



# SUBMISSION OF WRITTEN WORK

Class code: **1821019-Autumn 2016**

Name of course: **Adding context to online discussions through the generation of user profiles (Autumn 2016)**

Course manager: **N/A (Master's Thesis)**

Course e-portfolio: **N/A (Master's Thesis)**

Thesis or project title: **Adding context to online discussions through the generation of user profiles**

Supervisor: **Paolo Tell, Steven Jeuris, Lucian Leahu**

Full Name:

Birthdate (dd/mm-yyyy):

E-mail:

- |                      |                   |                     |
|----------------------|-------------------|---------------------|
| 1. <b>Simon Gray</b> | <b>02/06-1988</b> | <b>simg</b> @itu.dk |
| 2. _____             | _____             | _____ @itu.dk       |
| 3. _____             | _____             | _____ @itu.dk       |
| 4. _____             | _____             | _____ @itu.dk       |
| 5. _____             | _____             | _____ @itu.dk       |
| 6. _____             | _____             | _____ @itu.dk       |
| 7. _____             | _____             | _____ @itu.dk       |

# Adding context to online discussions through the generation of user profiles

## Abstract

Online platforms facilitate discussions with strangers, but some researchers argue that they also normalise incivility. This thesis explores a concept based on improving online discussions by building and displaying brief summary profiles of otherwise anonymous users on Reddit, one of the world's biggest online discussion platforms. Literature on psychology and online communication was used to inform the concept's design, while the area of Natural Language Processing (NLP) forms the technological foundation of the implemented system. To inform the design and establish the feasibility of the concept, an initial study was conducted with a mock-up of the concept and 10 participants. Two separate implementations of a summary profile creation algorithm were then built. The efficacy of the second and final summary profile creation algorithm was evaluated in a second study featuring 5 participants. The second study reaffirmed the feasibility of the concept and highlighted both the strengths and weaknesses of the system. It also revealed technical limitations that must be solved before building and testing a prototype of the full concept.

# Preamble

When I set out to write this thesis, I was in a vulnerable place. At the time, it was less than a year since I had quit my student job and requested medical leave from ITU. I was suffering from serious chronic stress. Once I started to regain my strength, I knew I wanted to make my thesis about helping other people deal with erratic emotion, the same way I had to learn to deal with my symptoms. After realising how much time I spent discussing – and even arguing – with strangers on the Internet, this thesis eventually evolved into a concept centred around just that.

## Acknowledgements

I am grateful for the supervision I received from my three supervisors: Steven, Paolo, and Lucian. I would also like to give my thanks to Frederik Klovborg, who helped me design the initial study, the people of pitlab who supplied laughs and interesting conversations along the way, and all of the participants who took time out to help with the studies.

## Notes

When reading this thesis, do note that references to specific Java classes or objects created from classes are always written capitalised and in CamelCase. When referencing conceptual or abstract representations, I use lower case. For example, the word “statement” is heavily used in this thesis. When I write “statement” (not capitalised) I am referring to a more fluid concept based around the dictionary definition, while the word “Statement” (capitalised) refers to a specific class or object of that class. I have tried as much as possible to follow words like “Statement” with the word “object” or “class” to make it clear when I am referencing a specific Java class.

The code for the project is available at Github. The version that is used in this thesis is v0.1: <https://github.com/simongray/StatementAnnotator>

# Table of Contents

[Abstract](#)

[Preamble](#)

[Acknowledgements](#)

[Notes](#)

[Table of Contents](#)

[1 Introduction](#)

[1.1 Concept](#)

[1.2 Research Focus](#)

[2 Theoretical Background](#)

[2.1 The Similarity Effect](#)

[2.2 The Hyperpersonal Perspective](#)

[2.3 Online Discussions](#)

[2.3.1 Reddit](#)

[2.3.2 Persuasion Dynamics](#)

[2.4 Summary](#)

[3 Related Work](#)

[3.1.1 ToneCheck](#)

[3.1.2 Reddit Profiling Tools](#)

[4 Design](#)

[4.1 Motivation](#)

[4.2 Real-time Opinion Matching](#)

[4.3 Using Reddit](#)

[4.4 Research Methodology](#)

[5 Initial Feasibility Study](#)

[5.1 Method](#)

[5.1.1 Participants](#)

[5.1.2 Commenting Interface](#)

[5.2 Procedure](#)

[5.3 Limitations](#)

[5.4 Results](#)

[5.4.1 High Efficacy of “High School” Info](#)

[5.4.2 Diverging Impressions of User 3](#)

[5.4.3 Different Perceptions of Relevance](#)

[5.4.4 Difficulty Understanding the First Comment](#)

[5.4.5 Other Results](#)

[5.4.6 Summary](#)

[5.5 Discussion](#)

[5.5.1 Concluding Remark](#)

[6 First Approach: Sentiment Targets](#)

[6.1 Literature Review](#)

[6.1.1 Sentiment Analysis](#)

[6.1.2 Sentiment Towards Entities](#)

6.2	<a href="#">Sentiment Target Algorithm</a>
6.2.1	<a href="#">The Framework of Choice</a>
6.2.2	<a href="#">Navigating the Parse Tree</a>
6.3	<a href="#">Evaluation of Approach</a>
7	<a href="#">Second Approach: Statement Extraction</a>
7.1	<a href="#">Literature Review</a>
7.1.1	<a href="#">Dependency Grammar</a>
7.1.2	<a href="#">Universal Dependencies</a>
7.2	<a href="#">Statements</a>
7.2.1	<a href="#">Statement Annotator</a>
7.2.2	<a href="#">Information Extraction</a>
7.3	<a href="#">The Statement Concept</a>
7.4	<a href="#">Statement Extraction Algorithm</a>
7.4.1	<a href="#">Component Extraction</a>
7.4.2	<a href="#">Component Reduction</a>
7.4.3	<a href="#">Component Linking</a>
7.4.4	<a href="#">Statement Creation</a>
7.5	<a href="#">Statement Patterns</a>
7.5.1	<a href="#">Patterns In Use</a>
7.6	<a href="#">Evaluating Statements for Display</a>
7.6.1	<a href="#">Interestingness</a>
7.6.2	<a href="#">Lexical density</a>
7.6.3	<a href="#">Quality</a>
7.6.4	<a href="#">Relevance</a>
7.7	<a href="#">Limitations</a>
7.7.1	<a href="#">Generalisability</a>
7.7.2	<a href="#">Lexical and Syntactic Assumptions</a>
7.7.3	<a href="#">Other Limitations</a>
8	<a href="#">Efficacy Study</a>
8.1	<a href="#">Method</a>
8.1.1	<a href="#">Participants</a>
8.1.2	<a href="#">Data Sets</a>
8.2	<a href="#">Procedure</a>
8.2.1	<a href="#">First Part: Personal Profile</a>
8.2.2	<a href="#">Second Part: Statement Selection</a>
8.3	<a href="#">Limitations</a>
8.3.1	<a href="#">Level of Abstraction</a>
8.3.2	<a href="#">Limited Number of Participants</a>
8.3.3	<a href="#">Data Set Size</a>
8.3.4	<a href="#">Language</a>
8.3.5	<a href="#">Simplified Presentation</a>
8.3.6	<a href="#">Vagueness of Definitions</a>
8.3.7	<a href="#">Errata</a>
8.4	<a href="#">Quantitative Results</a>
8.4.1	<a href="#">System Overlap</a>
8.4.2	<a href="#">Participant Overlap</a>

	<a href="#"><u>8.4.3 Shared Entities</u></a>
8.5	<a href="#"><u>Interview Results</u></a>
	<a href="#"><u>8.5.1 Partially Accurate Profiles</u></a>
	<a href="#"><u>8.5.2 Reviewing the System Results</u></a>
	<a href="#"><u>8.5.3 Selecting Based on Non-neutrality</u></a>
	<a href="#"><u>8.5.4 Topic Bias</u></a>
	<a href="#"><u>8.5.5 The Presence of Tone</u></a>
	<a href="#"><u>8.5.6 Summary</u></a>
8.6	<a href="#"><u>Discussion</u></a>
	<a href="#"><u>8.6.1 Viability of Small Data Sets</u></a>
	<a href="#"><u>8.6.2 System Efficacy</u></a>
	<a href="#"><u>8.6.3 Lessons Learned from Overlapping Sentences</u></a>
	<a href="#"><u>8.6.4 The Importance of Tone</u></a>
	<a href="#"><u>8.6.5 Concluding Remark</u></a>
9	<a href="#"><u>Conclusion</u></a>
	<a href="#"><u>9.1 Summary</u></a>
	<a href="#"><u>9.2 Threats to Validity</u></a>
	<a href="#"><u>9.3 Future Work</u></a>
	<a href="#"><u>Bibliography</u></a>

# 1 Introduction

In both popular and academic literature, online discussions have been singled out as less constructive – and perhaps even more hostile – than face-to-face conversations, particularly when conversing with strangers as opposed to friends and acquaintances. Hmielowski, Hutchens, and Cicchirillo (2014) argue that this online animosity helps to normalise incivility in our societies by socialising individuals to see flaming as acceptable behaviour.

Online discussions are often cleaned of antisocial comments by moderators and – more recently – by democratic voting systems to express approval or disapproval. Yet despite these measures, online discussions are still marked by negative emotional outbursts and a distinct lack of civility compared to face-to-face discussions.

This thesis is an attempt to apply psychology and computational linguistics to the issue. Using theories from the social sciences to shape my methodology, I have conceptualised and implemented a piece of software to help cultivate civility in online discussions. The software extracts, structures, and presents relevant information from the comment histories of users on online discussions platforms using technologies from the area of Natural Language Processing (NLP).

## 1.1 Concept

The purpose of the software is to display information to readers of comments on online discussion platforms. This information is a short summary profile of the author of any comment found on such a platform. The intent of the summary profile is to influence the impression of the comment author that is held by the reader, so that this impression becomes less negative. This should in turn lead the reader to reply in a more civil and constructive way to comments that the reader might disagree with.

## 1.2 Research Focus

This section documents the overall focus of this thesis. I have been working to answer the following research question:

- **What are the challenges in designing and implementing a tool that generates and displays information with the intention of inducing more civil discussions?**

All throughout the writing of this thesis, my focus has been on discovering how to arrive at a technological solution that could be used to implement the concept described in the [Concept](#) section.

The project is exploratory in nature and I have not been committed to any specific technological approach throughout the design and implementation phase. Nevertheless, all

of the technological contributions in this thesis fall somewhere within the fields of Natural Language Processing (NLP) and Information Extraction (IE).

The concept for the tool is grounded in theories from social psychology and communication. This theoretical grounding is described in detail in the [Theoretical Background](#) section.

The concept was tested in an initial qualitative user study: a short experiment using a mockup of the system and a follow-up interview involving a small number of participants.

Following the initial user study, two separate technological approaches were attempted.

The first approach was based on implementing an algorithm to combine sentiment analysis at the sub-phrasal level with named entity tagging. It was inspired by an existing algorithm (Biyani, Caragea, and Bhamidipati, 2015). The “sentiment targets” found using this approach would then be used to construct bullet points of likes and dislikes for the concept application. I abandoned the approach after developing an initial prototype, being unsatisfied with the accuracy of the sentiment analysis results and the limitations of only using named entities.

The second approach was based on structuring dependency parser output for pattern matching. This approach does not extract entities, but rather extracts entire relevant sub-phrases – called statements – and ranks them by perceived relevance to select which ones should be shown in the concept application UI. The ranking is based, in part, on lexico-syntactic pattern matching.

The second approach was finished and tested in another qualitative user study. This study focused on the efficacy of the algorithm at producing relevant content for the application UI.



## 2 Theoretical Background

I will now explain in more detail the two theoretical foundations of this project: the *Similarity Effect* and the *Hyperpersonal Perspective*. These two subsections draw on theoretical and experimental findings from the social sciences, with one concept coming from Social Psychology and the other from Computer-Mediated Communication (CMC).

In the third subsection I will introduce Reddit, the online platform used to source both participants and data for this project. With reference to recently published research, I show how uncommon it is for Reddit users to be persuaded in a discussion.

### 2.1 The Similarity Effect

*Interpersonal attraction* is a term used within social psychology to describe any kind of affinity for another person. It has been defined by Byrne (1961) as “*the direction and the strength of the affect engendered between the two participants in a dyad*”. Different kinds of relationships can be explained in terms of varying degrees of interpersonal attraction.

This attraction is determined by several different things, including physical and functional distance. Functional distance refers to the general likelihood that two people will come in contact with each other. These distance measures are commonly referred to as *propinquity* within social psychology. Humans are also attracted by a similarity of values or other psychological factors. Of course, physical attractiveness can in itself be one of the factors determining interpersonal attraction, particularly in romantic relationships.

A wide range of empirical studies have documented the link between interpersonal attraction and actual or perceived similarity of personality traits, attitudes, and hobbies. In fact, the evidence for this phenomenon – dubbed the “*similarity effect*” – is overwhelmingly strong within social psychology and the disagreement resides mainly with whether actual or perceived attraction is the contributing factor towards the attraction. In the meta-study performed by Montoya, Horton, and Kirchner (2008) the correlation between perceived similarity and interpersonal attraction in particular is shown to be quite strong.

There is also evidence pointing towards an established relationship between attraction, similarity and prosocial behaviour. Early studies into altruism support the notion that similarity in values leads to an increase in altruistic behaviour. Krebs (1975) showed that subjects could be induced into altruistic behaviour – sharing money and electrical shocks meant for another person – if they believed that the one they were sharing with had similar personal values.

Stürmer, Snyder, and Omoto (2005) take a group-level perspective and theorize, based on results of earlier studies and their own experiments, that interpersonal attraction towards outgroup members will modify typical ingroup-outgroup behaviour. Typically, ingroup members are less stereotyped and less discriminated against than outgroup members. However, interpersonal attraction modifies attitudes towards members of outgroups in such

a way that these members seem less prototypical, in effect becoming subject to fewer stereotypes and less discrimination than other outgroup members.

The experiments conducted by Stürmer, Snyder, and Omoto (2005) using sexual identity as the ingroup-outgroup separator provide support for this perspective. Interpersonal attraction is therefore shown to have a positive effect on prosocial behaviour towards the object of affection.

In early studies, attraction and similarity was shown to have a positive linear relationship (Byrne and Nelson, 1965, cited in Singh and Ho, 2000), although later studies have shown an asymmetric relationship between similarity and dissimilarity. Empirical evidence from multiple studies shows that dissimilarity has a stronger weight than similarity when it comes to social attraction, but that similarity and dissimilarity have equal weight when it comes to intellectual attraction (Singh and Ho, 2000). Intellectual attraction is defined as the evaluation of intelligence and knowledge, whereas social attraction concerns a liking for and enjoyment of the company of.

## 2.2 The Hyperpersonal Perspective

When communication moves online, the change in communication medium also affects the way people communicate. Much research in computer-mediated communication analyses the presence or absence of cues used to infer other people's characteristics. Some nonverbal behaviours present in face-to-face communication, such as facial expressions or tone of voice, are missing in online text-based communication. On the other hand, online text chat typically has added cues – e.g. in the form of emoticons, emojis or typing speed – that can not be found in face-to-face conversations.

Toma (2012) has researched online perceptions of trustworthiness taking into account these differences in cues. The research provided significant support for the Hyperpersonal Perspective of communication. The results of the research came from an experiment where participants had to evaluate online dating profiles.

