specified search terms to locate and highlight the word or phrase posted or used in the conversation.

After analyzing the ConVis interface and performing informal studies with participants we identified three main ways to extending it. More specifically, we introduced following new features in ConVisQA:

- Questions Answering: We have introduced the the Question Answering feature to help users fulfill their information needs. ConVisQA is capable to answer the questions asked in Natural language in search bar. The user can type a question and the interface then returns the answer by highlighting corresponding facets (e.g comments, authors and topics).

- Multi-faceted filters: We have introduced multi-faceted filters to enhance readability for larger conversations. For example, users can filter the conversation using positive or negative sentiment scores. It also allows users to filter comments based on number of sentences.

- Scalability: ConVis could only display about 100 comments originally. We have improved scalability by showing a larger dataset that have more than 500 comments. To achieve this, ConVisQA divides the thread overview into pages and a paginated list is displayed to switch to next pages of comments.

4.2 Visual Encodings and Interactive Features

We developed ConVisQA by extending ConVis [18], a visual interface for exploring an online conversation which was originally designed based on requirements of the blog-reading tasks. The ConVisQA interface is primarily designed as an overview+details interface as shown in Figure 4.2. The overview consists of the whole conversation thread as well as the discussion topics and authors who participated in the conversation. The Thread Overview visually represents each comment of the discussion as a stacked bar, where each stacked bar encodes three different metadata (comment length, the position of the comment in the thread as well as the depth of the comment within the thread).

A set of five diverging colors is used to visualize the distribution of sentiment orientation of a comment, ranging from purple (negative polarity) to orange (positive polarity). Thus, the distribution of colors in the Thread Overview can help the user to perceive the kind of conversation they are going to deal with. For example, if the Thread Overview is mostly in strong purple color, then the conversation has many negative comments. The primary facets of the conversations, namely topics and authors, are presented cyclically around the Thread Overview. By default, both topics and authors are positioned according to their chronological order in the conversation starting from the top, allowing the user to understand how the conversation evolves as the discussion progresses. However, such ordering may change based on certain questions (e.g. if the user asks to sort topics based on how controversial there are). To indicate the topic-comment-author relationship, the facet elements are connected to their corresponding comments in the Thread Overview via subtly curved links. Finally, the Conversation View displays the actual text of the comments in the discussion as a scrollable list.

We also made some extensions to the ConVis interfaces by including additional

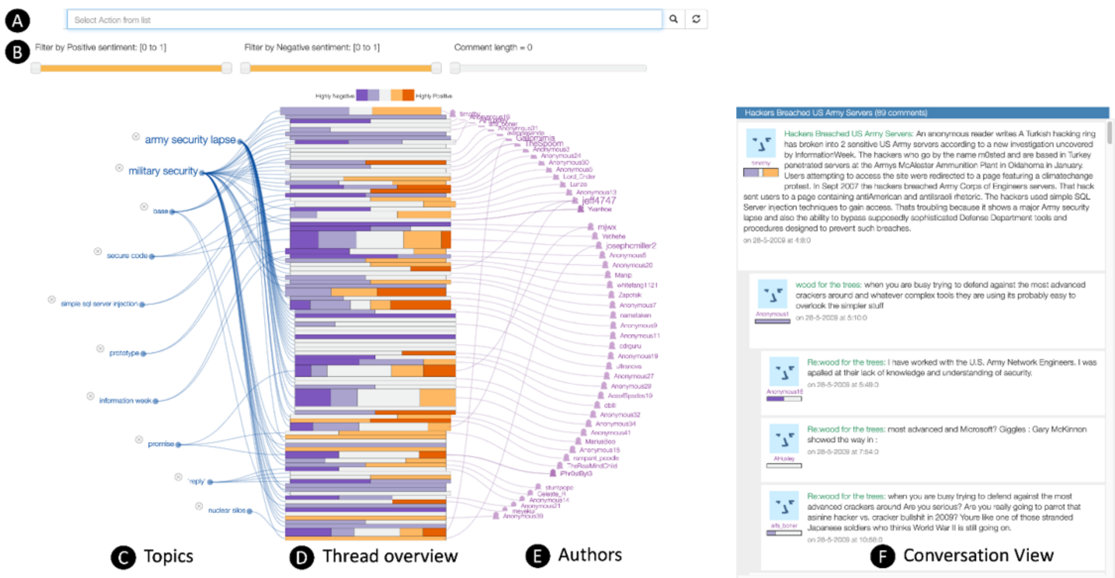


Figure 4.2: The ConVisQA user interface allows users to explore a blog conversation and get their questions answered about that conversation. Here, the user can ask a question using a textbox (A) and can filter comments based on sentiment and comment length (B). The Thread Overview visually represents all the comments using stacked bars (D) while topics and authors are arranged circularly around this overview (C, E). Finally, the Conversation View at the right presents the comments in a scrollable pane (F).

interactive features for filtering, sorting, and deletion. For example, the user can filter comments based on sentiment score Figure 4.3 this can be done for both positive and negative sentiment for whole conversation. The comment length filter helps the reader to filter the conversation for number of sentences in all comments. This feature enables users to filter the conversation for the desired number of comment length implying sometimes a reader is not interested in reading all lengthy comments. The user can also sort comments by comment length. They can delete topic(s) and participating comments as well whenever they find some of them as less interesting or irrelevant. The system will allow the users to delete the topic as soon as they hover the mouse on any topic.

4.3 Interactive Question Answering

ConVisQA allows the user to ask questions by typing in the query box at the top of the interface. As the user starts typing the system parses the current query input and provides the possible query suggestions that may follow the current input. Internally, the system uses the ANTLR parser to check what are the possible valid parse trees given the current input and the grammar that we have designed. The user can select the only question from the drop-down list at any point. The system is flexible to answer the short form of questions such as the system is capable to answer "What is the most negative topic" and the short variation "most negative

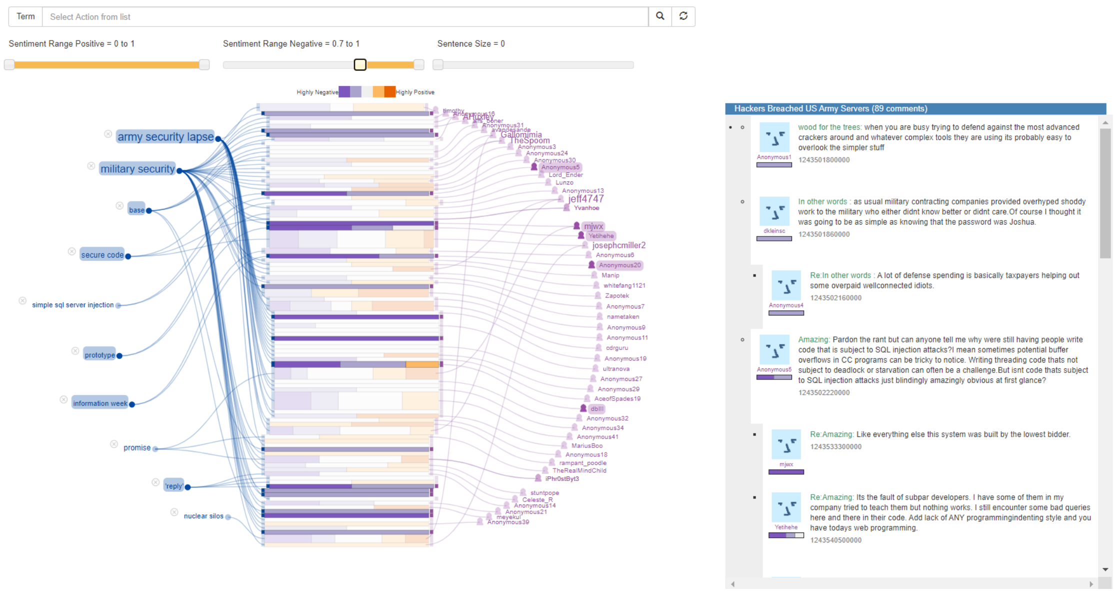


Figure 4.3: ConVisQA supports comments filtering by sentiment(Positive or Negative) and comment length (Number of Sentences in a comment)

question".