The Hyperpersonal Perspective stands in opposition to Social Presence Theory and Information Richness Theory – the “cues-filtered-out” perspective – which postulate that the more cues provided by a certain medium, the more conducive this medium will be to socio-emotional exchange in the form of interpersonal trust and “liking” (Toma, 2012), i.e. interpersonal attraction. Accordingly, the lack of certain cues in text-based computer-mediated communication was thought to result in more impersonal communication with a lack of affection and emotion.

The Hyperpersonal Perspective was developed by Walther (1996) to account for examples of comparatively high affection and emotion seen on bulletin boards, chat rooms, but also in business settings. Walther based his perspective on Spears & Lee and their Social Identity-Deindividuation Theory (1992, cited in Walther 1996).

In contrast to the “cues-filtered-out” perspective, the Hyperpersonal Perspective states that under the condition of visual anonymity – a common phenomenon in online communication – people will tend towards “filling in the blanks”, so to speak. That is, information scarcity leads people to exaggerate existing known information about the other part, thereby forming either strongly positive or strongly negative impressions. This process is known as overattribution (Walther, 1996; Toma, 2012).

In addition to the effects of overattribution, the asynchronous and semi-anonymous nature of most online communication allows people to selectively self-present, further diverging from the personal impression that they would give in a face-to-face encounter (Walther, 1996).

In the experiments conducted by Toma, participants found the individuals described in the dating profiles to be most trustworthy if they had *“limited but positive information about them”*. The addition of a profile picture in fact reduced trustworthiness. Text-only communication, when compared to other kinds of communication mediums, received the highest impressions of trustworthiness. However, this is to be contrasted with a general finding in social psychology showing that humans are in fact quite poor at accurately assessing the trustworthiness of others.

## 2.3 Online Discussions

Online discussions are not only shaped by the dominance of text-only communication, but also by the rules and norms of different online communities.

### 2.3.1 Reddit

Reddit (reddit.com) is one such community. In fact, Reddit may be characterised as a sort of umbrella community consisting of thousands of individual communities – on Reddit these are known as a “subreddits” – each of which, in addition to following Reddit’s general rules and guidelines – known as the Reddiquette – also enforce their own rules to varying degrees (kemitche, 2012). This makes Reddit a place where a wide variety of social interaction takes place, although most of it still taking the form of thread-based, text-only communication between relatively anonymous users only identified by their screen names.

All content on reddit is submitted by the community’s users. These users may also write comments about the content or comments about each other’s comments. Moderation of discussions on Reddit are mostly based on self-policing using Reddit’s upvote/downvote buttons which are available for all types of submitted content and comments. An algorithm sorts the content and user comments based on cumulative vote scores and time passed in order to present a an organically changing frontpage of Reddit as a whole, of each subreddit, and of the comment sections of each submission. This self-organising process is meant to ensure that the most relevant content always floats to the top of the page. In addition to this, the volunteer moderators of each subreddit may moderate the discussion taking place within it according to its own self-prescribed rules.

### 2.3.2 Persuasion Dynamics

One of the more unique communities on Reddit is called “Change My View” ([reddit.com/r/changemyview](https://www.reddit.com/r/changemyview)) and features an extension to Reddit’s ordinary format in the form of a virtual currency awarded to other users based on how persuasive their argumentation is. The discussions taking place on the “Change My View” subreddit are also quite formalised, with each submission being a statement of a case and an invitation to change the user’s view and each reply being an attempt to do so by another user acting as a challenger. These facts made it a target of quantitative analysis by Tan et al. (2016) who wanted to explore persuasiveness in online discussions.

Tan et al. find that users who are ultimately persuaded in some way still form a small minority. Users are more likely to be persuaded by an early challenger in the discussion and also more likely to be persuaded if the challenger engages in some back and forth discussion with the user. The persuasion rate increases from just above 3% with one reply to somewhere above 5% with further discussion. However, after 5 or more replies, the persuasion rate drops all the way down to 0%.

Persuasion is more likely when attempted by a single challenger, rather than by multiple challengers in the same comment thread. However, persuasion becomes more likely with more challengers joining the discussion at large.

In addition to the group dynamics, stylistic choices also affect persuasion rates. This is one factor that the challenger can actually attempt to control. For example

- persuasive arguments are more lexically dissimilar to the original argument, than non-persuasive ones
- longer replies are more persuasive
- persuasive arguments begin by using calmer words

The low rates of persuasion on Change My View corroborate the idea that online discussions are often a waste of time, assuming of course that persuasion is one of the primary goals of a discussion. Of course, different individuals might have different goals with a discussion. The finding that using calmer words can increase persuasiveness corroborates the idea that civility and politeness can make discussions more pleasant and more purposeful.

## 2.4 Summary

In this section, I introduced the concept of interpersonal attraction, focusing on the similarity effect and its influence on how other humans are perceived. I also introduced the Hyperpersonal Perspective on mediated communication and documented how this theory can be used to explain some of the incivility seen online. Finally, I introduced Reddit as an example of an online discussion platform and demonstrated its applicability as a case study

in this domain by referencing recent work by Tan et al. (2016). Altogether, this provides a focused lens to view the subject matter with.

## 3 Related Work

Having explained the theoretical background of the thesis, I now proceed to the area of related work, before moving on to define the design of the the thesis concept.

This section will briefly introduce existing software tools that are somewhat related to the concept of this thesis. I will first introduce ToneCheck, a tool aimed at modifying the tone of emails prior to sending. Then I shall introduce two online tools that both generate profiles of Reddit users.

Note that I have placed technological literature reviews in the two separate Approach sections (6 and 7) since these tie directly into the design and implementation of each the separate approaches taken in this thesis.

### 3.1.1 ToneCheck

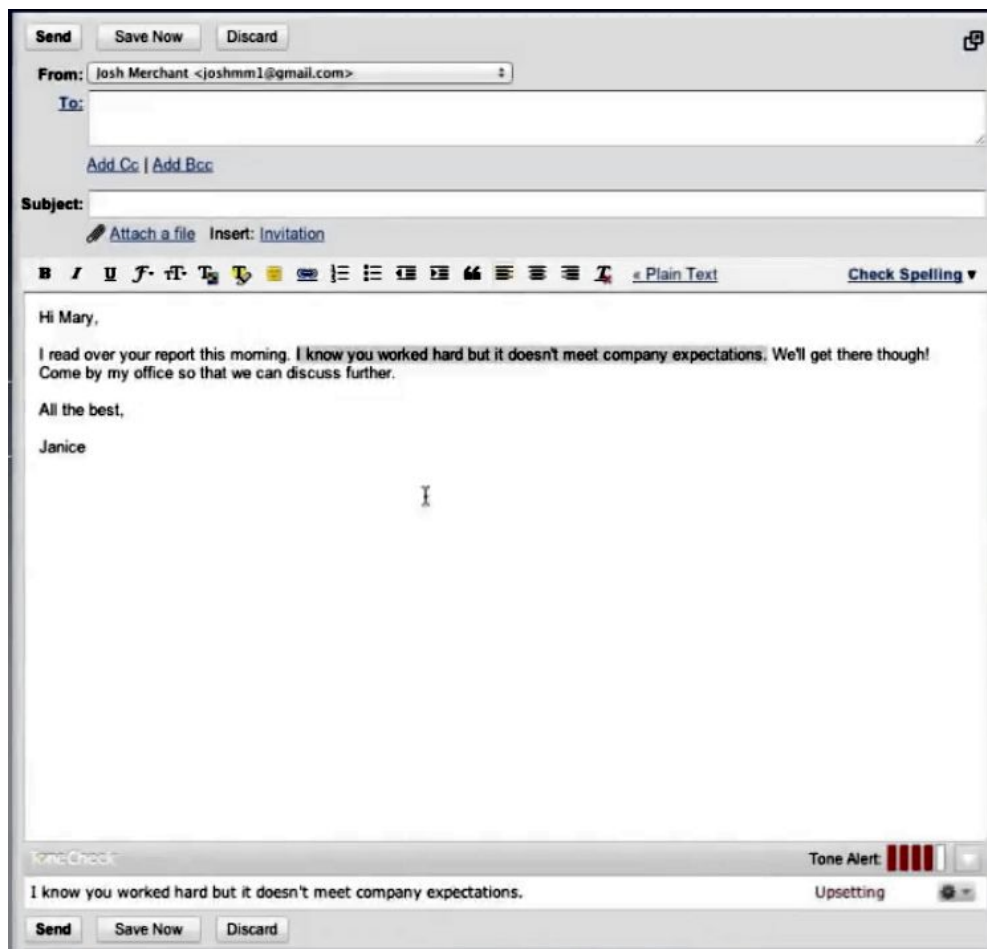


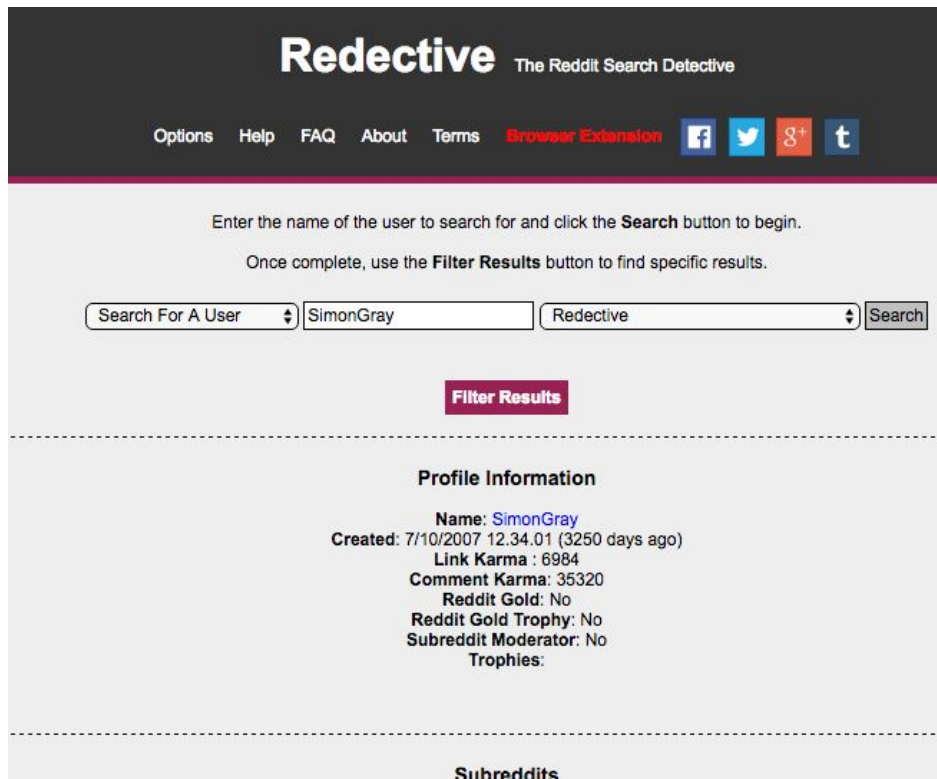
Illustration 3a:: ToneCheck for Gmail

ToneCheck (<http://tonecheck.com/>) is marketed as “Emotional SpellCheck for Email”. The purpose of the software is to find “*inappropriate sentences or phrases that are throwing the*

*tone of [the] message off*” (Arndt, 2012). The CEO of the company, Matt Eldridge was inspired to create the software by “*the increased frequency and lack-of-nuance in digital communications*”. The UI resembles a spell-checker that can find phrases with negative emotional content. It is then the task of the email author to review and possibly modify phrases with negative emotional content before sending the email.

### 3.1.2 Reddit Profiling Tools

I am aware of two separate tools that can generate profiles of Reddit users.



The screenshot shows the Redective website interface. At the top, the logo "Redective" is displayed with the tagline "The Reddit Search Detective". Below the logo is a navigation bar with links for "Options", "Help", "FAQ", "About", "Terms", and "Browser Extension", along with social media icons for Facebook, Twitter, Google+, and Tumblr. The main content area has a dark header with the title "Redective" and the tagline "The Reddit Search Detective". Below this, there is a search form with a dropdown menu labeled "Search For A User", a text input field containing "SimonGray", a dropdown menu labeled "Redective", and a "Search" button. Below the search form is a "Filter Results" button. The results section is titled "Profile Information" and displays the following data: Name: SimonGray, Created: 7/10/2007 12:34:01 (3250 days ago), Link Karma : 6984, Comment Karma: 35320, Reddit Gold: No, Reddit Gold Trophy: No, Subreddit Moderator: No, and Trophies: (empty). Below the profile information is a section titled "Subreddits".













Illustration 3b: Redective

The first tool is Redective (<http://redective.com/>) – The Reddit Search Detective – which performs basic statistical operations on a Reddit comment history: word frequencies are found, subreddits are ranked by activity (content and comment creation), and activity is presented by time of day. The profile is very basic and provides only a cursory glance at a Reddit user’s activity. There is no attempt to extract personal information or opinions.

## SYNOPSIS #

Accuracy or making sense not guaranteed. Results may be incorrect or misleading.

You can help by using the  and  feedback buttons below. Uncertain data is **in orange**. Follow # links for sources.

you are	<b>male</b>   #
you are in a relationship with your	<b>girlfriend</b>   # # # # # # # #
you live(d)	<b>in sinchon</b>   #
people in your family	<b>father</b>   #
things you've said you like	<b>music subscription services</b>   #
you are	<b>student</b>   #

*Illustration 3c: SnoopSnoo*

The second tool is SnoopSnoo (<http://snoopsnoo.com/>). This tool provides a much more detailed summary of a Reddit user, including personal information presented as tags in categories such as “you are”, “your locations of interest”, “your hobbies and interest”, “you like to play”, “you like to discuss”, and similar. SnoopSnoo also has subreddit recommendations for each user. In addition to that, SnoopSnoo features a comprehensive graphical presentation of the same kind of statistical information shown in Redective.



## 4 Design

The following section brings together the relevant social science theories and research findings from the [Theoretical Background](#) section, in order to motivate and inform the design of the thesis [Concept](#). I have defined a [Research Methodology](#) around this concept.

I chose to use Reddit as the target platform for reasons outlined in the section called [Using Reddit](#).

### 4.1 Motivation

The existing methods of moderation on online message boards cannot prevent fruitless, heated discussions full of animosity from taking place. It is therefore at the discretion of the individual participant to modify their behaviour in online discussions in order to prevent wasted time and rising stress levels.

In order to cultivate more civil discussion and fruitful dialog on online message boards, one must ask if it is possible to develop a method to foster a more prosocial approach to online discussions on the individual level. Such a method should tap into known psychological processes at play and exploit or change these processes in such a way that they produce the desired effects, in this case self-regulation of communication.

As is made evident by the research on the [Similarity Effect](#), sharing attitudes and values with another person produces a higher level of attraction for that individual than would otherwise be the case. This higher level of attraction is also linked to more prosocial and friendly behaviour by virtue of modifying the impression of an Internet stranger in such a way that he/she appears less stereotypical. Establishing a higher level of interpersonal attraction by creating a higher level of perceived similarity could therefore, in accordance with the research findings (Stürmer, Snyder, and Omoto, 2005; Krebs, 1975), lead to more friendly and civil discussions.

Under the Hyperpersonal Perspective on computer-mediated communication, the lack of many contextual cues in online text-based communication leads discussion participants to overinterpret and exaggerate the vices and virtues of other participants. Additional small cues could radically alter this – potentially hyperpersonal – perception of the other participant and consequently the level of interpersonal attraction towards that participant. If one has a negative impression of the other participant based on a single comment he/she has written, then this impression could be shifted by presenting relatively few pieces of information that provide a different impression.

As the research shows (Singh and Ho, 2000), the link between similarity and interpersonal attraction is likely to be asymmetric in such a way that dissimilarities have higher weights on the overall level of interpersonal attraction than do similarities. With individuals for whom there is an existing negative impression, it is therefore all the more necessary to counter

dissimilarities with multiple strong similarities, or at the very least neutral pieces of information.

## 4.2 Real-time Opinion Matching

One method of inducing self-regulation towards more civil discussions is to create an interface that integrates into the comment writing process on an online discussion platform. This integration can be accomplished by creating a web browser extension that modifies the pages of the message board in question. This modified interface would be supported by a backend system that is able to generate appropriate content for the interface at the appropriate time.

This interface should display similarities of attitudes and personal information between the user writing the comment and the user that the comment is addressed to. This is done in an attempt to exploit the hyperpersonal quality of impression-making on these discussion platforms. The main purpose of the interface is therefore to increase the perception of similarity and consequently interpersonal attraction, leading to self-regulation of potentially hostile replies towards more friendly and civil ones. A secondary aim is simply to present neutral information to act as a balance to negative information.

One of the techniques used for similarity studies in social psychology is the *phantom-other technique*, originally invented by Smith (1957, cited in Montoya, Horton, and Kirchner 2008) where participants are presented with a summary of the opinions and attitudes of another individual on a piece of paper. This type of profile is then shown to lead to higher evaluations in areas such as intelligence, knowledge of current events, morality, and general likability, when the information in the profile is similar to the participant's own opinions and attitudes. The interface described in this section is in effect a simple interactive design based on this type of psychological study.

## 4.3 Using Reddit

I have chosen to use Reddit as the target platform for this system. It is conceivable that the same system could be used with potentially any online discussion platform – and much care has been taken to make the results as generalisable as possible – but Reddit is a very good platform for testing just such a system:

1. Reddit has an open API for easy and legal access to the comment history of a user.
2. Reddit users are anonymous and do not have user profiles or profile pictures.
3. Being on a diverse discussion platform, an active user of Reddit will typically have large backlog of previous comments containing likes and dislikes of various topics as well as other personal information. Opinions and any relevant personal information can be extracted from this backlog of comments.

For these reasons, I decided to base the implementation of the system on data sets of Reddit comments, with a modified version of Reddit in mind as the application UI.



*Illustration 4a: What the modified Reddit interface might look like.  
In this mock-up, the contents of the comment have been omitted.*

In this modified version of Reddit, a generated summary profile is inserted into the area immediately between the comment and the text field used for replying to the comment. The summary profile takes up about as much space as the text field.

## 4.4 Research Methodology

This thesis has an open-ended focus on documenting the challenges of implementing a theoretical concept, which takes inspiration from social psychology and computer-mediated communication. Both the concept and any implementation of it will need to be evaluated with this theoretical background in mind. The psychological effects of the concept and any implementations of it can be theorised about, but no standard exists for measuring the actual effects quantitatively. For this reason, the methodology of the thesis is based primarily on qualitative methods, as these better accommodate this kind of open-ended, exploratory research. The two studies that I have conducted for this thesis are both based on gathering qualitative data through semi-structured interviews in an experimental setting. The first study was designed to establish the feasibility of the concept, while the second study was designed to evaluate the efficacy of an implemented system.

## 5 Initial Feasibility Study

This section outlines a study of the thesis concept. The research method has been designed based on section [4 Methodology](#). In this study, a mockup of the concept was presented to a group of participants to gauge their reactions. The book by MacKenzie (2012) was used to source good practices for HCI experiments.

The intention of this qualitative study was to establish the feasibility of the concept and gain valuable insights from the reactions of the users. The materials from the study are available in Appendix A.

### 5.1 Method

I investigated whether modifying reddit's commenting interface to add an info box would affect how participants interpreted controversial comments that I evaluated that they were likely to disagree with or disapprove of. I was interested in knowing whether the addition of a few lines of neutral-to-positive personal information would change the general perception of the author of the controversial comment. The intent was to guide the design of the system and evaluate whether the concept was feasible.

The object of the analysis are interviews made with the participants during the sessions. I have intentionally not considered the actual replies as material, as I have no prior knowledge of the typical tone and style of each participant when writing online. It is therefore not reasonable to compare the effect by looking at the comment itself as there is no basis for comparison. Instead, I have focused on the thoughts expressed by the participants about three different Reddit users, taking special note of any references to the influence of the concept, as well as the differences between the control group and the experimental group.

The participants were partitioned into two groups – a control group and an experimental group – for a between-subjects experiment. In this way, the task could be set up to fully discriminate the two test conditions, helping to avoid interference.

#### 5.1.1 Participants

10 participants were recruited through an announcement on the Denmark subreddit and an announcement on Facebook.

- 4 out of 10 of the participants were female and 6 out of 10 were male.
- They were aged from 21 to 31 years old.
- All of the participants indicated that they sometimes read discussions on the Internet.
- The majority - 8 out of 10 - found discussions to read on Facebook, while 6 out of 10 found discussions to read on Reddit. 5 out of 10 also found discussions elsewhere.

- 9 out of 10 participants indicated that they sometimes participate in discussions on the Internet; the same 9 participants indicated that these discussions are sometimes with strangers.
- 5 out of 10 of the participants participate in discussions specifically on Reddit.
- 9 out of 10 spoke Danish during the interview, 1 out of 10 spoke English.

### 5.1.2 Commenting Interface

The experiment differentiated participants by presenting two different simulated versions of the reddit commenting interface. This commenting interface represented the only independent variable of the experiment. For the control group the interface was an unmodified version of the standard reddit commenting interface. For the experimental group, the interface had been modified to include an info box immediately below the comment text.

For both versions of the interface, this was accomplished by using static HTML pages simulating real submissions on reddit.com. Each of the submissions only had a single comment on it and a comment field for inputting a reply to this comment. The intention behind doing it in a relatively representative way was to improve the external validity of the experimental results. A possible threat to the validity comes from the artificial setting of the study.

While the layout represented the current design of reddit.com (June 2016), the content of the page had been modified. The submission titles were real submission headlines from Reddit, but all comments had been removed except for a single one in each case and its score had been set to the default of 1. This was done to remove any indications of how the comment had been received by the community in its original context, as this might affect the perception of the participant. Each of the remaining comments were aggressive in tone and contained opinions that most Danish people were likely to be opposed to: denial of global warming, calling for a terror attack, and stating that religious people are more moral. The intention behind this was both to try to affect participants emotionally in the otherwise artificial setting, but also to make sure that there would be a high chance of disagreement.

For the modified version of the interface, an info box had been attached to the comment. This info box contained 3 bullet points of generally neutral-to-positive information about the user. An example of neutral information is the bullet point *“He likes to sing in the shower”* as it resembles a common everyday activity of many. An example of positive information is *“She donates blood at the Red Cross blood bank”* as it represents an activity that is well-received by many. Each comment had a different set of bullet points associated with it. Those participants who were presented with the modified interface were told that this information was not chosen by the author of the comment, but rather had been extracted and chosen by software. The bullet points in the info box were designed to be non-confrontational and something most Danish people would be likely to be unopposed to or even in favour of, for example: denial of global warming. The intention was to simulate the software concept as closely as possible.

ISIS 'will target' the Rio Olympics 2016 says Brazil's anti-terror chief (dailymail.co.uk)  
submitted 3 days ago by listentome5  
1 comment share save hide give gold report

[–] SimplyStranger 1 points 20 minutes ago

Fuck it, let it happen 2016 Rios Olympic.

The world is going nowhere with these political correctness, and we need another tragedy to evolve.

**Info about user:**

- SimplyStranger is in high school.
- He thinks our leaders are not doing enough about global warming.
- He likes to sing in the shower.

permalink embed save report give gold reply

*Illustration 4b: The title, comment, and info box for the above mock-up Reddit submission. The mock-up used for control group did not feature the yellow info box and its contents, but was otherwise the same.*

## 5.2 Procedure

Before performing the task, participants were asked to give consent and to fill out a short questionnaire. On the questionnaire it was made clear that the results of the experiment would be kept entirely anonymous and that the participant could choose not to perform a task or not to answer a question, if the participant did not want to do so. The questionnaire was designed to collect basic demographic information as well information about the participant's familiarity with online discussions. It is available in the appendix together with the other material from this study.

The participants for each group were brought in separately to fill out the questionnaire, perform the task, and answer follow-up questions.

The task was implemented in a representative way, by having the participant read 3 comments in a simulated version of reddit and asking them to write a reply for each one.

For the control group the unmodified commenting interface was used, while the other group was presented with the modified interface. There were 4 people in the control group and 6 people in the experimental group. As the interviews with the experimental group were more likely to provide good feedback for the implementation of the system, I did not balance the size of the two groups.

After writing the replies, the participant was queried in a semi-structured interview about their feelings about the comments and their replies. The group for whom the commenting interface had been modified were also asked questions pertaining to the info box. The interview guide for the experiment is available in the appendix.

## 5.3 Limitations

The participants were asked to write replies to comments that were intended to be aggressive in tone and opinion. Most reddit users use screen names and have a reasonable expectation of privacy when writing comments or replies to comments. In this experimental setting, there was no expectation of privacy and this may have affected the actions of the participants.

## 5.4 Results

Many participants across both groups expressed resignation at the thought of engaging in discussions online, regardless of whether they commonly engaged in discussions or not. A few participants chose to not write replies in some cases and several participants stated that they would not have written certain replies if they were not part of the experiment. Since the replies were not themselves the object of my analysis, this did not affect the findings. It does highlight the variance in engagement between different participants.

Only Participant 1 spoke English during the sessions, the rest of the participants spoke Danish. All quotes presented here are translated from Danish except those of Participant 1. The users that are evaluated by the participants are referred to as User 1, 2 and 3 in the order of the appearance of their comment (and associated info box) in the session. Participants are referred to as Participant 1 through 10 in the chronological order of their scheduled sessions. Participants 3, 5, 7, and 8 were in the control group, while Participants 1, 2, 4, 6, 9, and 10 were in the experimental group.

Interview quotes are presented with timestamps from the recordings. The same quotes can all be found in untranslated form in the Appendix to this thesis. The comments written by the three users, the associated information in the info box, and the replies of each participant are also available in the Appendix.

### 5.4.1 High Efficacy of “High School” Info

Participants with the info box were all seemingly affected by or took strong note of the information about User 2 – allegedly – being in high school: *“SimplyStranger is in high school”*. Participant 2 (17:30), Participant 4 (13:32), Participant 6 (15:42), and Participant 10 (33:20) all state directly or give strong implications that they are affected by the bullet point about User 2 being in high school.

For example, Participant 6 states (15:42): *“But yeah, you can’t not be affected by the fact that it says that he’s in high school, then you think he’s young”*. The participant continues at 16:35: *“I think that there’s no reason to get tough with younger people”*.

Participant 2 notes that she (19:53) *“thinks that if it hadn’t been there – the yellow box – then [she] think that [she] would have thought that he was different from how he was”* and that she instead *“is giving him a longer leash”* (20:11).

Participant 4 also took note of the high school fact. When asked in the interview, if this fact had changed how she had evaluated the comment, she replied (13:32): *“Most definitely. I think it’s different when it’s a young person as opposed to an adult”*. She comments on the same fact later in the interview (17:03; 17:33): *“Now that I know that he’s young I would probably ignore it. (...) If I didn’t know this I would probably reply”*.

Participant 10 offers an even more colourful description of User 2 (33:20): *“He’s probably in the back row in Classical Studies and is just sitting there writing ‘whatever man, kill the world, doesn’t matter we just need a hard reset’ and I in many ways agree with him here and I’ve incorporated that into my reply”*.

Participant 1 might also have been affected by the same factoid when the participant calls the user *“some little kid hating the world”* (9:03) although the participant later states (12:05): *“I actually didn’t even read the info box for this second one. I guess I did see this through passing.”*

Participant 9 *did* have the info box available in their session, but stated that she did not read the information in it when replying to comments (15:46). At 19:03, when talking about User 2, she again states that she did not read the info box, but also took note of the high school information in that same moment: *“I will honestly admit that I actually didn’t read that box at all, so it wasn’t even something that was under my consideration. I would probably – seeing that he’s in high school – possibly consider replying differently, but I think it would still be the same”*.

The results of the participants with the info box are noteworthy when comparing them to participants without the info box. As an example, Participant 3 did not have the info box and forms a less clear – and negative – impression compared to the participants who had the info box (14:19) simply calling the user: *“some american idiot”*.

Participant 5, who does not have the info box either, also forms a less clear impression of the user (20:29): *“one might immediately think that it is someone who is angry at the World, but I don’t think so”*. Participant 5 also mentions that he would typically not reply to a message like this (21:39).

Participant 8 also did not have the info box and formed a dismissive impression of the user. The participant did not even reply (17:18): *“There really isn’t anything to say. It’s obviously someone writing to provoke a reaction”*.

Participant 7, also without the info box, states at 28:48: *“It might be a right-winger, it might be a left-winger, it might be an extremist... it might be anyone who has lost something”*. However, after being introduced to the concept at the end of the session and while still remaining a bit sceptical towards the system’s applicability, Participant 7 remarked that he would also be affected by the high school info for User 2 (36:14): *“Yes, at least the top one saying that he’s in high school. Then you would think: ‘He’s just a young kid, so we need to teach him something’, right?”*.



The impressions made by the participants without the box are all – perhaps with the exception of Participant 5 – clearly negative impressions, while the participants who had the info box were arguably more neutral in their impressions of User 2 as well as being more willing to approach the user, building a strong case for the efficacy of this specific bit of information.

#### 5.4.2 Diverging Impressions of User 3

Participant 2 notes that the information in the info box changed their impression of User 3 (20:46): *“before reading the box, I tell myself what kind of person this is. Then I’ll have some quick prejudice. I’m kind of thinking that it’s some man who is 50+ or 40+, but then I notice that it’s a lady, she’s into urban gardening. And then it says that she loves animé, so I’m thinking she’s kind of young. ‘Donates blood’ so that means she’s not a Jehova’s Witness. Then I form a new impression, right? But I still think it’s incredibly stupid what she’s written.”*

Participant 6, when queried about whether the info box for User 3 made a difference, says the following (18:02): *“Not the part with Anime, but like ‘urban gardening’ then I’m thinking ‘oh, it’s someone living in the big city’ and I also think, with her donating blood, that she must... she’s someone who’s willing to do something to help others, right? She is someone trying to be a good person”,* later adding (18:25): *“yeah, it’s probably also something that makes you think she’s someone you can talk to, right?”*.

Participant 4 did not think the information for User 3 was relevant at all (15:02).

Participant 10 generally formed strong impressions of all three users based on the info boxes, but with User 3 this impression did not result in a notably neutral or positive view of the user, despite the participant also partaking in the some of the same activities as User 3. The participant states (36:29): *“I also watch animé and donate blood, but as soon as I look at this, I’m seeing a person who’s just polishing her halo on a daily basis and who comes riding in with a cross in their hand and is just some kind of righteous knight”*.

Like with User 2, several participants without the info box had arguably less clear and/or more negative impressions of User 3. Participant 5 said (22:20): *“I initially thought [that it was] some Southern State redneck’ish one, not particularly intelligent, because nothing is said other than – this is taken out of context, but – ‘God is great and right, halleluja’ and other points of view have not been considered at all.”*

#### 5.4.3 Different Perceptions of Relevance

Participant 10 was a daily reddit user with a large comment backlog and, of all the participants, wrote the longest replies and also had the most elaborate descriptions of the different users in the follow-up interview. In all 3 examples, the participant incorporated the information of the bullet points into his replies.