Once the user provides a query string and clicks on the search button, the system analyzes the query and then automatically presents the answer by highlighting or filtering the relevant facets e.g. topics, comments, or authors in the overview as well as in conversational view. For instance, given the question "What topic is generating more discussions?", the system finds that the "People" topic has the most number of comments therefore it highlights this topic as well as relevant comments and authors. This response to the user's question is also mirrored in the Conversation View which automatically scrolls to the corresponding comment

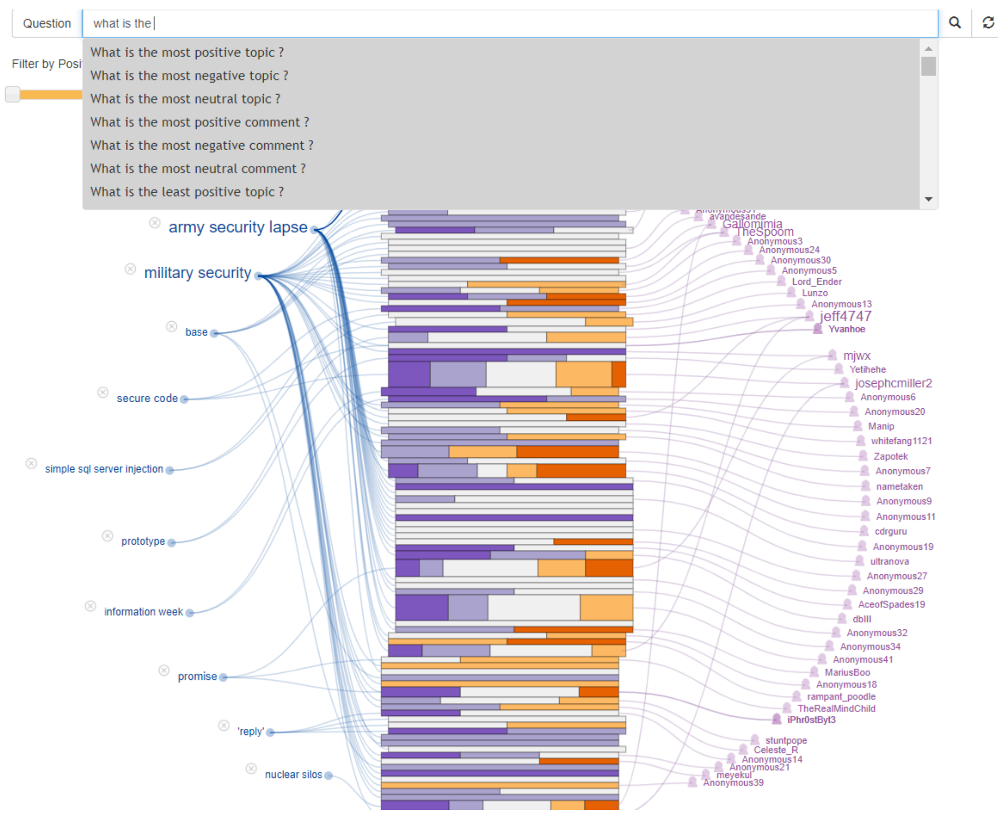


Figure 4.4: Auto-completion (Interface suggests list of questions based on the current input.)

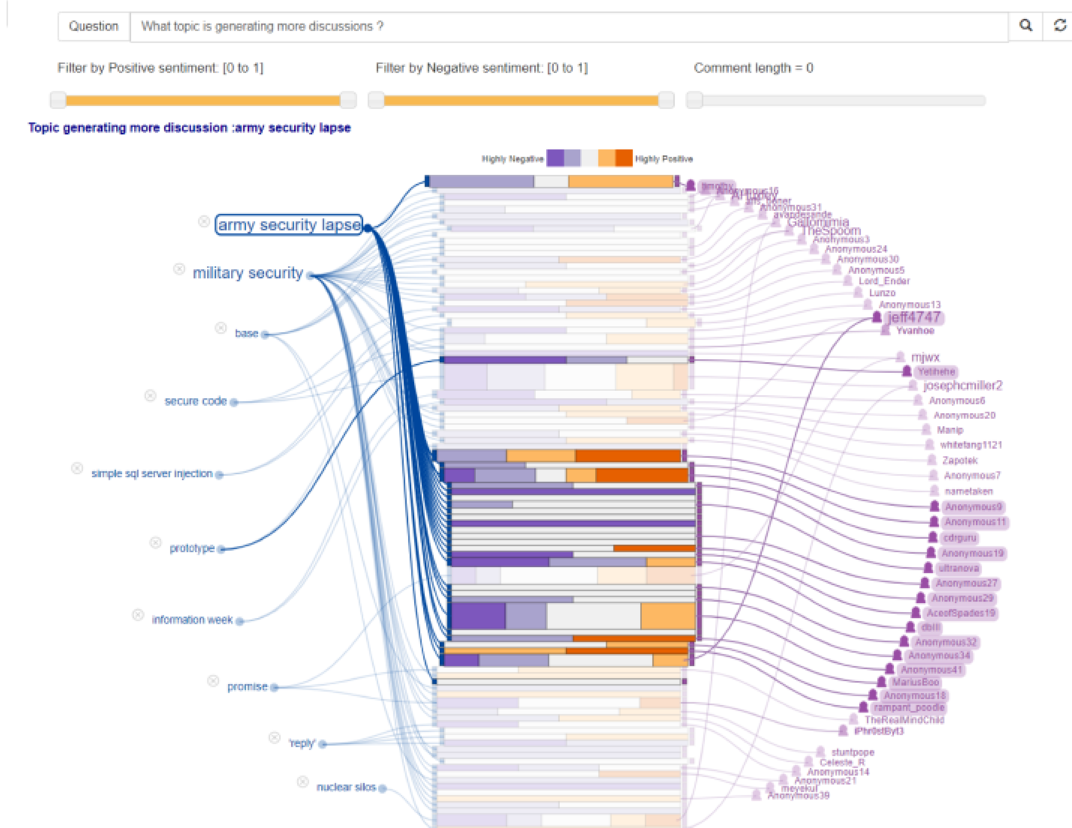


Figure 4.5: The interface is showing the topic generating the most discussion.

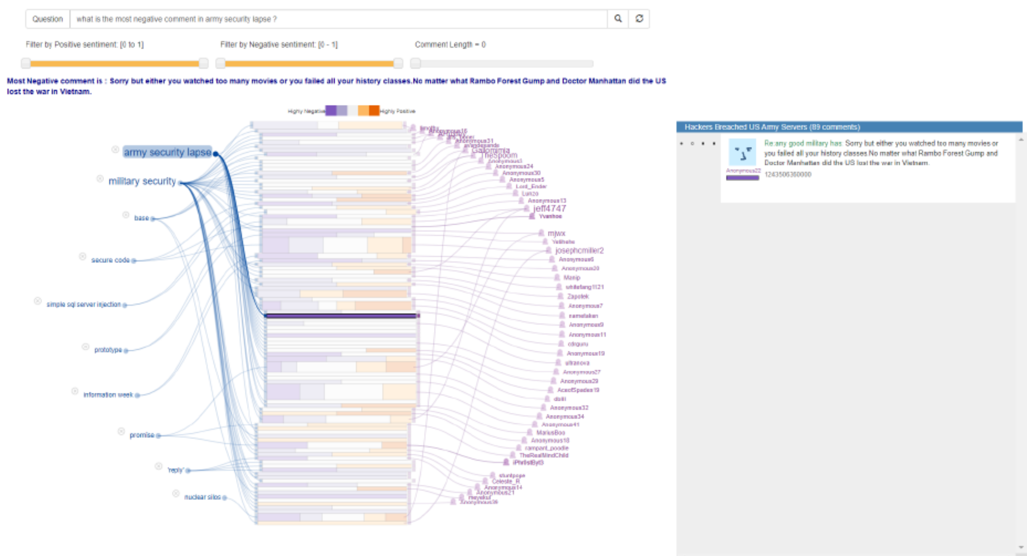


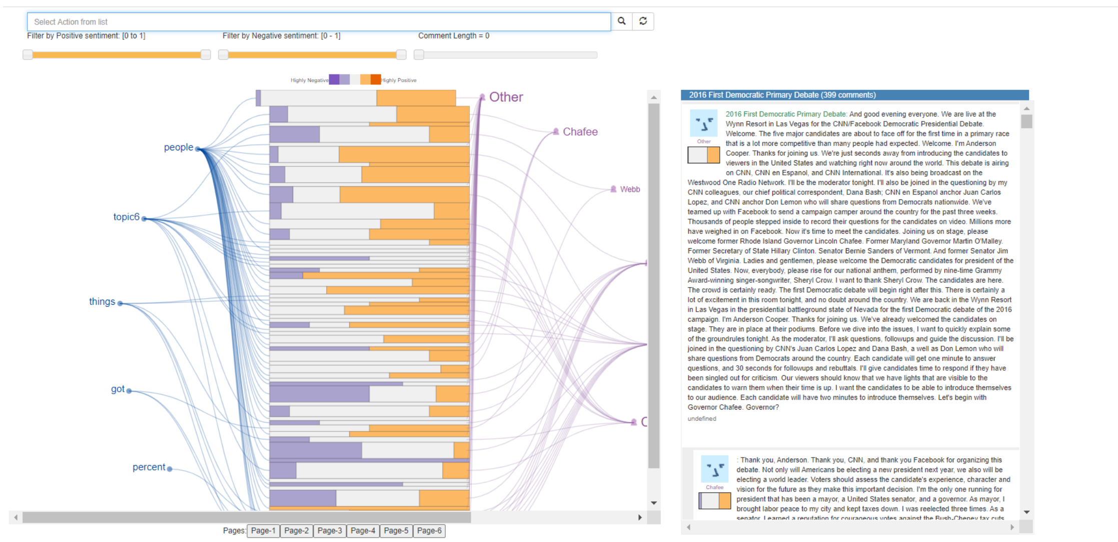
Figure 4.6: The interface is highlighting the answer for the question: what is the most negative comment in 'army security lapse'?

for immediate access. In this way, the user can locate the comments that match her information needs without having to navigate through all the comments.

Figures 4.3, 4.4, and 4.5 illustrate a series of interactions with ConVisQA. In Figure 4.3, the system helps the user to complete a query through the auto-completion feature. The user continues to ask several questions and in response, the system provides the answer in the text as well as highlights the relevant comments in the Thread Overview (Figure 4.4 and 4.5).

In addition to question answering, ConVisQA also supports users to search through various facets such as comments, author or comment+author and system locates and highlights the comments that matched the search criteria. For example, if the user searches for the words ‘security’, the system highlights all the comments that match this word in the Thread Overview as well as in the Conversation View.

ConVis was designed to display a conversation with one hundred comments only. We have extended ConVisQA to support larger conversational data containing over 500 comments. We have displayed a larger dataset in the paginated view. Both thread and conversation views are updated concurrently when a user moves to a new page within the visualization.



4.4 Technical Implementation

A server-side component (in PHP) retrieves conversations that are annotated with topics, comments, and sentiment scores. The visualization component is implemented in JavaScript (using the D3, JQuery, and bootstrap libraries). An ANTLR library 2 footnotes https://github.com/antlr/antlr-php-runtime is used in a PHP environment to automatically parse the question.

5 Online User Study

In this chapter, we present the online user study we performed to evaluate the new question-answering system ConVisQA. In order to understand the usefulness of the questing-answering feature, we compare ConVisQA with ConVis which is a visual text analytic system that serves as a baseline.

ConVis is a system that integrates interactive visualization with novel text mining and summarizing techniques to fulfill the information needs of users in exploring an online conversation through posing questions asked in natural language. Both interfaces are similar in design but the main difference is that ConVisQA has additional question-answering capability and support filtering comments based on different criteria To better understand the potential usefulness of natural language interaction system for conversation and proposed system in real-world scenarios, we undertook a mid-scale, online user study. The primary aim of our study was to examine how users would use ConVisQA and what would be their reactions to such a visual search interface for Question Answering.

5.1 Goals

We ran a formal user study to evaluate the effectiveness and usability of the ConVisQA interface compared to a similar interface that represents a conversation in a similar design. The aim of the user study was to answer the following two questions:
(1) When we compare ConVisQA with the ConVis for retrieving answers to users' questions, is there any difference in user performance and subjective reactions?
(2) What specific features of the ConVisQA interface are perceived as more/less beneficial by the potential users (e.g., Question Answering, Interface Filters)?

5.2 Methodology

After developing ConVisQA we wanted to understand how the system may support users in performing question answering with conversational data. Since our first research objective requires a comparison between two interfaces. This study was based on two interfaces as conditions:

1. ConVis, a similar interface displaying the same list of conversations.
2. ConVisQA, showing the same list of blog conversations, for finding answers to User's questions with additional features.

A within-subject design was used to conduct this study with these two interfaces as the within-subject factor, allows us to directly compare the measures of each participant for both interfaces.

5.3 Dataset and Preprossessing

We used the ClaimBuster data sets, extracted from U.S. general election presidential debates 2016 [1]. All transcripts are similar in length (number of sentences) (1530 and 1378 and 1572 sentences, respectively).

We pre-processed these data sets to convert it into the required form of conversation for integration with our system. This preprocessing of data was performed using an open-source python environment Anaconda [36]. We selected two data sets for testing with a similar number of turns (306 and 399 respectively) to avoid potential variations due to the conversation length.

5.4 Procedure and Tasks

We performed an online summative study to compare the two interfaces [28]. The study was designed with two interfaces: ConVis, and ConVisQA. The primary reason for including ConVis as baseline interface was to evaluate whether any potential improvements in performance and better user interaction over a typical Question Answering and multi-party blog conversation search and reading through the conversation is due to common visualization design space between ConVis and

ConVisQA, or due to the interactive question answering (which is only present in ConVisQA). For a fair comparison, different interface parameters such as screen size and font size were kept the same across both interfaces. Users were asked to explore the interfaces according to their preferences.

This research was conducted in an online experiment setup to avoid any potential risks associated with the pandemic. The candidates were contacted via email and a meeting time was scheduled. This meeting was conducted over the internet using a video conference tool (e.g. zoom).

We used a within-subject design for this study as a within-subject factor, allows us to directly compare each participant with respect to both interfaces. Each participant used both interfaces and all ten participants were introduced to both interfaces in a different order. To avoid order bias we shuffled the order of interfaces. Both interfaces display a list of debates, where each line-item represents a set of metadata of the conversations, such as the title of conversation and the length of conversation (number of comments).

At the beginning, a pre-study questionnaire was administered to capture demographic information and prior experience with blog reading. Participants accessed the search interface online through a web-link. They were asked to share their computer screen during their tasks. Then, both interfaces were demonstrated to the participants. The experimenter explained the interface actions by following the written script. After that, they were allowed to choose any conversations of their interest from a set of three multi party conversation and explore the selected interface for a given task. Participants were asked to perform a specific task within a certain time period and also explore the conversations according to their own interests. At the end, the experimenter provided an online post-study questionnaire followed by a brief exit interview regarding their experience about search interface. The participants were required to share key insights (if any) gained while exploring each conversation in short exit interview. They performed required task(s) for approximately 60 minutes. During the study,we primarily focused on gathering qualitative data such as observations and semi-structured interviews.