For example, at 28:08 Participant 10 explains how he incorporated the information that User 1 was a chess player into his reply and therefore should respond better to analytical thinking:

*"It's a totally different way to read this, because if I read it as if the extra info isn't there, then I would have seen... then I don't think I would have given him the reply that I've given him here. Then I wouldn't have been able to justify using my time to come up with with arguments [against] this, but because he's a chess player then I assume that he is more analytical".* Participant 10 later elaborates (29:12): *"if this was a real situation, then I don't think I would have entered a discussion here and replied, if I hadn't seen this information".*

Participant 10 continues describing User 1, incorporating more of the facts from the info box (30:38): *"I'm guessing that he's in the kindergarten, but still has a hobby which is doing... you know, reading more scientific articles, only based on chess and on the fact that he's on a forum".*

To Participant 10, the information that User 1 is chess player was highly relevant. In fact, the participant even based his reply around it.

To Participant 4, however, the same info box did not contain any sort of relevant information (12:44): *"With global warming one might have been interested in knowing whether he has some kind of agenda. This is just... 'likes playing chess', that's not so relevant here".* The same participant notably also did not find the information for User 3 relevant either (15:02).

#### 5.4.4 Difficulty Understanding the First Comment

Many participants had trouble understanding the first comment: Participant 2 (16:29), Participant 3 (11:02), Participant 4 (11:52), and Participant 5 (18:48). Many expressed confusion as to what the user's position was on global warming and several users chose not to answer: Participants 2, 3, and 6 all did not reply to the first comment.

Participant 6 did not seem to have trouble understanding the message, but chose not to answer for the following reason (13:43): *"It was the way he wrote. And how he seemed a bit angry, a bit aggressive in his rhetoric, in which case I've had previous experience where if you write to people like that, then it's not going to be a nice experience anyway, so I didn't".*

The participant also mentioned that she briefly considered using a fact in the info box against the user (14:02): *"The only thing I could ever wish to end up writing is: are you bitter after your divorce? And then I thought 'no, I'm not going to sit here and attack people over something like that'."*

#### 5.4.5 Other Results

Some participants had mixed feelings about the concept or the concept's applicability to their own situation. Participant 1 explains (12:51): *"People use reddit so they can be anonymous".* The participant continues (13:16; 13:27): *"On facebook when people write stuff, I make sure to see who wrote it. On reddit, I've never actually clicked on a user."* Participant 7 stated that they liked the idea, but that they were not the target group (37:56).

Participant 6 thinks the concept is effective (20:28), but that she would rather be without it for the following reason (20:45): *"I think it would harder to keep the eyes on the ball instead of*

*the man, meant in both a good and a bad way, because it might be that it says ‘oh, he’s young and he needs to get a chance to hear some other opinions’, but it also might go into the other direction: ‘Oh, but you’re just some idiot Republican anyway’.*”

Participant 5 compared the concept to face-to-face conversations (27:39): *“It’s the same concept that we use out in the real world. You typically speak far nicer to people you know.”*

Participant 10 made an interesting remark about one of the factoids written about User 2 (34:06): *“He likes to sing in the shower’ - that doesn’t say much about a person, we all do that if we have the right song on our mind”*. In this comment, he dismisses the information presented in the factoid, while simultaneously demonstrating the relatability that it was intended to express.

### 5.4.6 Summary

Summing up, all of the six participants in the experimental group took note of the information about User 2 being in high school and 4 out of 6 had notably changed impressions. The two participants in the experimental group that did not have notably changed impressions also did not pay much attention to the info box, indicating some possible UI design issues.

It is apparent from several of the interviews that there is a variance not only in what types of information affect the different participants, but also in how it affects them, e.g. Participant 4’s reaction was to disengage knowing that User 2 is in high school.

## 5.5 Discussion

The high efficacy demonstrated by the high school fact for User 2 provides some clues as to what might be good content for the info box. While the obviously positive facts – such as “donates blood” – did not result in many strongly modified impressions – apart from Participant 6 – this more neutral fact had a strong effect on all participants it was shown to and their subsequent impressions of the user were markedly similar. The participants felt a need to moderate their language and approach the user differently – seemingly more friendly – than would otherwise have been the case.

Conversely, the impressions of the participants without the info box were less clear and more negative. This could be taken as an example of the Hyperpersonal perspective in action, the theory introduced earlier in this thesis, where a basic negative impression is turned into a more neutral impression in the light of not only positive, but also just strong neutral information. This information can be used to broaden and soften the hyperpersonalised negative impression given by the comment on its own by reducing overattribution.

Strong facts like this – e.g. a user’s age – can be indicated by other bits of information, such as occupation. Having occupation or some other signifier of age and location in the info box might then have a reasonably strong effect. An implementation should therefore weigh this type of information higher. Of course, this example was informing participants that the user

was younger than them, so it is possible that any indications of other ages might have no efficacy or would perhaps even give a more negative impression – or it might be a general phenomenon and would also apply to e.g. an old lady. There is not enough evidence to make a strong conclusion in this regard, but it can be used to build a hypothesis that can be tested when a prototype of the system has been implemented.

The fact that Participant 10 can respond strongly to certain bits of information and other participants do not react at all – or even lament the fact that the information is irrelevant to them, such as what Participant 4 expressed – can perhaps be explained by the similarity effect from social psychology, also introduced in the [Theoretical Background](#) section. This well-established phenomenon predicts that shared attitudes, shared hobbies, and other kinds of similarities often lead to more positive impressions of other people, all other things equal.

The fact that these similarities can affect certain people more strongly is an indication that the system implementing this concept, should weigh information according to relevance towards the reader. If the author of the comment and the reader of the comment have a similar relationship to some objects or activities, then these objects and activities should be weighed higher. The counterpoint to this idea is represented by Participant 10 too: the participant didn't form a neutral or positive impression of User 3, but did form a very strong and mainly negative impression, despite the info box mentioning two activities that they allegedly share: watching animé and donating blood.

Participant 6 also builds a case for this, when she talks about the dangers of the summary profile showing information that the reader would disagree with based on certain things, e.g. if the other is a Republican in Participant 6's case. Adjusting by relevance is also a case of being careful not to show certain strong dissimilarities that would worsen the existing impression. To Participant 6, knowing that the other user is a Republican, could potentially create a worse impression than was the case without the info box.

Of course, this is the part where the ethics of the concept are most challenged. Is it ethical to reveal and conceal facts in this way? Some participants would clearly have issues with the knowledge that the information has been adjusted to fit their own worldview better. In this case, I would use the Hyperpersonal Perspective to argue, that an impression that hinges solely on a single negative comment is arguably more biased than an impression that is based on a comment as well as a set of facts that have been adjusted by relevance to the reader, even if this adjustment is itself introducing bias. We have to make the assumption that strangers are decent people most of the time for this concept to work. If the basic mindset of a person is misanthropic and if the inherent bias explained by the Hyperpersonal Perspective is denied by this person, then the concept will seem both unethical and not very useful.

It is interesting to note that the actual effect on the different participants also differs. While, for example, some participants with the info box became seemingly more engaged in the discussion through the knowledge provided by the info box – e.g. Participant 10 – Participant 4 said that he would use such information to avoid starting a discussion in the case of User

1. While the exact effect of the concept is not strongly documented in this study, this is still an important point. It shows that the effect of adding additional context cannot always be said to inspire any engagement. It could have the opposite effect, as was noted by Participant 4, and that should indeed happen in cases where any civil discussion is impossible. For other participants, the system could perhaps be an enabler of civil discussion.

The points made by the participants were diverse and interesting. However, being the researcher for this project, I am at the discretion of interpreting the general theme of their comments and actions. It is my impression that the participants were indeed affected by the presence of the info box.

Even in cases where they dismissed the content outright – e.g. Participant 10 dismissing the fact about User 2 singing in the shower – I would argue that there is a possibility that the factoid has some influence on the reader. The facts do not necessarily need to put the user in a more positive light, they simply need to show a few indications of an alternative identity from that which was indicated by the comment alone. For some participants, knowing that the user is in high school was enough to change their overall impression. For other participants it was about the user being a fellow chess player or being into animé and urban gardening. Some people will not be swayed by either of these facts and this will require adjusting the profiles to be maximally relevant to every reader.

### 5.5.1 Concluding Remark

The study has shown that targeted exposure to certain types of information about a comment author is enough to change that author's impression in the eyes of the participants. In several cases in the study, an initial impression changes from a more negative impression towards a more neutral impression. The type of information that was most effective in the study was not stereotypically positive facts, but neutral personal information such "He is in high school". I hypothesize, based on the Hyperpersonal Perspective, that this event occurs because the increase in available information about the user, as provided by the info box, helps reduce overattribution. The reduction in overattribution leads to a more neutral impression of the user.

There was a noticeable difference between the impressions held by different participants; some participants took special note of certain kinds of information, while other participants dismissed the same factoids. While this effect was not as strongly evident as the reduction in overattribution, I hypothesize, based on the interviews and on the methodology of this study, that interpersonal attraction can be used, in part, to explain this difference. Specifically, interpersonal attraction based on functional distance and similar hobbies. There were several cases of participants who shared an activity or a hobby with a comment author, who noted that these shared activities contributed towards changing their impression of the author to be less negative.

An implementation of the software should therefore not be concerned with whether the facts presented in the info box put the author in a more positive light. The info box content can be

neutral information about the author. Additionally, preference should be given to information based on overlap with the information known about the reader. To sum up, the info box should present the information about the comment author that has the most overlap with the information about the reader in order to obtain good results.

## 6 First Approach: Sentiment Targets

Sections 6 and [7](#) document the two separate approaches I took in trying to answer the research question: “*What are the challenges in designing and implementing a tool that generates and displays information with the intention of inducing more civil discussions?*”. The goal of both approaches was to implement software that could be used to generate content for the concept described in section [1.1](#).

In this first approach, I was not concerned with trying to extract opinions or other information directly from the comments, but rather to assign a sentiment polarity to any named entities that could be found in a comment. I intended to use these “sentiment targets” to construct bullet points of information for the application UI. The intention was to prefer entities that overlapped in sentiment between two users over entities that did not. This was done in order to generate relevant sentences such as “*This user **also** likes Germany*” or “*This user doesn’t like Baseball **either***” to put in the info box (note the emphasis).

This section is about the first approach. I was not satisfied with the results of the system for reasons outlined later in this section, so I decided not to pursue it further.

### 6.1 Literature Review

In this subsection I present research from the field of Sentiment Analysis. This field was the basis of my first approach to implementing the thesis concept.

#### 6.1.1 Sentiment Analysis

Sentiment analysis is the “field of study that analyzes people’s opinions, sentiments, appraisals, attitudes, and emotions toward entities and their attributes expressed in written text” (Liu, 2015). Sentiment analysis is also known as opinion mining and both terms first appeared in 2003.

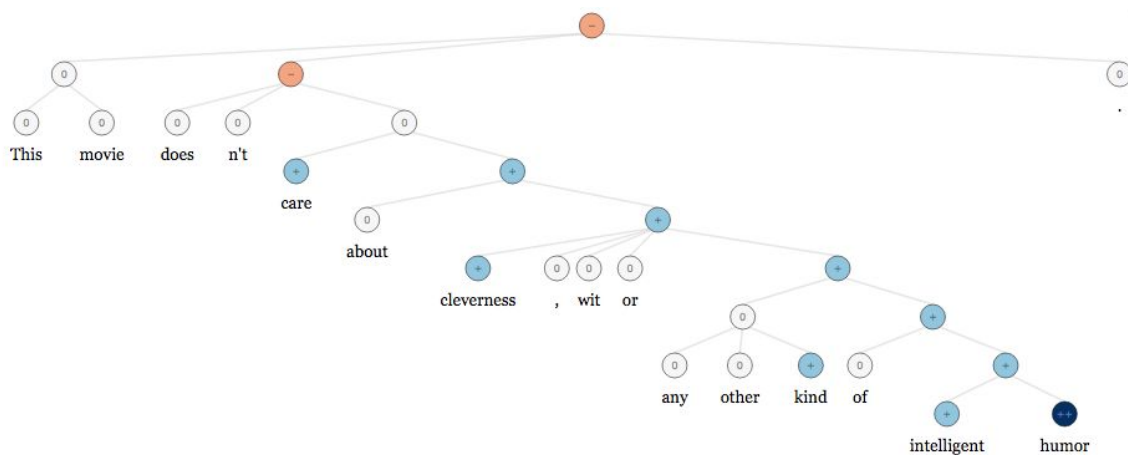
Sentiment analysis is focused on detecting sentiment at different levels: the document level, the sentence level, and the aspect level. Much research goes into document-level sentiment analysis, for example by building tools to assign a sentiment polarity – positive or negative – to a piece of text such as a review or a tweet.

Sentiment analysis is popular for predictions, for example stock market predictions, election outcome predictions, or predicting product success in the general population. It is also used to classify documents such as reviews into sentiment categories (Feldman, 2013; Liu, 2015). Extracting precise semantic meaning is less common in the prior research since typical applications using sentiment analysis do not require this level of fine-grained opinion extraction.

In academia, polarity detection is most often performed using statistical methods based on machine learning. These approaches typically train a classifier on some annotated,

opinionated text such as reviews, which makes them useful for detecting positive or negative valence at the document- or paragraph-level.

Unfortunately, classifiers do not work well with smaller inputs such as sentiment expressed at the sentence-level. These approaches also typically do not handle semantic information well and cannot extract sentiment targets from the text (Cambria et al., 2015). Indeed, one must assume that the text only contains a single target entity and a single accompanying sentiment (Feldman, 2013). Another issue is that statistical approaches are domain-dependent. They can be biased from the specific sentiments expressed in the training data, e.g. a political bias based on the general political opinion in the data (Novielli, Calefato, and Lanubile, 2015).



*Illustration 6a: An example of Socher et al's sentiment analysis system.  
Each node represents the sentiment of the subphrase spelled out by its child nodes.  
Blue is positive sentiment and red is negative sentiment.*

In 2013, Socher et al. lifted the state-of-the-art significantly through the use of *Recursive Neural Tensor Networks*, using sentence parse trees as input to a training algorithm. Rather than limiting the object of analysis to sentences, this approach allowed for sub-phrasal sentiment analysis into a 5-value polarity spectrum going from “very negative” to “very positive”. This approach increased the accuracy of sentiment polarity predictions considerably, however like other machine-learning approaches, it is domain-bound.

Other approaches mentioned by Cambria et al. (2015) include keyword spotting, lexical affinity, and concept-level polarity detection. These are called knowledge-based approaches by Cambria and Hussain (2015).

Keyword spotting simply assigns a polarity based on the presence of sentiment-carrying keywords such as “happy” or “sad”, while lexical affinity assigns sentiment probabilities from a lexicon of words or concepts and can be considered a more evolved version of keyword spotting; they are both lexical approaches. Words carry different sentiment in different contexts or the sentiment expressed can be flipped using various grammar patterns, which can make lexical approaches imprecise if they do not take syntactic patterns into account (Novielli, Calefato, and Lanubile, 2015). Lexical approaches require a sentiment lexicon



which can either be a domain-independent one or a domain-specific one created from a corpora (Feldman, 2013).