The participants were asked to perform a task on a given conversation (a set of same conversations was provided for each interface). Participants were not asked specific questions, rather they were asked to perform an open-ended task to reflect the exploratory nature of blog reading. In addition, there were two specific questions each debate transcript:

1. Which topics have generated a lot of disagreements(any two topics)?
2. Who disagreed with Bernie about topic X?
3. Which discussion deemed negative, overall?
4. Which author used the most negative words?

5. Who disagreed with Trump about topic X?

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 were asked different questions on the following aspects on a 5-point Likert scale in a post-study questionnaire:

1. Usefulness: 'I found this interface to be useful for finding answers to my questions about the given conversation.
2. Ease of use: 'I found this interface to be easy to use'.
3. Enjoyable: 'I found this interface enjoyable to use'.
4. Find Insightful Comments: 'This interface enabled me to find more insightful comments.
5. Effectiveness: ' I found the interface to be effective for finding answers to my questions about the given conversation'.

5.5 Participants

We conducted this study with ten users (aged 18-50, 5 females and 5 males) who have considerable experience of exploring and reading online conversations and discussions. The participants held a variety of occupations ranging from IT professional, Business, Management, Engineering and Information Systems and students from both graduate and undergraduate levels. The focus audience for this user study is from various backgrounds who are familiar with online platforms and either participate or read frequent users of blogs. The age range was between 18-50. They were rewarded $15 Tim Hortons gift cards for their time.

5.5.1 Results and Analysis

We analyzed the user study sessions and results by triangulating between multiple data collection methods, including observations, notes taken by participants during the analysis session, and semi-structured interviews at the end of each session. We now present our key findings from the sessions.

In the post-study questionnaires, participants evaluated each interface (ConVis and ConVisQA) on a 5-point Likert scale. They provided their ratings for various criterion such as whether the interface was helpful and faster in finding the answers to the questions, how satisfied was the participant with the results displayed by the interface, and if interface filters (such as sentiment and sentence length) were effective in filtering visualization as shown in Figure 5.1

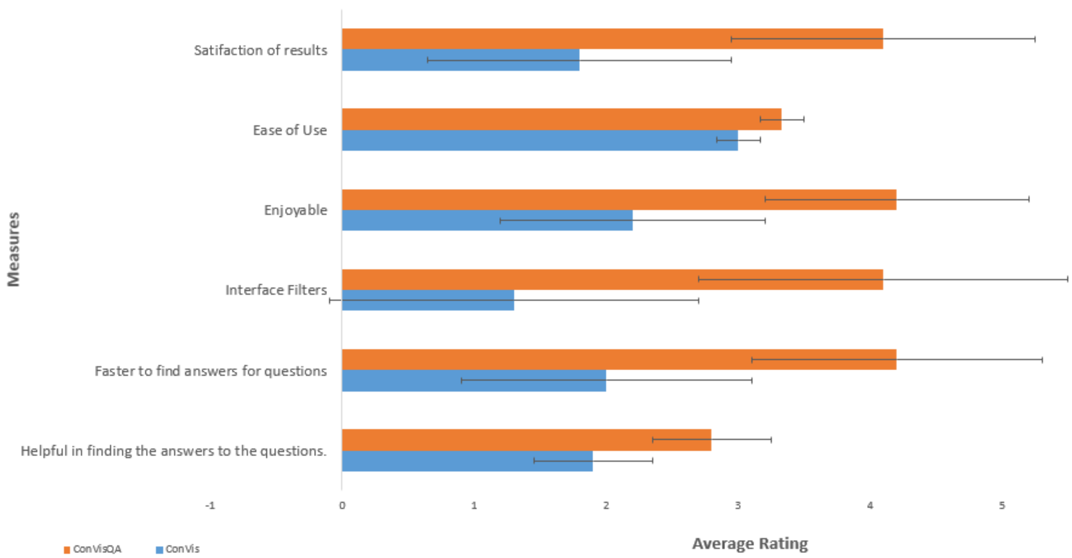


Figure 5.1: Average rating of both interfaces by the participants on six different measures. Longer bars indicate higher rating.

Participants also rated each interface (ConVis and ConVisQA) in a post task questionnaire with respect to four measures (*Usefulness*, *Ease of use*, *Satisfaction*, and *Relevance*) on 5-point Likert as shown in Figure 5.2. Our interface received significantly higher rating on *Usefulness* (Mann-Whitney $U = 18; p < 0.05$), *Satisfaction*($U = 7.0; p < 0.005$), and *Relevance* ($U = 0, p < 0.005$). For the *Ease of use* measure the difference was not significant ($U = 31.5; p = 0.147$). Overall, these results suggest that in comparison to the baseline participants found ConVisQA to be more useful and they found the interface to be more satisfactory that provided

Figure 5.2: User study responses to post-task questionnaires.

more relevant results to their questions.

5.5.2 Subjective Measures

At the end of the study, we asked the participant to fill-up a post-study questionnaire and a short exit interview regarding the usability aspects and suggested improvements of the interface. Five participants suggested that overall ConVisQA was useful for exploring conversations, however, one participant found this interface a bit complex to understand.

Two participants found the interface very easy to use for exploring an online conversation. For the effectiveness of question-answering ability of the interface, they found it to be very useful in retrieving answers given by the interface in analyzing and exploring multi-party conversations (P1 and P3). P4 added that coordination between the views can be more enhanced and interactive that can be evaluated when the interface can answer a question about conversation text .It was also suggested that it would be better if a question can display answers from all pages as in the current demonstration users had to select a question on each page. P5 suggested that Question answering is very interesting and can be enhanced further, for example, I would like to find answers about the conversation such as what are Trump's most abusive lines or comments in this debate. Further, P7 added the search bar should be flexible enough to take any questions or tell explicitly what kind of questions it can answer. Overall eight out of ten participants found this interface to be very helpful in finding answers to their questions. However, there is a need for supporting a larger list of questions and to include questions about the content of blog conversation such as ”What is trump's supported claim in the debate”.

5.5.3 Interaction Patterns

Three participants who first used ConVisQA, started by typing Questions to see how the interface responds to their queries. As they stated that this seems different to have this kind of meta-questions instead of factoid Question Answering. They tried different questions such as two participants started by typing questions in the search bar. As the suggestions for auto-completion are shown they selected one of the suggestions without typing further. Some of the common questions that they selected include “What is the most negative topic?”, “What topics are generating more discussions?”, and “Which user is most dominant in the discussion?”. They suggested that the way ConVisQA presents the results by highlighting topics, comments, and authors in the Thread Overview is very effective. However, one participant suggested that it would be better if Conversational View can only display the comments posted in response the question asked to avoid the need for scrolling down through the Conversation View, For example, one participant selected a question “what is the most negative topic?” and the relevant comment(s) was almost at the end of the conversational view and had to scroll down to read related comments. The third participant started by performing the keyword search, where the participant typed a term and in response the system highlights all the relevant comments containing that term within conversational view. It was found very effective and easy to view the comments of interest for a particular term. The other three participants started exploring the topics and their relevant comments. They explained that they want to get an overview of the whole conversation. Then they explored the visualization by selecting different questions from the list of questions. The rest four participants started with the given task by exploring questions and analyzing their answers.

5.5.4 Reactions to Interface Features

In general, all participants agreed that the way ConVisQA is showing the answers of various questions by connecting all participating facets e.g. authors, topics and that makes it easy to get the answers. It is relatively clear which authors are participating in a discussion or topic while answering a particular question. P1 found "the sentiment filter will be very helpful especially for huge conversations with many comments". P1 also confirmed that "deleting topic is a convenient feature to get rid of the unwanted topic(s) and the user can enjoy only the topics of interest." P2 found that "filtering by sentence length is very helpful to avoid huge conversation trails". P3 also reported "Author search is very useful in locating comments posted by a particular author. The system filters the comments posted by a particular author as most blog readers tend to search for someones' (friends or relative) comments in a blog conversation." P4 found that sentiment overview is very helpful to observe how a debate or discussion has evolved in terms of sentiments just at a glance. Overall, all participants found these interactive features very helpful in quick filtering and visualizing the required amount of data.

5.6 Discussions

We now discuss the implications of our results and the general observation we have from the online user study.

5.6.1 Summary of Findings

Based on the analysis of the results and discussion with participants, we now revisit our evaluation objective mentioned at the beginning of this Chapter. The first question was " Whether ConVisQA helps users in performing information-seeking tasks from multi-party conversations in form of question answering? We observed from the feedback data, the majority of participants who filled up the questionnaires found the interface is useful and felt that it enabled them to find the relevant answers to their questions with ease. Also, the qualitative feedback from participants revealed that their overall impression was quite positive with some further improvements suggested in the questions list. With regards to the second research objective, What specific features of the ConVisQA interface are perceived as more/less beneficial by the potential users. We learned from exit interviews with participants that they were able to find comments of their interest such as most negative comment(s) in a topic in large conversation. They found this feature to be very interesting that enables them to perceive a lot of comments whether they are positive or negative. They suggested it can be improved by adding more sentiment categories such as whether a comment is abusive or hateful.

P7 added that the design of visualization is perfect but can be good if network graphs are added here that can enhance filtering and zooming of conversation.

According to P8, the search bar can be enhanced by some sort of guide or can be made more flexible to take other types of questions. P8 also suggested that the sentiment overview could have been made easier to understand. For example, maybe a filter for sentiment range would help by letting users choose if they want to see it by comment length or sentiment color (positive to negative). In the end, their feedback suggests that add on features within the visualization were found helpful in filtering the visualization according to their needs.

5.6.2 Suggested Improvements

We now reflect upon on our design and evaluation of the ConVisQA interface to summarize the lessons learned that can be generalized to other conversational domains to enhance the usability for novice users. P7 observed it could be difficult to figure out what kind of questions can be asked within this interface without an explanation for a novice user who is not familiar with this interface or similar visualizations and that the interface should be self-explanatory for any task and any user. P7 also added that the sentiment bar reflects three or more sentiment colors that indicate one comment has many sentiments and it can be improved by extracting overall sentiment or opinion of a comment instead of sentence-level sentiment to keep users at ease. Some participants were not familiar with data visualizations, novice users, and initially found it difficult to understand and navigate through such complex visualizations. It might be helpful to explore how can we further simplify the ConVisQA interface without losing key functionalities.

5.6.3 Limitations

There may be some possible limitations in this user study. We ran the study with ten participants only. While this study provides an initial evidence that our system is useful, we still need to run a larger scale study with sufficient number of participants to further confirm the hypothesis.

The possible constraint in recruiting more participants was due to current pandemic situation as we could not recruit enough participants for the user study as we planned initially. However we will further evaluate this system with sufficient number of users to confirm the effectiveness and usability of our system.

6 Conclusion and Future Work

In this dissertation, we combine natural language processing and information visualization to support users in exploring a larger online conversation by asking questions and getting answers in visual form. Our work was motivated by emerging research avenues of question answering and of the challenges arising from the volume and complexity of online conversational data and the shortcomings of existing approaches in dealing with such a challenging research problem. To address the problem, we developed the ConVisQA system which is described in Chapters 3 and 4. Subsequently, we conducted an online user study to demonstrate that our solution can be successfully personalized to a new system for question answering such as problems faced by users in finding a meta overview of a conversation. In this final chapter, we revisit our approach for tailoring an existing visual text analytics systems (Section 6.1) and indicate open research questions and directions for future work (Section 6.2). We conclude the dissertation with some closing remarks and an overview of future work.

6.1 Summary of Contributions

We have presented ConVisQA, an interactive question answering system for analyzing online conversations. We believe that this work provides an initial step towards building an effective natural language interface for exploring and analyzing a large number of online conversations. The primary contributions of this thesis are:

i) We developed a natural language interaction technique for exploring online conversations that allow the user to ask questions about the conversations and get the answer presented in a visual interface. To the best of our knowledge, this is the first attempt at building a natural language interaction system for exploring and making sense of online conversations.

ii) We incorporated our natural language interaction technique by designing ConVisQA which extended a visual interface for exploring conversations ConVis [18]. Given a question, the interface highlights the answer in a visual overview of the conversations and help her to rapidly filter navigate through the relevant comments to fulfill their information needs.

iii) Our online user study with ten participants demonstrated how the ConVisQA interface may be helpful for quickly locating comments of interests to fulfill specific information needs that the user may have.

6.2 Future Work

There are several avenues of this research that we plan to explore in the future. First, currently, we are handling limited types of questions that involve meta-data like topics, authors, and sentiment. In the future, it would be useful to handle more varieties of questions such as factoid and semantic questions related to the content (e.g. why people do not follow social distancing?). More content analysis techniques such as argument mining [30] and question-answer similarity measures [20], [29] could help the system to answer such questions. Second, we would like to handle additional challenges while parsing the question including the ambiguities in natural language. We would like to enhance auto-completion features to help users formulate the questions as user types and resolves ambiguities. Third, while this work provides an initial idea of question answering using hand-crafted grammar, we are building large corpora question-answer pairs so that we can apply more advance deep learning models to automatically learn grammar rules in a supervised fashion. On scalability, while ConVisQA can deal with conversations with hundreds of comments, additional techniques are required for larger conversations. We would like to explore advanced visualization techniques to deal with such large scale conversations.

We would also like to further explore who are the other types of users who might be willing to use such an advanced interface? For this purpose, we aim to conduct user studies with more focused pool of users such as journalists, fact checkers, content screeners, editors and law enforcement officials to evaluate the interface to its real potential and usability. Finally, we would like to validate our system extensively among real users through longitudinal studies and field trials.

Bibliography

[1] Fatma Arslan, Naeemul Hassan, Chengkai Li, and Mark Tremayne. A benchmark dataset of check-worthy factual claims. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 821–829, 2020.

[2] David Bawden and Lyn Robinson. The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of information science*, 35(2):180–191, 2009.

[3] Tony Bergstrom and Karrie Karahalios. Conversation clusters: grouping conversation topics through human-computer dialog. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2349–2352, 2009.

[4] David M Blei. Probabilistic topic models. *Communications of the ACM*, 55 (4):77–84, 2012.

[5] Weiwei Cui, Shixia Liu, Li Tan, Conglei Shi, Yangqiu Song, Zekai Gao, Huamin Qu, and Xin Tong. Textflow: Towards better understanding of evolving topics in text. *IEEE transactions on visualization and computer graphics*, 17(12): 2412–2421, 2011.

[6] Kristina Toutanova Dan, Dan Klein, Christopher D Manning, and Yoram Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *In Proceedings of HLT-NAACL 2003*. Citeseer, 2003.

[7] Kushal Dave, Martin Wattenberg, and Michael Muller. Flash forums and forumreader: navigating a new kind of large-scale online discussion. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 232–241, 2004.

[8] Nicholas Diakopoulos, Mor Naaman, and Funda Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *2010*

IEEE Symposium on Visual Analytics Science and Technology, pages 115–122. IEEE, 2010.

[9] Marian Dörk, Daniel Gruen, Carey Williamson, and Sheelagh Carpendale. A visual backchannel for large-scale events. *IEEE transactions on visualization and computer graphics*, 16(6):1129–1138, 2010.