These alternative approaches might be also grouped together as rule-based approaches since they emphasise using rules to assign sentiment rather than relying solely on machine learning. Chiticariu, Li, and Reiss (2013) argue that there is a disconnect between industry and academia when it comes to information extraction methods. Machine learning is by far the preferred method for information extraction tasks in academic papers. The rule-based approaches are both fewer in number and the use of rules is obfuscated in favour of machine learning in papers using mixed approaches.

In industry, however, rule-based approaches dominate. Chiticariu, Li, and Reiss speculate that rule-based approaches might be more accepted in industry since they are “*easier to adopt, understand, debug, and maintain in the face of changing requirements*”. Furthermore, they argue that there is a strong tendency to ignore the time-consuming tasks involved in machine learning approaches, emphasising instead the labor costs of creating rules. On the other hand, rule-based systems are seen as a dead-end in academia, while machine learning is seen as the technical frontier, according to Chiticariu, Li, and Reiss.

### 6.1.2 Sentiment Towards Entities

Sentiment Analysis is in many ways an unsolved problem. The most fine-grained sentiment analysis algorithms typically output sentiment at the sentence-level. Very little work has been done in the area of extracting sentiment expressed towards entities (objects or persons) mentioned in these sentences.

Most of the related work in this area focuses on two core types of applications: extracting information from reviews and documenting trends on social media. In either case, the entity is known in advance. Mining opinions towards unknown entities in a heterogeneous dataset is a task that has not been attempted by many researchers.

One attempt at mining entity sentiment from heterogeneous sources was done by Biyani, Caragea, and Bhamidipati (2015). Their data source for this paper was Yahoo News comments. They note in their paper that “*despite the evidence of strong value in analyzing the sentiment of users tied to specific entities, there have not been any reported works on this problem.*” and continue by stating that “[t]he problem of identifying the sentiment polarity of these comments remains inherently difficult due to several main challenges, including irrelevant entities and implicit sentiment.”

Their approach is one of combining entity extraction with sentence-level sentiment analysis. They use the Stanford Named Entity Recogniser (SNER) to extract named entities from phrases and train a classifier based on a variety of features, including syntactical features, to predict sentiment. Named entities typically include persons, organisations, and places, although the exact types of entities that are found using different algorithms – or in the case of the SNER, with how the training data has been labeled – vary a great deal.

Biyani, Caragea, and Bhamidipati (2015) apply a heuristic to associate a sentiment context with each relevant entity. A sentiment context is defined as the range of words around a named entity. In their algorithm, this context is decided using a heuristic. For example, a clause initiated by the word “but” is a piece of context for the entity inside the clause, as are the two parts of a sentence split around a comparative term. For all other sentences with multiple entities, the context is simply the (up to) 3 nearest words in either direction from the entity. For single-entity sentences, the entire sentence is used as context. The researchers also apply simple rules for anaphora resolution between sentences.

Anaphora resolution refers to finding the antecedent entity of some pronoun, e.g. “she” might be an anaphor of the previously introduced antecedent entity, “Hillary Clinton”, so that any subsequent uses of the word “she” in a piece of text refers to “Hillary Clinton”. Anaphora resolution is then the process of linking or replacing an anaphor with its antecedent.

Next, the researchers apply syntactic heuristic to determine which of these sentence contexts express a sentiment and which do not, e.g. *“an entity of the type person is more likely to be polar than an entity that is of non-person type”*.

Finally, they train a classifier to predict the polarity of each entity using a variety of features including syntactic features, occurrence of sentiment words from a subjectivity lexicon, and sentiment polarity scores of the words in the context provided by SentiStrength, a lexical sentiment analysis system.

## 6.2 Sentiment Target Algorithm

The following section describes the algorithm implemented for the first approach. The algorithm’s design is based on the algorithm proposed and implemented by Biyani, Caragea, and Bhamidipati (2015) for finding entity sentiment in Yahoo news comments.

It differentiates itself from the former approach by relying more implicitly on grammatical structure and sentiment analysis results as provided by the Socher et al annotator, rather than using a set of heuristics to assign context for each entity. It also applies the sentiment analysis results of the contexts directly to the associated entities, rather than using it as input data to train a classifier.

To be clear, the purpose of this approach was to develop an algorithm that could extract a set of entities from a user’s comment history. Each entity would have a polarity score to be able differentiate negative, neutral, and positive sentiment. These entities could then be compared to another user’s entities and the matching entities with matching polarity between two users were then to be used to construct the summary profiles.

### 6.2.1 The Framework of Choice

The Socher et al (2013) paper moved the state-of-the-art of sentiment analysis significantly when the paper was published by providing better results than other algorithms on the same dataset and by allowing for Sentiment Analysis down to the subphrase level. In addition to

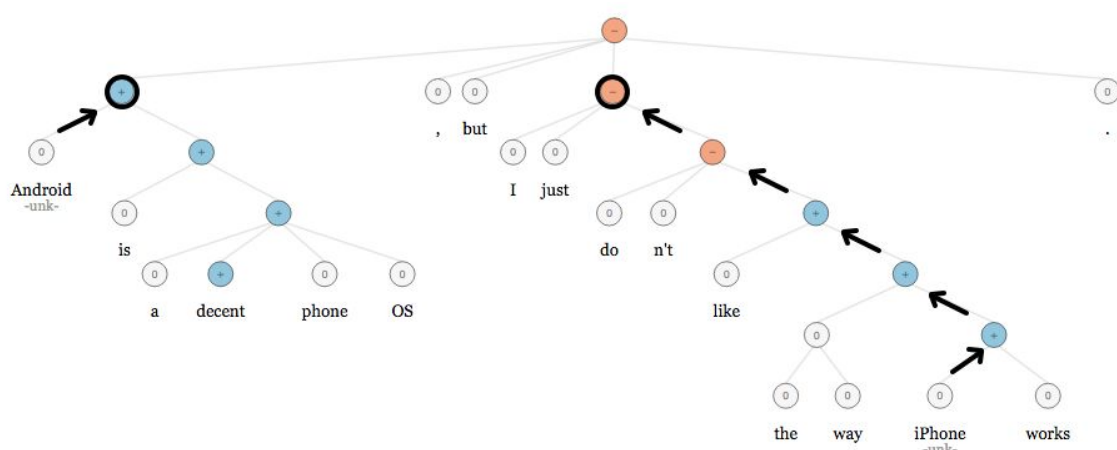
that, Stanford University has open-sourced the code (Manning et al., 2014) implemented by Socher et al. (2013) for sentiment analysis as a part of their CoreNLP package. The university also supplies a pre-trained classifier for plug-and-play sentiment analysis at the sentence level. Based on the paper and the immediate availability of the source code, I decided to base my approach around CoreNLP. Specifically, around the Sentiment Analysis and Named Entity Recognizer annotators.

The more interesting aspect of Socher et al's implementation is that it assigns a polarity to every part of the phrase-structure parse tree of a sentence, not just the sentence itself. This means that all possible grammatically independent sub-phrases are analysed and scored separately from the sentence root. Assuming that the fine-grained results are mostly accurate, this type of sentiment analysis should lend itself well to an algorithm inspired by the one created by Biyani, Caragea, and Bhamidipati (2015). For this reason, I chose to use the CoreNLP sentiment analysis annotators results directly instead of training a classifier.

While it is possible to retrain the sentiment analysis annotator for higher domain-specific accuracy, using e.g. Reddit comments as the dataset, this would require spending considerable resources on compiling, not to mention annotating, a dataset by hand. I did not have the resources to build a comprehensive dataset so I used the included model that had been trained on a data set of film reviews, one of the standard data sets for sentiment analysis in academia. Unfortunately, the domain-specificity of the included model is likely a major cause of the inaccuracy I found with my prototype.

Due to ease of implementation I also chose to use the Named Entity Recognizer (Finkel, Grenager, and Manning, 2005) included with CoreNLP for finding named entities. This is the same system used by Biyani, Caragea, and Bhamidipati (2015), so in this case there is no differentiation.

## 6.2.2 Navigating the Parse Tree



*Illustration 6b: determining entity sentiment by navigating the parse tree. In this example, the named entity Android is assigned a positive sentiment, while iPhone is assigned a negative sentiment.*

The original algorithm used heuristics to assign context to each entity. This is the part where my approach differs the most from the original. As in the original approach, single-entity

sentences are used in their entirety to provide context for the entity, but in all other cases context is found by navigating the grammatical parse tree produced by CoreNLP.

The phrase-structure grammar of the parse tree is the classical way to structure natural language. Originally coined by Noam Chomsky, phrase-structure grammar divides a sentence into a binary tree of subject and predicate nodes. This same tree is used to define sub-phrases within the sentence and each of these sub-phrases have been assigned a sentiment polarity score by CoreNLP. The implication is that any syntactically connected sub-phrase has its own sentiment polarity score.

Assuming that these sentiment polarities are indeed correct, then the heuristics used in the original approach to determine context (Biyani, Caragea, and Bhamidipati, 2015) could be replaced with a more general approach that simply navigates along the path from the entity node to the root node of the entire tree. If two or more paths intersect, then the node immediately before the node where they intersect is used as the root node of the specific context for that entity. Since each node in the tree has already been annotated with a sentiment polarity score by the CoreNLP Sentiment Analysis annotator, the score at this node can then be used as the sentiment of the entity.

Assigning sentiment to an entity is then just a case of navigating the parse tree to find the highest level node in the un-intersected path to the root node. The subtrees that are used as context for entities are not grammatically meaningless, since subtrees are usually guaranteed to consist of a subject and a predicate.

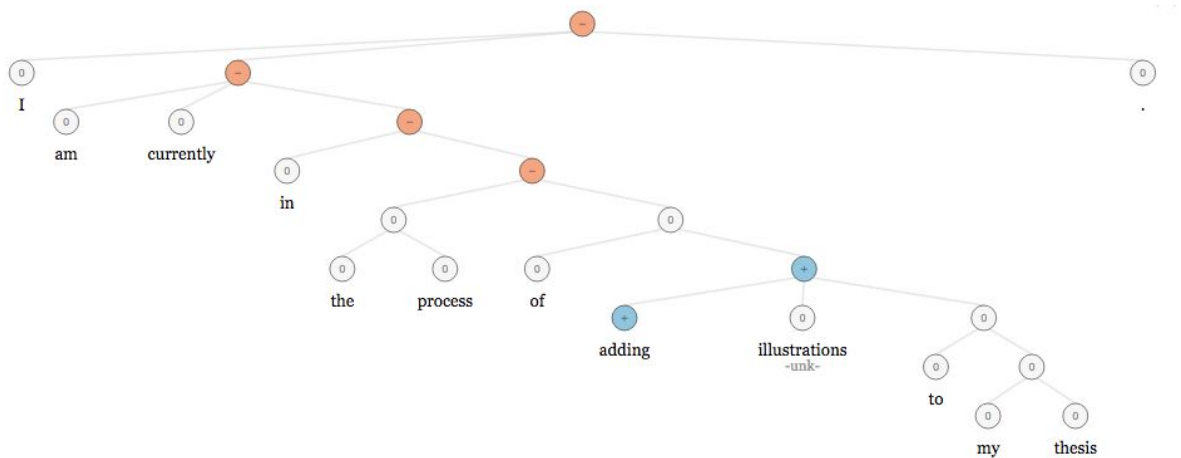
A simple anaphora resolution algorithm was also implemented to connect pronouns with their antecedent(s) in the previous sentence. This feature is similar to the anaphora resolution implemented in the original (Biyani, Caragea, and Bhamidipati, 2015), but also uses gender information for increased accuracy through the gender annotator of CoreNLP (Stanford University, 2015). CoreNLP features two anaphora/coreference resolution annotators – coref and dcoref – however, they were too slow and resource-intensive for practical use in this application.

## 6.3 Evaluation of Approach

As mentioned earlier, I abandoned this approach. I did this for two reasons:

1. the set of entities found was not very numerous
2. the sentiment analysis results were not very accurate, possibly due to domain dependence

To be more specific, the sentiment analysis results at the sentence level greatly overestimated the level of negativity in the comments of the test data set, most of which were neutral in tone. This is based on evaluating the comment histories of a few reddit users. I suspect that this is due to the model being trained on a data set based on film reviews, which might be too domain-dependent.



*Illustration 6c: an example of a neutral sentence that is incorrectly annotated as expressing negative sentiment by the Socher et al system (red = negative).*

The named entities were generally correctly tagged, however using only named entities and not any other relevant entities, was too limiting. The purpose of extracting entities was to compare entities with other users, and having a very limited set of entities available for each user reduced the chance of overlap by too much.

While the approach of navigating the parse tree to find context and sentiment polarity scores might be well-merited, it will not produce good results if the sentiment analysis it builds upon is inaccurate. Furthermore, with a very small set of entities, entity overlap was simply too uncommon to be a practical way to select entities to generate content from. This also makes it hard to even evaluate the algorithm in a research context. For these reasons, I chose to explore another approach to generate content for the application.

## 7 Second Approach: Statement Extraction

Like the [First](#) implementation, this approach builds on Stanford's CoreNLP for ease and speed of implementation. However, this implementation mainly uses the features of the dependency parser and the Part-of-Speech tagger, not the named entity tagger or the sentiment analysis annotator. It further differs from the first implementation by outputting actual sentences from comments written by the author, rather than generating sentences for the info box based on overlapping entities and sentiment polarities, as was the case with the first approach.

### 7.1 Literature Review

There are two major techniques in use to interpret and parse natural language. The most traditional technique is by way of phrase-structure grammar, where a sentence is recursively divided into a binary tree consisting of subject and predicate nodes. This way of parsing a sentence was, for example, used by Socher et al (2013) to derive all sub-phrases of a sentence, so that they could be annotated with sentiment polarity scores.

The other major technique used to interpret natural language is dependency grammar. I will now go over the ideas behind dependency grammar, as this second approach builds heavily on the dependency graphs produced with a dependency parser.

#### 7.1.1 Dependency Grammar

According to Debusmann (2000, p. 2), the intuition behind dependency grammar is the following: *"In a sentence, all but one word depend on other words. The one word that does not depend on any other is called the root of the sentence"*. In practice, verbs are often both the root component of a sentence and the governing component of any sub-phrases within a sentence.

Debusmann (2000, p. 3) continues: *"Dependencies are motivated by grammatical function, i.e. both syntactically and semantically. A word depends on another either if it is a complement or a modifier of the latter"*. In dependency grammar, meaningful phrases form tightly interconnected units in a sentence based on dependency relations. These relations can be represented as directed edges in the so-called dependency graph of a sentence, with the words of the sentence forming the vertices of the graph and their grammatical relations forming the edges.