[10] Mennatallah El-Assady, Valentin Gold, Carmela Acevedo, Christopher Collins, and Daniel Keim. Contovi: Multi-party conversation exploration using topic-space views. In *Computer Graphics Forum*, volume 35, pages 431–440. Wiley Online Library, 2016.

[11] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1175–1184, 2010.

[12] Ronen Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.

[13] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G Karahalios. Datatone: Managing ambiguity in natural language interfaces for data visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 489–500, 2015.

[14] Pollyanna Gonçalves, Matheus Araújo, Fabrício Benevenuto, and Meeyoung Cha. Comparing and combining sentiment analysis methods. In *Proceedings of the first ACM conference on Online social networks*, pages 27–38, 2013.

[15] Xia Gong, Xiangyi Kong, Zhujun Zhang, Lin Tan, Zixiong Zhang, and Bing Shao. Customer service automatic answering system based on natural language processing. In *Proceedings of the 2019 International Symposium on Signal Processing Systems*, pages 115–120, 2019.

[16] Susan Havre, Beth Hetzler, and Lucy Nowell. Themeriver: Visualizing theme changes over time. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, pages 115–123. IEEE, 2000.

[17] Lynette Hirschman and Robert Gaizauskas. Natural language question answering: the view from here. *natural language engineering*, 7(4):275, 2001.

[18] Enamul Hoque and Giuseppe Carenini. Convis: A visual text analytic system for exploring blog conversations. In *Computer Graphics Forum*, volume 33, pages 221–230. Wiley Online Library, 2014.

[19] Enamul Hoque and Giuseppe Carenini. Interactive topic hierarchy revision for exploring a collection of online conversations. *Information Visualization*, 18 (3):318–338, 2019.

[20] Enamul Hoque, Shafiq Joty, Màrquez Lluís, and Giuseppe Carenini. CQAVis: Visual text analytics for community question answering. In *Proceedings of the ACM conference on Intelligent User Interfaces (IUI)*, pages 161–172, 2017.

[21] Enamul Hoque, Vidya Setlur, Melanie Tory, and Isaac Dykeman. Applying pragmatics principles for interaction with visual analytics. *IEEE transactions on visualization and computer graphics*, 24(1):309–318, 2017.

[22] Quentin Jones, Gilad Ravid, and Sheizaf Rafaeli. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research*, 15:194–210, 06 2004. doi: 10.1287/isre.1040.0023.

[23] Quentin Jones, Gilad Ravid, and Sheizaf 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.

[24] Shafiq Joty, Giuseppe Carenini, and Raymond T Ng. Topic segmentation and labeling in asynchronous conversations. *Journal of Artificial Intelligence Research*, 47:521–573, 2013.

[25] Daniel Jurafsky and James H Martin. Speech and language processing: An introduction to speech recognition, computational linguistics and natural language processing. *Upper Saddle River, NJ: Prentice Hall*, 2008.

[26] Amirkiarash Kiani, Sameh Al-Natour, and Ozgur Turetken. A comparison of sentiment analysis tools. *Association for Information Systems*, 2018. URL `https://aisel.aisnet.org/amcis2018/DataScience/Presentations/27`.

[27] Jayant Krishnamurthy, Pradeep Dasigi, and Matt Gardner. Neural semantic parsing with type constraints for semi-structured tables. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1516–1526, 2017.

[28] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE transactions on visualization and computer graphics*, 18(9):1520–1536, 2011.

[29] Md Tahmid Rahman Laskar, Xiangji Huang, and Enamul Hoque. Contextualized embeddings based transformer encoder for sentence similarity modeling in answer selection task. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 5505–5514, 2020.

[30] John Lawrence and Chris Reed. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818, 2020.

[31] Sweta P Lende and MM Raghuwanshi. Question answering system on education acts using nlp techniques. In *2016 world conference on futuristic trends in research and innovation for social welfare (Startup Conclave)*, pages 1–6. IEEE, 2016.

[32] Isa Maks and Piek Vossen. A lexicon model for deep sentiment analysis and opinion mining applications. *Decision Support Systems*, 53(4):680–688, 2012.

[33] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

[34] Adam Marcus, Michael S Bernstein, Osama Badar, David R Karger, Samuel Madden, and Robert C Miller. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 227–236, 2011.

[35] Paloma Moreda, Hector Llorens, Estela Saquete, and Manuel Palomar. Combining semantic information in question answering systems. *Information Processing & Management*, 47(6):870–885, 2011.

[36] NA. Anaconda software distribution, 2020. URL `https://docs.anaconda.com/`.

[37] Terence Parr. *The definitive ANTLR 4 reference*. Pragmatic Bookshelf, 2013.

[38] Victor Pascual-Cid and Andreas Kaltenbrunner. Exploring asynchronous online discussions through hierarchical visualisation. In *2009 13th International Conference Information Visualisation*, pages 191–196. IEEE, 2009.

[39] Panupong Pasupat and Percy Liang. Compositional semantic parsing on semi-structured tables. *arXiv preprint arXiv:1508.00305*, 2015.

[40] Muthukrishnan Ramprasath and Shanmugasundaram Hariharan. A survey on question answering system. *International Journal of Research and Reviews in Information Sciences*, 2(1):171–179, 2012.

[41] Warren Sack. Conversation map: An interface for very large-scale conversations. *Journal of Management Information Systems*, 17(3):73–92, 2000.

[42] V. Setlur, Enamul Hoque, D. H. Kim, and Angel X. Chang. Sneak pique: Exploring autocompletion as a data discovery scaffold for supporting visual analysis. *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020.

[43] Vidya Setlur, Sarah E Battersby, Melanie Tory, Rich Gossweiler, and Angel X Chang. Eviza: A natural language interface for visual analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 365–377, 2016.

[44] Arjun Srinivasan and John Stasko. Orko: Facilitating multimodal interaction for visual exploration and analysis of networks. *IEEE transactions on visualization and computer graphics*, 24(1):511–521, 2017.

[45] Internet Live Stats. Twitter usage statistics. URL `https://www.internetlivestats.com/twitter-statistics/`.

[46] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

[47] Pew Research CenterInternet Technology. Social media fact sheet. URL `https://www.pewresearch.org/internet/fact-sheet/social-media/`.

[48] Y Teh. W, jordan m, i, beal m, j, et al. *Hierarchical dirichlet processes*, 101 (476):10–1198, 2006.

[49] Maria Vargas-Vera and Enrico Motta. Aqua–ontology-based question answering system. In *Mexican International Conference on Artificial Intelligence*, pages 468–477. Springer, 2004.

[50] Julita Vassileva and Carl Gutwin. Exploring blog archives with interactive visualization. In *Proceedings of the working conference on Advanced visual interfaces*, pages 39–46, 2008.

[51] Gina Danielle Venolia and Carman Neustaedter. Understanding sequence and reply relationships within email conversations: a mixed-model visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 361–368, 2003.

[52] Fernanda B Viégas, Scott Golder, and Judith Donath. Visualizing email content: portraying relationships from conversational histories. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 979–988, 2006.

[53] Xiting Wang, Shixia Liu, Junlin Liu, Jianfei Chen, Jun Zhu, and Baining Guo. Topicpanorama: A full picture of relevant topics. *IEEE transactions on visualization and computer graphics*, 22(12):2508–2521, 2016.

[54] Zhaoxia Wang, Chee Seng Chong, Landy Lan, Yinping Yang, Seng Beng Ho, and Joo Chuan Tong. Fine-grained sentiment analysis of social media with emotion sensing. In *2016 Future Technologies Conference (FTC)*, pages 1361–1364. IEEE, 2016.

[55] Martin Wattenberg and David Millen. Conversation thumbnails for large-scale discussions. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 742–743, 2003.

[56] Furu Wei, Shixia Liu, Yangqiu Song, Shimei Pan, Michelle X Zhou, Weihong Qian, Lei Shi, Li Tan, and Qiang Zhang. Tiara: a visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 153–162, 2010.

[57] Yingcai Wu, Shixia Liu, Kai Yan, Mengchen Liu, and Fangzhao Wu. Opinionflow: Visual analysis of opinion diffusion on social media. *IEEE transactions on visualization and computer graphics*, 20(12):1763–1772, 2014.

[58] Bo Yuan, Ying Liu, and Hui Li. Sentiment classification in chinese microblogs: lexicon-based and learning-based approaches. *International Proceedings of Economics Development and Research*, 68:1, 2013.

[59] Jian Zhao, Nan Cao, Zhen Wen, Yale Song, Yu-Ru Lin, and Christopher Collins. # fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE transactions on visualization and computer graphics*, 20 (12):1773–1782, 2014.

[60] Inc. 1996 2020 © Alexa Internet. The top 500 sites on the web. URL `https://www.alexa.com/topsites`.

7 Appendix

Supplementary Materials for Chapter 4 This appendix contains supplemental materials for Chapter 4.The script used by the experimenter to run the study and questionnaires used during the study.

1. Appendix A: Online Consent Form
2. Appendix B: Script for User Study
3. Appendix C: Pre Study Questionnaire
4. Appendix D: Post Study Questionnaire
5. Appendix E: Post Task Questionnaire

7.1 Appendix A: Online Consent Form

Online Consent Purpose of the Research: In this research, we are interested in studying how a user can explore information available on micro blogging sites. It's is quite common now a days that people from around the globe share information on such platforms, hence, these micro blogging websites carry proliferated data volume resulting in huge threads of discussion over one topic or subject. It could be an overwhelming experience for user to find the relevant information. We are interested to build a search Tool, demonstrating generic behavior to find and display specific information to answer user's questions. The interface can display information relevant to search query and present only relevant communication threads. This interface will provide some additional features of sorting and filtering.

What You Will Be Asked to Do in the Research: As a research participant, you are requested to complete this Pre study questionnaire survey by providing answers that best match your opinion. The study should take approximately 60 minutes to complete. After completing this survey, each participant will be asked to perform a task on a given conversation for two interfaces ConVis and ConVisQA. A set of same conversations will be provided for each interface. Participants will be asked to perform an open-ended task to reflect the exploratory nature of blog reading. Also, you will be given one or two specific questions to explore the conversation. You are required to explore the interfaces and complete the task. After completing this task, you ae required to write a summary about your experience and the experimenter will take your feedback as informal interview. Risks and Discomforts: We do not foresee any risks or discomfort from your participation in the research.

Benefits of the Research and Benefits to You: By completing this experimental study students or participants may learn about different techniques about data visualization. They might experience how easy information visualization could be in a particular form of bisualizaton. Also they may enhance their understanding of how a case study is performed.

Voluntary Participation and Withdrawal: Your participation in the study is completely voluntary and you may choose to stop participating at any time. Your decision not to volunteer, to stop participating, or to refuse to answer particular questions will not influence the nature of the ongoing relationship you may have with the researchers or study staff, or the nature of your relationship with York University either now, or in the future.

In the event you withdraw from the study, all associated data collected will be immediately destroyed wherever possible. The participants will receive a 15 dollar gift card even if a participant decides to withdraw from the study.

Confidentiality: The survey itself will not ask for any identifying information.

The data is completely anonymous. The data will be securely stored. The data will be stored until August 31, 2021. At which point, only aggregate data will be kept. The electronic data will be removed, and any paper copies will be shredded.

Unless you choose otherwise, all the information you supply during the research will be held in confidence and unless you specifically indicate your consent, your name will not appear in any report or publication of the research. Your data will be safely stored in a locked facility and only the researcher will have access to this information. Confidentiality will be provided to the fullest extent possible by law.

The data collected in this research project may be used – in an anonymized form - by members of the research team in subsequent research investigations exploring similar lines of inquiry. Such projects will still undergo ethics review by the HPRC, our institutional REB. Any secondary use of anonymized data by the research team will be treated with the same degree of confidentiality and anonymity as in the original research project.

The researcher(s) acknowledge that the host of the online survey (e.g., Qualtrix, Survey Monkey Microsoft forms etc.) may automatically collect participant data without their knowledge (i.e., IP addresses.) Although this information may be provided or made accessible to the researchers, it will not be used or saved without participant's consent on the researcher's system. Further, "Because this project employs e-based collection techniques, data may be subject to access by third parties as a result of various security legislation now in place in many countries and thus the confidentiality and privacy of data cannot be guaranteed during web-based transmission.

Questions About the Research? If you have questions about the research in general or about your role in the study, please feel free to contact me at dipro@yorku.ca or my supervisor, Marin Litoiu at marin@yorku.ca and/or 416 736 2100 x20987. You may also contact the Program in Information Technology at lapsitec@yorku.ca and/or 416 736 2100 x40797.

This research has received ethics review and approval by the Delegated Ethics Review Committee, which is delegated authority to review research ethics protocols by the Human Participants Review Sub-Committee, York University's Ethics Review Board, and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines. If you have any questions about this process, or about your rights as a participant in the study, please contact the Sr. Manager Policy Advisor for the Office of Research Ethics, 5th Floor, Kaneff Tower, York University (telephone 416-736-5914 or e-mail ore@yorku.ca). Legal Rights and Signatures:

I ____

consent to participate in A natural language question answering system for exploring online conversations conducted by Nadia Siddiqui. I have understood the nature of this project and wish to participate. I am not waiving any of my legal rights by signing this form. My signature below indicates my consent.

Signature ______________________________ **Date** __________

Participant

Signature ______________________________ **Date** __________

Principal Investigator |

Figure 7.1: User Consent Signature

7.2 Appendix B: Script for User Study

1. STEP 1: PARTICIPANT GREETING

   Tell Participant:

   "Thank you for participating in our study. The whole process today will last approximately 60 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 multi party conversations and writing short summaries about your experience (Writing short summary is an optional task). We have two interfaces to evaluate in this study (ConVis and ConVisQA). At the end of the study, you will be given a short post-study questionnaire and a brief exit interview regarding your experience about the search interfaces. Please make sure to enter correct email address as your reward gift card will be sent you using the email address provided."

2. STEP 1: PRE-STUDY QUESTIONNAIRES

   Tell participant: " Action: Give and overview of online consent and tell them that they will have to agree before starting pre-study questionnaire. Participant's consent form will be used as the Record of Participation in this study. Now we will have you fill a pre-study questionnaire."

   Tell participant: "Please fill up the following questionnaire."

   Action: "Open up user pre study questionnaire that contains consent form, signs online consent form and provide the name and email address. The user will fill up the pre-study questionnaire, then select interface".

3. STEP 2: 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: Open the interface with a sample dataset and demonstrate the key features of the interface.

   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 Conversation Mode: - Hovering the mouse over a facet element - related comments and facets are highlighted - tool tips 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 - related comments are highlighted in conversation view

The participants will be given some time to explore the conversation.

4. STEP 3: Select Task

   Please read the following task that is for any debate related dataset:

   Now you have to perform an open-ended task to reflect the exploratory nature of multi party conversation reading. Can you answer following questions:

   a. Which topics have generated a lot of disagreements (any two topics)? b. Who disagreed with sanders about topic X? c. what is your perception about overall sentiment of the conversation (e.g. positive, negative)?

   Participants will have about 15-20 minutes to perform this task. After this task they will answer few questions of researcher.

5. STEP 4: Switch Interface

   Now the participants will be asked to switch interface and will open the data sets list for ConVisQA. As soon as participant selects a dataset from the list, the list is replaced by ConVisQA.

6. STEP 4: POST STUDY QUESTIONNAIRE At the end of all the tasks, the participant will fill up post-study questionnaire and a brief exit interview.