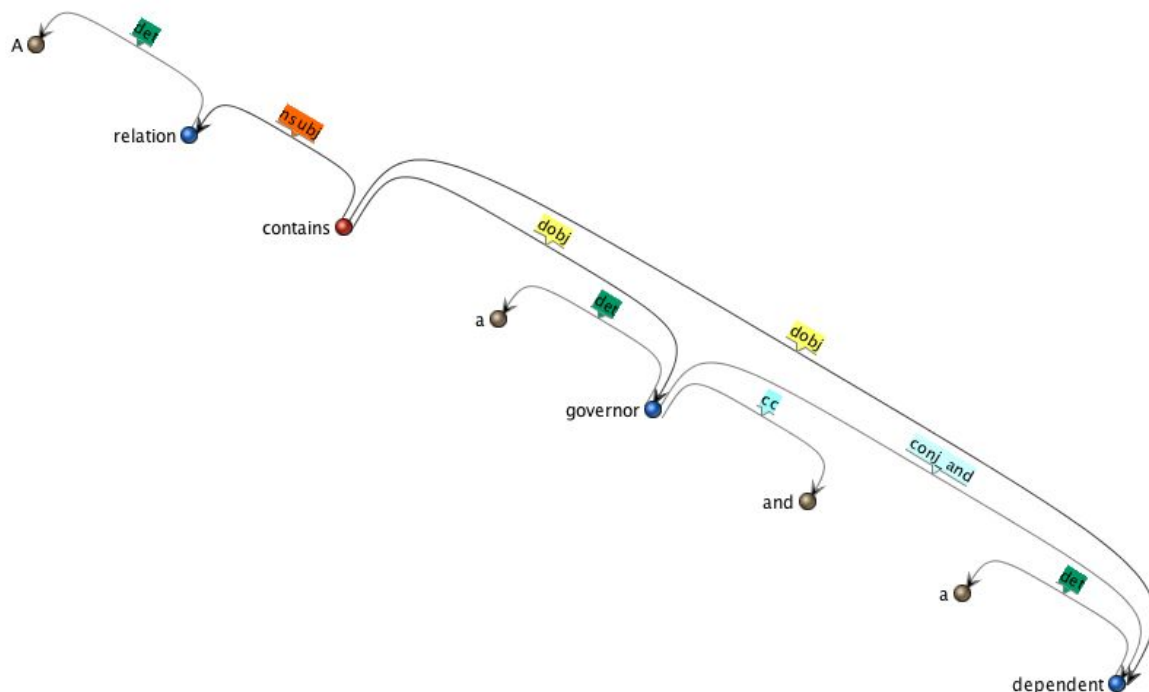
Other components, such as subject(s) and object(s), are typically "dependents" of the "governor" verb or adjective. Less important words, such as determiners or modifying adjectives, are dependents of these "dependent" components and so on, until all of the words in a sentence form a scope of inherited dependencies around the "governors".



*description of the grammatical relationships in a sentence that can easily be understood and effectively used by people without linguistic expertise who want to extract textual relations".* There are approximately 50 different types of grammatical relations in use.

Stanford Dependencies have since been superseded by Universal Dependencies, an open standard that closely resembles Stanford Dependencies. This standard is also used by other universities and organisations that are researching dependency grammar, e.g. Google's SyntaxNet (Google, 2016b).

The dependencies describe relations between two words. One word is always the "dependent" and the other word is the "governor". In a directed graph, each dependency is represented by a directed edge from the governor to the dependent; for example, the *nsubj* relation is from a *verb* to a *noun subject* or from an *adjective* to a *noun subject*. This relation, along with other relations such as *csubj* or *nsubjpass*, can be used to find the grammatical subject of a phrase.



*Illustration 7b: another dependency graph of a sentence*

Above I have included another dependency graph of the sentence "A relation contains a governor and a dependent". The root node is the verb "contains" which has outgoing relations to a grammatical subject – the noun "relation" – and to two direct objects: the nouns "governor" and "dependent". In the same graph it is apparent that the subject and objects of the phrase have their own separate dependents and that the two direct objects are also interconnected through the *conj\_and* relation.



## 7.2 Statements

This section documents the design of the Statement Annotator and its output: the Statement class, a data structure with some useful lexico-syntactic properties. Lexico-syntactic pattern matching of Statement objects is facilitated by a complementary StatementPattern class that accesses these lexico-syntactic properties. Statement objects are collected in the Profile container class, which uses StatementPatterns to produce a ranking of statements based on various desirable lexico-syntactic features and based on overlap with another Profile. The top statements of this ranking are selected to be displayed in the info box. These core classes together form the second approach to generating content for the info box.

### 7.2.1 Statement Annotator

The Statement Annotator is a custom annotator for Stanford CoreNLP. The dependency graphs produced by *nndep* form the primary input for the Statement Annotator. By examining the dependency relations between words in a sentence, the Statement Annotator is able to extract meaningful components from the sentences and organise them into Statement objects. These Statement objects can be used for various computational purposes, including lexico-syntactic pattern matching. The StatementPattern class was developed to perform pattern matching on Statement objects by considering the grammatical and/or lexical properties of the Statement.

The motivation for creating the Statement Annotator was to introduce a more formal structure to the relatively complex set of relations contained in a dependency graph. While a dependency graph contains a lot more grammatical information than the binary tree produced by applying phrase-structure grammar, its complexity also makes it harder to interpret computationally. By layering a more formal data structure on top of the dependency graph, I was hoping to structure the information in a Reddit comment history, so that sentences could be ranked based on certain lexico-syntactic patterns and in relation to other comment histories. Furthermore, using grammatical patterns to isolate for example the subject of a sentence would be a much less limiting way to discover entities compared to just using named entities. I was inspired to go in this direction because I suspected that this approach would yield a much higher pool of entities than the previous approach.

### 7.2.2 Information Extraction

On the surface, the type of information extraction needed for the application can be compared to the common task of extracting triples from a corpus to create knowledge databases. Examples include Stanford's own OpenIE (Angeli, Premkumar, and Manning, 2015). Generally, these triples take the form of

**relation(object\_x, object\_y)**

For example

**president\_of(Barack Obama, USA)**

I briefly considered using OpenIE or another of these more common information extraction solutions to save time on the software implementation. The initial idea was simply to extract triples from one set of comments and find identical triples in another set of comments, to show shared opinions or find interesting statements.

Unfortunately, despite the resulting triples being sourced from the user's own comments, they lack any context and therefore cannot be used to compare opinions in a reliable manner. It is simply not possible to know whether a person agrees or disagrees with the statement in a given triple when the syntactical structure is lacking, e.g. the user could be expressing another person's opinion or could be negating the statement elsewhere in the sentence.

## 7.3 The Statement Concept

The Statement objects included with the Statement Annotator are an attempt to solve the aforementioned issue with a lack of context. Statements are very similar to relation triples, but comprise a varying number of components – not just 3. They are also designed to be composable and may therefore contain embedded statements. This allows for expressions such as

**“she thinks bananas taste bad”**

to be structured as

**{ she, thinks, { bananas, taste, bad } }**

While a more traditional system based on triples might have just extracted { bananas, taste, bad }, this construction provides the context for the triple { she, thinks, ... } which allows one to disregard the opinion about bananas, since it has not been expressed by the writer him-/herself, but by a third party referenced by the writer.

The individual parts of conjunctions within a sentence are also extracted into separate statements, so that

**“I love playing football and basketball”**

becomes

**{ I, love, { playing, football } } and { I, love, { playing, basketball } }**

The Statement class and the basic Statement component classes that I have implemented – Verb, Subject, DirectObject, IndirectObject – include methods for accessing lexical and syntactic information. Determiners, negations, dependent clauses, possessives, and other elements are separated out from the main compound of a Statement component and can be accessed programmatically. Components also include methods for accessing relevant

grammatical information, such as the point of view of a component, the part-of-speech type, or the negation state.

## 7.4 Statement Extraction Algorithm

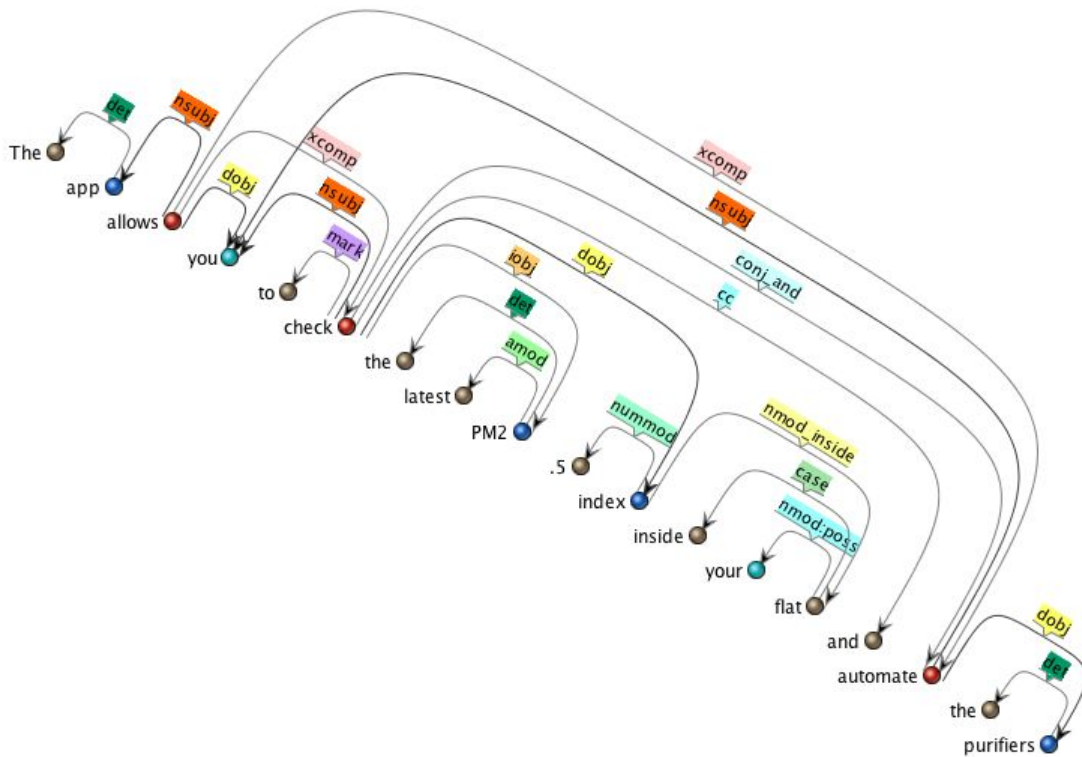
This section details the algorithm I have designed to extract Statement objects from a corpus of text, e.g. from a user's Reddit comment history. The algorithm comprises four main steps. In the following four subsections, I will describe what happens during each step of this algorithm.

```
sentence: "The app allows you to check the latest PM2.5 index inside your flat and automate the purifiers."
|
|_ statement: {S+V+DO+E: "The app allows you to check the latest PM2 .5 index inside your flat"}
|   |_ component: {Subject: "app"}
|   |_ component: {Verb: "allows"}
|   |_ component: {DirectObject: "you"}
|   |_ statement: {V+DO+IO: "to check the latest PM2 .5 index inside your flat"}
|       |_ component: {Verb: "check"}
|       |_ component: {DirectObject: "latest PM2 .5 index"}
|       |_ component: {IndirectObject: "flat", preposition: "inside", possessive: "your"}
|
|_ statement: {S+V+DO+E: "The app allows you automate the purifiers"}
|   |_ component: {Subject: "app"}
|   |_ component: {Verb: "allows"}
|   |_ component: {DirectObject: "you"}
|   |_ statement: {V+DO: "automate the purifiers"}
|       |_ component: {Verb: "automate"}
|       |_ component: {DirectObject: "purifiers"}
```

*Illustration 7c: How the system interprets an input sentence.*

The output of the algorithm can be seen in the above illustration, which was created by applying the algorithm to the dependency graph for the relatively complex sentence: "The app allows you to check the latest PM2.5 index inside your flat and automate the purifiers".

This sentence contains two separate statements made about the subject of the sentence – "the app" – and also includes two dependent clauses. The resulting output is two Statement objects that each have an embedded Statement object along with multiple constituent components.



*Illustration 7d: The dependency graph of the same sentence reveals a more complex understanding of the sentence structure.*

I have also included the dependency graph for the same sentence. I hope that these two illustrations help to illuminate how the complexity of the dependency graph is reduced in the Statement representation of the same sentence. At the same time, I would like to point out how the Statement object and its constituent component objects provide a less simplified representation of the information in the sentence than the traditional triple representation described in section [7.2.2](#).

The following is a summary of the four-step algorithm:

1. **Component Extraction:** each of the basic component types – Verb, Subject, DirectObject, IndirectObject – are extracted from a sentence by interpreting the dependency graph of the sentence.
2. **Component Reduction:** the components created in the previous step are removed from the set of extracted components if their words overlap with any of the other components.
3. **Component Linking:** the components left from the previous step are placed into separate sets based on the dependency relations of their words to the words of the other components.
4. **Statement Creation:** Statements objects are created from the sets of components that were created in the previous step.

### 7.4.1 Component Extraction

In the first step of the statement extraction algorithm, the four basic component types – Verb, Subject, DirectObject, IndirectObject – are found by examining the directed dependency graph produced by *nndep*. Each dependency graph only covers a single sentence!

The directed edges of this graph – dependency relations – give hints as to what purpose the different words of the sentence serve. A relation between two words in a sentence is represented in the graph as a governor vertex, a dependent vertex, and a labeled directed edge going from the governor to the dependent.

While designing this part of the algorithm, I used the Stanford Typed Dependencies Manual as a reference (De Marneffe and Manning, 2008). The more recent Universal Dependencies standard is based on this original standard and the standards generally overlap.

Depending on the input data, different sets of relations can be used to train the neural network of *nndep*. I used the default pre-computed model trained on English sentences annotated with Universal Dependencies and used the default “CCProcessed” representation of the graph. In the Typed Dependencies Manual this representation is defined as “*collapsed dependencies with propagation of conjunct dependencies*” (De Marneffe and Manning, 2008). I decided to use this representation as it includes more postprocessing than the basic representation does. For instance, conjunction relations are annotated with the specific type of conjunction, e.g. the *conj\_and* relation or the *conj\_or* relation, rather than just the basic *conj* relation. Dependencies are also propagated so that, for example, multiple direct objects are connected directly to their governing verb through the *dobj* relation, rather than indirectly through a sequence of dependencies. This makes it slightly simpler to process.

In the case of finding grammatical subjects, the two relations *nsubj* and *nsubjpass* were used to find the primary word of each Subject compound. These relations typically emanate from a governing verb to a dependent subject. The outgoing relations from this primary word of the Subject compound are then used to define the entire scope of the Subject compound, taking care not to include relations that are likely to lead to other individual sentence components. As a reference to understand the different relations mentioned here, I suggest downloading the aforementioned Stanford Typed Dependencies Manual (De Marneffe and Manning, 2008)

This process is repeated for each of the other three basic component types of a Statement object – Verb, DirectObject, and IndirectObject – with separate relations considered for finding the primary word and for limiting the scope in each case.

When a primary word of a component has been found, the words comprising its entire grammatical compound are stored within a data structure that can dynamically include/exclude certain words from the compound, e.g. negations or dependent clauses, based on the current computational needs.

The outcome of this step is a single set of components. These components can be either of the four types of basic Statement component: Subject, Verb, DirectObject, and IndirectObject.

### 7.4.2 Component Reduction

Unfortunately, *nndep* is not an entirely accurate parser. As a result of that, the scopes of the basic Statement components found in a sentence will sometimes overlap.

In order to have perfectly separated components, components that overlap are removed entirely based on a heuristic: a ranking of the importance given to different components. In this ranking, Verbs are considered before Subjects, which in turn are considered before DirectObjects. The lowest ranked component is the IndirectObject. Whichever component has the lowest rank is removed in case of overlap.

The outcome of this step is a reduced set of basic components that do not overlap.

### 7.4.3 Component Linking

In the third step of the algorithm, components are organised into sets based on their relations to other components. These relations are found by examining the outgoing relations of the words contained within each component.

Special care is taken to identify and separate dependent clauses from the root sentence through relations such as *xcomp* or *ccomp*. Components found within these clauses can only be linked to other components within the same clause and components within the root sentence can only be linked to other components within the root sentence. The purpose of this separation is to prevent statements from having multiple components of the same type.