7. STEP 6: DEBRIEFING

   Tell Participant: “Thank you very much again for your participation. Would you have any other comments or questions?”

8. Action: Send Payment by email and receive email for receiving the compensation of participation.

7.3 Appendix C: Pre Study Questionaire

ConVisQA- Pre-Study Questionnaire

A Natural Language Interface for Exploring Online Conversations

* Required

Consent to participate in this study

Researcher: Nadia Siddiqui
Student,
School of Information Technology,
York University
3052 Victor Phillip Dahdaleh Building (DB)
Toronto, Ontario, M3J 1P3 Email: nasidiqi@yorku.ca (mailto:nasidiqi@yorku.ca)

Purpose of the Research: In this research, we are interested in studying how a user can explore information available on microboring sites. It's is quite common now a days that people from around the globe share information on such platforms, hence, these microblogging websites carry proliferated data volume resulting in huge threads of discussion over one topic or subject. It could be an overwhelming experience for user to find the relevant information. We are interested to build a system, demonstrating generic behavior to locate and visualize specific information to answer user's questions.

In this research, we are extending ConVis, a visual interface that was originally developed for visually exploring conversations we are extending an existing natural language interface ConVis, for enhanced functional features to support natural language interaction for exploring online conversations that enables the user to ask questions about the conversations and get the answer presented in a visual interface. To the best of our knowledge, this is the first attempt of building a natural language interaction system for exploring and making sense of online conversations. We expect that ConVisQA interface may be helpful for quickly locating comments of interests to fulfill specific information needs that the user may have

For this purpose, we need to evaluate and validate the new functional features incorporated in ConVisQA. During the study, we will collect quantitative data such as the time it takes to complete the search task, the interaction log data (mouse and keyboard interactions) as well as subjective reactions collected in the form of questionnaire. Next, we are interested to analyze the data from the study to further improve our new search interface for information search.

Risks and Discomforts: We do not foresee any risks or discomfort from your participation in the research.

Benefits of the Research and Benefits to You: Include a statement regarding any benefits of the research as well as benefits to the research participants.

Voluntary Participation and Withdrawal: Your participation in the study is completely voluntary and you may choose to stop participating at any time. Your decision not to volunteer, to stop participating, or to refuse to answer particular questions will not influence the nature of the ongoing relationship you may have with the researchers and York University either now, or in the future. If you stop participating, you will still be eligible to receive the promised pay/compensation for agreeing to be in the project. In the event you withdraw from the study, all associated data collected will be immediately destroyed wherever possible.

Consent

Confidentiality: All personally identifying information you supply during the research – mainly: your name and email address -- will be held in confidence and your name will not appear in any report or publication of the research. Your data will be collected though the use of this on-line instrument. Your personally identifying data will be safely stored in a locked facility and a password protected digital medium (USB key or DVD) and only research staff/research team members will have access to this information. The personally identifying data will be destroyed by the end of 2020. Confidentiality will be provided to the fullest extent possible by law.

Your non-personally identifying responses -- including: demographics (gender, age-range, educational background), interaction log data and post-study questionnaire responses -- may be published on the PI's web-sites (personal, lab) and shared with the research community as data files. The PI will screen the data and use his own judgement to identify accidentally or indirectly personally identifying features and will redact the parts in question or, if necessary, the entire response prior to availing it to third parties. The anonymized non-personally identifying data collected in this research project may also be used by members of the research team in subsequent research investigations exploring similar lines of inquiry.
The researcher(s) acknowledge that the host of the online survey (e.g., Surveygizmo, survey monkey) may automatically collect participant data without their knowledge (i.e., IP addresses.) Although this information may be provided or made accessible to the researchers, it will not be used or saved without participant's consent on the researchers' system. Further, because this project employs e-based collection techniques, data may be subject to access by third parties as a result of various security legislation now in place in many countries and thus the confidentiality and privacy of data cannot be guaranteed during web-based transmission.

Questions About the Research? If you have questions about the research in general or about your role in the study, please feel free to contact Ms. Nadia Siddiqui by e-mail (nasidiqi@yorku.ca (mailto:nasidiqi@yorku.ca)). This research has received ethics review and approval by the Human Participants Review Sub-Committee, York University's Ethics Review Board and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines. If you have any questions about this process, or about your rights as a participant in the study, please contact the Sr. Manager & Policy Advisor for the Office of Research Ethics, 5th Floor, Kaneff Tower, York University (telephone 416-736-5914 or e-mail ore@yorku.ca (mailto:ore@yorku.ca)).

1. • I have read and understood the above and consent to participate in this study.
   • I consent to participate in User Study of a system "ConVisQA: A Natural Language Interface for Exploring Online Conversations", conducted by Nadia Siddiqui. I have understood the nature of this project and wish to participate. I am not waiving any of my legal rights by signing this form. By clicking 'I Agree', below I am indicating consent. *

   ◯ I agree

2. Please select your age range *

- ◯ 18-24
- ◯ 25-34
- ◯ 35-44
- ◯ 45-54
- ◯ 55+

3. Please indicate your Gender. *

- ◯ Male
- ◯ Female
- ◯ Prefer not to say

4. What is your occupation? *

5. What is your field of study (if you are a student): *

6. How often do you read blogs and social media sites (e.g. reddit, facebook, twitter)? *

- ◯ Never
- ◯ Rarely (several times a year)
- ◯ Occasionally (several times a month)
- ◯ Frequently(everyday)
- ◯ Very frequently (several times a day)

7. What is the purpose of visiting these sites (blogs and social media)? *

- To obtain knowledge
- Entertainment
- Seeking Information
- Seeking guidance/ opinion
- Other

8. How often do you successfully find the required information on traditional blogs ? *

- Always
- Usually
- Sometimes
- Never

9. How often do you read other people's comments in a blog conversation ? *

- ◯ Daily
- ◯ Weekly
- ◯ Sometimes
- ◯ Never

Name and email address

Your name and email address will be held in confidence and your name will not appear in any report or publication of the research. It might be used only for remuneration.

10. Name *

11. Email Address *

7.4 Appendix D: Post Study Questionnaire

ConVisQA- Post Study Questionnaire

A Natural Language Interface for Exploring Online Conversations

* Required

This questionnaire is about your experience of using this interface to explore a blog.

1. How easy to use this user interface for reading blogs?

- ◯ Very Difficult
- ◯ Difficult
- ◯ Neither easy nor hard
- ◯ Easy
- ◯ Very Easy

2. How useful is this interface for reading blogs? *

- Not useful at all
- Slightly Useful
- Very Useful
- Extremely useful
- Somewhat useful

3. This interface makes it faster for me to find the information I was looking for.

- Strongly Disagree
- Disagree
- Somewhat Agree
- Agree
- Strongly Agree

4. How helpful were the answers given by the interface in analyzing multi-party conversations?

- ◯ Not helpful at all
- ◯ Somewhat helpful
- ◯ Neither helpful nor unhelpful
- ◯ Moderatly unhelpful
- ◯ Very helpful
- ◯ Extremely helpful

5. I found this interface to be useful for finding answers to my questions about the given conversation ?

- ◯ Not useful at all
- ◯ Slightly useful
- ◯ Somewhat useful
- ◯ Very useful
- ◯ Extremely useful

6. I found this interface enjoyable to use.

- ◯ Not enjoyable at all
- ◯ Slightly enjoyable
- ◯ Somewhat enjoyable
- ◯ Very enjoyable
- ◯ Extremely enjoyable

7. Would you prefer ConVisQA interface over traditional blog websites to read and explore conversation? *

- ◯ Yes
- ◯ No
- ◯ Maybe

8. Do you have another question in your mind that you would like ConVisQA to answer? *

9. Do you have any suggestions for improvement (design or features) in ConVisQA *

10. Name *

11. Email address