The outcome of this step is multiple sets of related components.

### 7.4.4 Statement Creation

From these sets of components, initial Statement objects are created. A Statement object is – put simply – a data structure that contains components, but also potentially other Statement objects.

At this point, the basic components are organised into sets typically consisting of 1-4 components and these sets are used to create initial Statement objects.

In certain cases – often due to errors in the dependency graph – the same components can be present in two or more statements. Statements with shared components are recursively queried to remove the shared components from all except one of the implicated statements. The combination of statements that minimises the amount of duplicate component types is chosen as the optimal combination. This process resolves an issue where *nndep* will misinterpret a part of an independent clause as a part of another independent clause, typically in the presence of a conjunction word.

At this point, when there's a conjunction of several components, the remaining Statement objects will still contain multiple objects of the same component type, e.g. multiple Subjects. The Statements are recursively queried and split into multiple Statements based on this fact, until every statement at most has one single component of each type. This is the process that separates the phrase *"I love playing football and basketball"* into two separate phrases as was explained in section [7.3](#).

Lastly, the Statement objects that came from dependent clauses are embedded as a components into the individual Statement objects that they depend upon. The embedded Statement objects can themselves embed Statement objects.

The outcome of this step is a set of Statement objects.

## 7.5 Statement Patterns

This section is an introduction to lexico-syntactic pattern matching using the StatementPattern class. I developed these patterns to sort statements by quality based on desirable grammatical features, as well as for capturing and categorising relevant entities and activities found in the statements. These entities and activities form the basis for ranking statements in relation to another profile.

StatementPatterns can match and capture Statement objects and their constituent components. On the surface, they somewhat resemble regular expressions. However, while regular expressions are used to match and capture symbols in strings, StatementPatterns match grammatical properties of Statement objects and can also be used to capture the components from these Statement objects.

They are written in Java. Here is one example that I used in the project:

```
StatementPattern IDENTITY_PATTERN_1 = new StatementPattern(  
    new SubjectPattern().firstPerson(),  
    new VerbPattern().copula(),  
    new DirectObjectPattern().partsOfSpeech(Tag.noun,  
        Tag.properNoun).capture().optional().notWords(UNINTERESTING_NOUNS)  
);
```

The above pattern is used to match statements of the sort *"I am [smn]"* or *"we were [sth]"*. One of the advantages that this type of pattern matching has over regular expressions, is that the above pattern will match any statement written in the first person, not just first person singular. It also will match any verb tense, not just the present tense. It also requires that any direct object contained in the statement has certain grammatical properties and certain lexical properties, e.g. it cannot be in the list of stopwords. Furthermore, these patterns are all limited by the basic structure of the Statement class and its constituent components, making them less prone to logic errors.

Focusing on matching grammatical properties – rather than strings – makes it possible to match a broader range of expressions than would be *realistically* possible using regular expressions. StatementPatterns are also much simpler to write than any comparable regular expression. The code was also simple to implement, as it just applies basic boolean logic to the lexico-syntactic properties of the Statement class and its constituent components to perform the matches.

### 7.5.1 Patterns In Use

The following list of 24 patterns is taken from the Profile class. The full implementations are available in the *Profile* class of the *statements.profile* package in the system source code.

1. EMBEDDED\_INTERESTING\_PATTERN: interesting embedded statements
2. INTERESTING\_PATTERN: interesting statements
3. INTERESTING\_ANTIPATTERN\_1: uninteresting statements
4. CITATION\_ANTIPATTERN: citations
5. PERSONAL\_PATTERN: first person or first person possessive
6. ADVERB\_ADJECTIVE\_PATTERN: components with adverb or adjective modifiers
7. EMBEDDED\_ACTIVITY\_PATTERN: captures activities
8. LIKE\_PATTERN\_1: liked entities
9. LIKE\_PATTERN\_2: liked entities
10. DISLIKE\_PATTERN\_1: disliked entities
11. DISLIKE\_PATTERN\_2: disliked entities
12. WANT\_PATTERN: wanted entities
13. FEEL\_PATTERN: feelings
14. PROPER\_NOUN\_PATTERN: proper nouns
15. STUDY\_PATTERN: studied entities
16. WORK\_PATTERN: workplaces, projects or work identities
17. IDENTITY\_PATTERN\_1: identities
18. IDENTITY\_PATTERN\_2: identities
19. ACTIVITY\_PATTERN: activities
20. LOCATION\_PATTERN: locations
21. POSSESSION\_PATTERN\_1: possessions
22. POSSESSION\_PATTERN\_2: possessions
23. OPINION\_PATTERN\_1: opinions
24. OPINION\_PATTERN\_2: opinions

Some of these patterns are referenced in section [7.6](#) where I explain precisely how they are used. Please do note that this list of patterns is somewhat arbitrary and that I have not made any extensive evaluation of each pattern separately other than whether the pattern matches the desired pattern and captures the desired components from a test data set. In the future it would be a good idea to either evaluate the patterns separately or discover patterns using data mining as was done by Murray and Carenini (2010) for use with subjectivity detection. Unfortunately, for this thesis, I did not have the time or the relevant annotated data sets to discover patterns in this way.

Based on my experience with the limits of using Named Entities in the first approach, these patterns were intentionally designed to capture a much wider range of entities. These are not random nouns, but are instead captured based on the personal relationship of these entities



to the author. For example, locations that the author has some relationship with are captured using this pattern:

```
StatementPattern LOCATION_PATTERN = new StatementPattern(  
    new SubjectPattern().firstPerson(),  
    new VerbPattern().words(Common.LOCATION_VERB),  
    new ObjectPattern().preposition(Common.LOCATION_PREPOSITION)  
        .partsOfSpeech(Tag.noun, Tag.properNoun).capture()  
);
```

This pattern combines lexical and syntactic features to match locations that are personally relevant to the author. The lexical part consists of two word sets, LOCATION\_VERB and LOCATION\_PREPOSITION, comprising relevant synonyms for location-related verbs and prepositions found using the WordNet lexical database. It is designed to match statements such as “I once stayed in a big mansion” or “we have been to Ibiza twice over the last few years”. In such cases, the entities “a big mansion” and “Ibiza” would be captured as Statement components. Normalised string representations such as “mansion” and “ibiza” can then be extracted from these components and added to a list of location entities in a Profile object for later comparison with other Profiles.

## 7.6 Evaluating Statements for Display

This section documents the criteria used to select which of the extracted statements should be shown in the info box. There is limited space in the info box, so the displayed statements should preferably have a high probability of containing relevant information.

Statements are sorted and ranked based on lexico-syntactic patterns implemented in the Statement pattern class described in Section [7.5](#). It is probably relevant to note that these patterns neither capture all of the available information in their respective categories, nor do they necessarily capture the right information in all cases. The patterns used here are designed to match very general statements, generally favouring statements containing personal information and more opinionated language.

### 7.6.1 Interestingness

Many statements are not considered for display at all. I call these *uninteresting* statements. Conversely, statements that *are* considered for display are called *interesting* statements.

Uninteresting statements include statements with obvious grammatical deficiencies, statements that contain anaphora, or statements that are part of citations or questions, since they typically reference third party opinions.

For instance, if a statement contains anaphora in the second or third person – e.g. “she”, “you”, or “they” – or localising determiners – e.g. “this” or “their” – then it is not possible for reader to know the exact context of the statement. While this is not an issue when reading a comment on Reddit, it makes a sentence hard or impossible to understand when displayed in the info box. For that reason, all interesting statements must be contextualised.

Interesting statements are matched using the following pattern:

```
StatementPattern INTERESTING_PATTERN = new StatementPattern(  
    new SubjectPattern(),  
    new VerbPattern().negated(null),  
    new NonVerbPattern().person(Person.first, Person.third).local(false)  
        .notWords(UNINTERESTING_NOUNS).all(), EMBEDDED_INTERESTING_PATTERN  
).question(false).minSize(3);
```

## 7.6.2 Lexical density

Ideally, the statements in the interface should present as much information as possible in order to act as a heavy counterbalance to the content of the comment. This is based on the theoretical implications of the Hyperpersonal Perspective that I wrote about earlier in this thesis; namely, that reducing information scarcity can help to reduce overattribution. It follows that a measure based on information density is a good potential baseline for reducing information scarcity.

This thesis does not attempt to define what number of statements is the optimal amount to display in the interface, but realistically it is going to be limited by the attention span of the reader. In the initial feasibility study, the info box contained just 3 bullet points, while in the efficacy study outlined in Section 8, the content for the info box comprised 5 sentences.

I have chosen the measure of lexical density as a baseline measure to rank and select the statements for display. Lexical density is a simple way to score a piece of text based on the proportion of information-carrying words to the total amount of words (Analyze My Writing, 2016). The score is a value between 0 and 1, with a higher value indicating higher lexical density. The other measures used to rank statements are Quality and Relevance. The Quality measure is a series of adjustments to the baseline lexical density, while the Relevance measure is a series of adjustments to the Quality measure.

A fair criticism – and one that I would agree with – is that this might bias the system towards pre-selecting shorter and less complex statements due to these statements being more likely to score 1 in lexical density. However, I would argue that shorter sentences – provided that they are relevant – are also a better fit for an interface that is mostly intended for providing a quick cursory overview of another person.

Using the part-of-speech tagged contents of the Statement objects, calculating the lexical density is matter of dividing the number of lexical words of a statement with the total number of words. Lexical words are defined as words that are tagged as with a Part-of-Speech that falls into the categories of either verbs, nouns, adjectives, or adverbs.

## 7.6.3 Quality

Quality is a measure of a Statement's overall applicability for reducing overattribution.

Quality is derived by making small adjustments to the baseline lexical density of a statement

based on matches with the patterns listed in section [7.5.1](#). These patterns heavily favour statements in the first person and statements that contain personal information.

The Quality measure is used in this thesis to generate the 50-sentence profile shown to participants in the study described in section [8](#).

#### 7.6.4 Relevance

Relevance is a measure of a Statement's applicability for reducing overattribution *when taking the reader into account*. As such, Relevance is a relative quality measure. Relevance is derived by making small adjustments to the baseline Quality of a statement based on the presence of overlapping entities. These overlapping entities are found by comparing the entities in the same categories of two separate Profiles.

It is used in this thesis to select sentences for display in the info box based on their relation with the Profiles of the participants of the second study as described in section [8](#).

### 7.7 Limitations

The following section outlines various limitations of the system as a whole.

#### 7.7.1 Generalisability

In order to make the results generalisable across Internet discussions platforms and in order to simplify the statement selection process, I decided to only use pure text comments as the data source. This entailed not using valuable metadata, such as the subreddit the comment was written in or the karma score of the comment. I also did not include Reddit's so-called self-posts as part of the data set. I deliberately chose not to consider these features since they are all Reddit-specific phenomena. Avoiding this kind of domain-specificity helps build a stronger case for the generalisability of the algorithm.

The downside of not using such features in the input data, means that I cannot favour e.g. the highest scored and therefore most well-received comments of a user or comments from shared subreddits.

#### 7.7.2 Lexical and Syntactic Assumptions

Statement objects are data structures that encapsulate lexical and syntactic information in sentences. For this reason, they are subject to certain assumptions about the nature of the lexical and syntactical input data. When these assumptions are not supported by the data, the quality of the output Statement objects also degrades.

Lexical data comprises the words that are contained in the Statement objects. Words are assumed to be spelled correctly and consistently across the data set. Misspelled or unconventionally spelled words increase the chance that the part-of-speech tagger in the CoreNLP pipeline will assign an incorrect part-of-speech tag. Incorrectly tagged words lead to errors directly in the Statement object for those operations that rely on part-of-speech

information. Incorrect tags also increase the likelihood that the dependency parse of the sentence contains errors, indirectly causing errors in the Statement objects. The Statement Annotator assumes that words are spelled correctly.

Syntactic data comprises the grammatical relations of the sentence. Incorrect or unconventional grammar increases the chance that the dependency parse of the sentence contains errors. The Statement annotation algorithm assumes that proper English grammar is used throughout. Informal grammar, such as leaving out the subject in a sentence, nearly always causes errors in the dependency parse.

When these assumptions do not hold, the resulting Statement objects are of reduced quality.

### 7.7.3 Other Limitations

Useful Statements are all factual. Statements that are ironic, sarcastic, deceitful, or otherwise not factual will decrease the efficacy of the Profile since they inaccurately reflect the true thoughts and feelings of a person. Sarcasm detection is currently not a solved issue in NLP. Data sets where e.g. comments contain a liberal use of sarcasm are therefore not reliable sources for constructing Profiles.

Statements that contain anaphora, such as pronouns in the second or third person, are not useful for building Profiles either. As explained earlier in the thesis, anaphora are references to antecedent entities. Without knowledge of the antecedent, the anaphor can technically refer to anything, only bounded by basic attributes such as gender or singular/plural. This is due to the current implementation of the Statement annotator lacking any kind of anaphora resolution. Anaphora resolution is currently not a solved issue in NLP, although certain implementations are useful in some cases, e.g. my own implementation used in the first approach. Anaphora resolution has not been included in this second approach due to time constraints. In a future version of the software, it might be possible to expand the pool of interesting statements by including limited anaphora resolution.

## 8 Efficacy Study

This section outlines the second study I did for this thesis. In this study, participants evaluated the efficacy of the statement selection algorithm; specifically the statement ranking methods of the Profile object in the current implementation, that was used to generate summary profiles. I will be referring to the software being tested as **the system** for the sake of brevity.

As was the case with the initial feasibility study, the book by MacKenzie (2012) was used to source good practices for HCI experiments. Unfortunately, I did commit a few errors during the study which I have noted in the subsection [Errata](#) under Limitations. I have also referenced these errors in the Results sections in the places where they might have affected any of the results.

The intention of this second study was both to determine the efficacy of the system in its current state, but also to gain valuable insight into what general patterns and strategies that humans – the participants – would use to create summary profiles. The interview sections of the study were mainly focused on collecting information about these patterns and strategies, as well as evaluating the current state of the system by comparing the system results and random results to the participant's own efforts. The materials from the study are available in Appendix B.

This study is explorative and qualitative, however to get an overview I also included a section with some [Quantitative Results](#).

### 8.1 Method

In this within-subjects study, I investigated the efficacy of the system for selecting sentences for display in the info box. In the first part of the study, I was interested in knowing how well the participant felt represented by the summary profile created by the system as well as familiarising the participant with the concept of a sentence-based summary profile in preparation for the second part of the study.

In the second part of the study, I wanted to see how the statements selected by the system compared to the statements selected by the participants themselves in 3 separate cases. The participant's own selections and interview responses were used to evaluate the current efficacy of the system. To reduce the possibility of bias, the participant was presented with both the system result and a random result in each case. By not revealing which result was random and which was the system's, I expected participants to be more likely to give objective evaluations of the system.

As a side effect of this study, I wanted to see whether a much smaller data set – which is faster to process than a large data set – could still produce results with the intended effect; namely, to present a summary profile of a comment author that accomplishes the stated

purpose of the concept as outlined in [Section 1.1](#): *“to influence the impression of the comment author that is held by the reader, so that this impression becomes less negative”*. The initial study showed us that such a summary profile could have a modifying effect on how the participants interpreted and replied to a controversial comment. By using the smaller dataset, I hoped to show whether it is also realistic to perform these operations in real time: downloading a comment history, producing statements from the comments, producing a summary profile from the statements.

### 8.1.1 Participants

5 participants were recruited through an announcement on the Denmark subreddit and an announcement on Facebook. Unlike the initial feasibility study, there was no subgrouping of participants. Participants were required to be Reddit users with a sizable comment history in English, specifically around 100+ comments in the English language. This was necessary in order for the algorithm to have enough data points to generate content from.

- 4 of the participants were male, 1 was female
- They ranged in age from 26 to 36, with three being 30 years old
- All 5 participants read and participated in discussions online with strangers
- Aside from finding discussions to read on Reddit, all 5 participants also read discussions on Facebook

### 8.1.2 Data Sets

For the first part of the study, I created data sets consisting of the (up to) 1000 most recent Reddit comments written by each individual participant. These data sets were downloaded through a Python script using the PRAW library for connecting to the Reddit API. While these data sets were up to 1000 comments each, only the English-language comments were used. English language detection was performed automatically using Apache Tika, rather than manually annotating the language of each comment.

For the second part of the study, I compiled 3 data sets with 50 sentences in each. Each of the 3 data sets was based on a different Reddit user. These three users were randomly selected after a surface glance revealed that they wrote their comments in relatively proper English and seemed to have different interests. The data sets were constructed by downloading the (up to) 1000 most recent comments written by each of the three Reddit users using the same Python script. These were then reduced in size to only 50 English-language sentences, again using Apache Tika to automate language detection.

The sentences that were extracted for the 50-sentence data sets came from the first and final sentence of each comment, beginning with the most recent comment and proceeding with the second most recent comment and so on, until 50 sentences had been extracted in total. This was done so that the 50 sentences covered a wider range of topics. Otherwise, a few long-form comments in the beginning of the comment history could potentially make up the entire 50-sentence data set.

## 8.2 Procedure

Before performing the task explained below in sections [8.2.1](#) and [8.2.2](#), participants were asked to give consent and to fill out a short questionnaire. On the questionnaire it was made clear that the results of the experiment would be kept entirely anonymous and that the participant could choose not to perform a task or not to answer a question, if the participant did not want to do so.

The questionnaire was designed to collect basic demographic information as well information about the participant's familiarity with online discussions. The questionnaire is available in the appendix together with the other material from this study.

The participants were brought in separately to fill out the questionnaire, perform the tasks, and answer follow-up questions.

### 8.2.1 First Part: Personal Profile

In the first part of the study, participants were presented with a list of sentences produced by the system to constitute a profile of them. Please note that what was presented at this point was not a *summary* profile to be used as content for the info box! It was a much longer profile created using the same process. The list was produced by using the participant's comment history data as input and selecting the first 50 statements ranked by Quality from highest to lowest of the resulting output.

The participant was given a few moments to examine the results. This was then followed up by a few questions about their profile. The purpose of this first part was to familiarise the participant with the concept of profiles based around sentences, as well as hearing their evaluation of it.

### 8.2.2 Second Part: Statement Selection

In the second part of the study, participants were presented with the 3 data sets compiled from comments of anonymous Reddit users. The data sets each consisted of 50 sentences extracted from a Reddit user's comment history as described in section [8.1.2](#). A different Reddit user was the source of the comments in each of the 3 cases and they are referred to in this study as Case 1, Case 2, and Case 3, respectively.

Please note that while they also consist of 50 sentences, these data sets are *not* related to the profiles shown to participants in the [First Part](#) of the study! They are simply data sets that each contain 50 sentences from a specific Reddit user.

For each data set, participants were asked to select 5 sentences that they *personally* thought provided the "best impression" of the user. They were not told initially what criteria to use to find this "best impression". This was deliberate, as I did not want to reveal what strategy the system was using to select sentences. Another reason for not stating any criteria beforehand, was to observe what strategy participants would naturally utilise to select

sentences. There were some issues with this approach which I have outlined in the [Limitations](#) section.

After the participant had completed the first task, they moved on to the second task. Here they were presented with 5 sentences selected by the system to provide the “best impression”, referred to here as the **system result**, as well as 5 randomly selected sentences, referred to here as the **random result**. Participants were not told which sentences were randomised and which were selected by the system. They were asked to compare the 3 different results in each of the 3 cases. They were also asked to guess which result was the system result and which was randomly selected sentences, based on the assumption that the system had to accomplish the same task as the participant. This was done in order to establish whether the system had a higher efficacy than a random result.

I followed up this task with a few general questions, after revealing which of the results were created by the system in all 3 cases. I have used a semi-structured interview format throughout the study.

## 8.3 Limitations

The primary ethical consideration of this study is the fact that the participant’s own personal data is required. Participants may feel uncomfortable with the statements selected by the system or the process of having to judge and evaluate statements made by others.

Limitations include the various technical limitations of the system, as well the notable caveats which may impact the validity of the study’s result.

### 8.3.1 Level of Abstraction

While participants were instructed to consider sentences for display, the system used Statements as its level of abstraction. In each case, exactly 5 sentences were to be selected. Compared to the system, participants may then be overvaluing longer sentences because they carry more total information than shorter sentences do, while still satisfying the 5 sentence limitation. The system does not consider sentence length at all.

### 8.3.2 Limited Number of Participants

I had originally scheduled sessions with 8 different participants, but due to time constraints, I had to limit myself to 5 participants. The individual sessions ended up taking around twice the amount of time I had originally planned for. While this has resulted in a lot of quotes per participant, it has also added so much to the time spent interviewing and transcribing interviews, that I needed to cancel the remaining sessions in order to ensure that I had enough time to analyse the data properly. The low number of participants lowers the generalisability of the study’s result.



### 8.3.3 Data Set Size

The limited data set size is a major limitation of the study. The three 50 sentence data sets used for the comparisons do not represent normal data sets, but rather had to be limited in this way for the statement selection task performed by the participant. Normal data sets are several magnitudes larger and using a normal data consisting of many thousands of sentences was too impractical. Using limited data sets limits the chances of finding overlap with the participants and therefore would limit the efficacy of the system in this study.

### 8.3.4 Language

The fact that all of the participants were Danish-speaking, meant that their personal data sets were limited in size, as all non-English comments were sorted out prior to running the algorithm. It is also conceivable that their English-language comments were farther from the sort of proper English grammar required for the [Statement Extraction Algorithm](#) to work best than would have been the case if only native English-speakers were sourced as participants.

### 8.3.5 Simplified Presentation

Using patterns to capture relevant information, it is possible to construct more varied and possibly more effective information. For instance, Profiles register important locations and possessions using StatementPatterns. This information could, if deemed relevant, be presented in another manner than simply showing the sentences that the information has been derived from. However, a more rich arrangement of information would have greatly complicated the task given to the participants of this study. In the interest of having more directly comparable data with respect to the participant's choices and the system's choices, I deliberately chose to simplify the UI presented in this study so that the task would be accomplishable.

### 8.3.6 Vagueness of Definitions

All of the participants asked me to clarify what “best impression” meant. In hindsight, I should have been clearer about the meaning of the word “best” to get results that could be more easily compared to the results of the algorithm. However, asking participants to define “best” themselves also produced variation in strategies for choosing sentences. In all cases, the participants went with the definition of “best” that is more synonymous with “contains the most accurate information” from their point of view rather than “contains the most positive information”.

### 8.3.7 Errata

Unfortunately, in this study I committed a few errors that may have influenced the results. I have listed the errors in this section as well as in the results sections when they are relevant.

1. Some of the participant profiles included duplicate sentences.
2. In two cases, with two different participants, the system results for one of the 3 users included a duplicate sentence. Both participants noted this and stated that it had

affected their perception of which result was random and which was selected by the software, possibly biasing them towards choosing the random result over the non-random, when guessing the system result.

3. For all participants, the random result for Case 1 had one sentence that had actually appeared in the data set for Case 2.

The duplicate system results are particularly unfortunate, as they possibly shifted the results of the study.

## 8.4 Quantitative Results

This section contains the quantitative results of the efficacy study. The three tests with their respective data sets and results are referred to as Case 1, 2, and 3. The five participants are referred to as Participant 1 through 5.

While this was not a quantitative study – primarily since the number of participants was too low to get conclusive quantitative data – the choices made by the participants and the overlaps between the different participants and between the participant and system results provide a good starting point before the qualitative analysis of the interview data.

### 8.4.1 System Overlap

<b>Guesses</b>	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	System total	Random total
Case 1	<i>Random*</i>	<b>System</b>	<b>System</b>	<b>System</b>	<b>System</b>	4	1
Case 2	<b>System</b>	<i>Random</i>	<i>Random*</i>	<b>System</b>	<b>System</b>	3	2
Case 3	<b>System</b>	<i>Random</i>	<i>Random</i>	<b>System</b>	<b>System</b>	3	2
System	2	1	1	3	3	<b>10</b>	0
Random	1	2	2	0	0	0	<b>5</b>

Table 8a: comparing the guesses of the different participants

Overall, the system result was chosen over the random sentences by a ratio of 10:5 – or more succinctly, 2:1. The system performed best in Case 1, where 4 out of 5 participants chose the system result over the random result.

It should be noted that the results presented to Participant 1 in Case 1 contained a mistake: two sentences in the system result were duplicates. This caused Participant 1 to posit that the system result might be the random result. Participant 1 guessed as the sole Participant that the random result were in fact generated by the system in Case 1.

It should also be noted that the same type of mistake appeared when the results were presented to Participant 3 in Case 2: two sentences in the system result were duplicates. Participant 3 also posited that the system result was actually the random result. Participant 3 also guessed that the random result was in fact generated by the system in Case 2.

I have marked both of the mistakes with a star (\*) in the table above.

All of the random selections for Case 2 also, through a copy-paste mistake, featured a sentence from the data set of Case 1.

Assuming that both of the mistakes on the system results were enough to influence the participants to choose the random results over the system results when guessing the system result, the ratio of system choices to random choices could potentially have been 12:3 – or more succinctly: 4:1, a much stronger performance by the system. Assuming that the mistake on the random result for Case 2 was enough to influence the rest of the participants to select the system results over the random results, the ratio becomes 8:7 in favour of the random results. Unfortunately, it is not possible to know what choices participants would have made in the absence of these mistakes.

Moving on to the actual sentences, participant selections did not overlap notably with the system results. In fact, only 10 out of 50 selections overlapped with the system selections. This compares with 8 out of 50 selections that overlapped with the random selections. It should be noted that I am using the selections as they were presented to the participants and that these selections included 2 mistakes in the system results and 1 mistake in every random result, as outlined above and in the [Errata](#) section, so this overlap does not entirely reflect the system performance. However, any slight adjustments done post-session would not yield a significantly different result.

<b>System overlap</b>	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Total overlap
Case 1	#3*	#1, #28	#28	#3, #28	#28	7
Case 2	-	-	- *	#28	#28	2
Case 3	-	-	#23	-	-	1
Total	1	2	2	3	2	<b>10</b>

*Table 8b: the overlap with the system results as they were presented*

When looking at the overlap with the system selections, it is interesting to note that most of the overlap in Case 1 is caused by sentence 28:

**28:** “Islam is terrible even without the terrorism.”

The overlap in Case 2 comes from the same number, again sentence number 28:

**28:** “Stalin wasn't a fanatic; he was as pragmatic as they come.”

<b>Random overlap</b>	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Total
Case 1	~*	-	-	-	#41	1
Case 2	#5	-	~*	#5	-	2
Case 3	#39	#42	#39	#39, #48	-	5
Total overlap	2	1	1	3	1	<b>8</b>

*Table 8c: the overlap with the random results as they were presented to the participants.*

Sentence number 39 in Case 3 was selected by several participants and coincidentally overlapped with the random result:

**39:** “I'd say being born into a part of the world where this is an option, where you have the financial means, free time and personal freedom to do this... isn't a luck we should take for granted.”

Overall, the results show that participant overlap with the system result is perhaps indicative of a more successful system result. Case 1 had the most system overlap and the system result also selected by 4 out of 5 participants, while Case 3 had the most random overlap and the system result was only selected by 3 out of 5 participants and Case 2 had little overlap and was also selected by just 3 out of 5 participants. However, the sample is too small to make any conclusions.

#### 8.4.2 Participant Overlap

Analysing overlap between participant selection can help illuminate which key sentences participants feel create the best impression of each user. Of course, participants were not told the exact criteria behind the system selections, i.e. ranking by the Relevance score, so their selections must also be viewed with that in mind.

In this section, I have highlighted significant overlap between participants. I will go into more detail on the unique choices made by participants later in the [Interview Results](#) section.

<b>Case 1</b>	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Total
Sentence 1		x				1
Sentence 3	x			x		2
Sentence 4					x	1
<b>Sentence 16</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>		<b>4</b>
Sentence 18			x			1
Sentence 19		x				1
Sentence 20	x					1
Sentence 23				x		1
Sentence 25			x		x	2
<b>Sentence 28</b>		<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>4</b>
Sentence 30	x					1
Sentence 32				x	x	2
Sentence 35	x					1
Sentence 36			x			1
Sentence 40		x				1
Sentence 41					x	1

*Table 8d: comparing the choices of the different participants for Case 1*

In the first case, 4 out of 5 participants picked the same two sentences:

**16:** "I can't hear you over the sound of how retarded you are."

**28:** "Islam is terrible even without the terrorism."

Both sentences display negative sentiment. One is negative in tone while the other is stating a negative and controversial opinion on the religion of Islam.

<b>Case 2</b>	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Total
Sentence 5	x			x		2
Sentence 6		x				1
Sentence 9			x	x	x	3
Sentence 10	x					1
Sentence 16	x	x			x	2
Sentence 18			x			1
Sentence 23		x				1
Sentence 25	x					1
Sentence 26		x		x	x	3
Sentence 28			x	x	x	3
Sentence 34				x		1
Sentence 36	x					1
Sentence 39			x			1
Sentence 40					x	1
Sentence 45			x			1
Sentence 48		x				1

*Table 8e: comparing the choices of the different participants for Case 2*

In the Case 2, 3 out of 5 participants picked the same three sentences:

**9:** “Speaking as someone who have calculated stuff using computers for a living for several years, you still need to be able to do simple arithmetic in your head.”

**26:** “I think the biggest problem with getting it is that Americans generally don't seem to know what a speed limiter is, and so react negatively to the suggestion that the Tesla should need one.”

**28:** “Stalin wasn't a fanatic; he was as pragmatic as they come.”

Both sentence 9 and 26 are quite long, while 28 is fairly short. These sentences from Case 2 are arguably more neutral in sentiment than the sentences that had the most overlap for Case 1.

<b>Case 3</b>	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Total
Sentence 6				x		1
Sentence 9		x			x	2
Sentence 12	x	x	x	x	x	5
Sentence 16				x		1
Sentence 22	x					1
Sentence 23			x			1
Sentence 32	x					1
Sentence 33					x	1
Sentence 36	x					1
Sentence 39	x		x	x		3
Sentence 41		x	x			2
Sentence 42		x				1
Sentence 46					x	1
Sentence 48				x		1
Sentence 50		x	x		x	3

*Table 8f: comparing the choices of the different participants for Case 3.*

In Case 3, all 5 participants selected the same sentence:

**12:** “As a Mustang owner on European roads, who comes at the GT from a background of a Focus RS and an E90 M3, I have to say your impressions are probably down to the individual car you rented.”

And 3 out of 5 participants selected the same two sentences:

**39:** “I'd say being born into a part of the world where this is an option, where you have the financial means, free time and personal freedom to do this... isn't a luck we should take for granted.”

**50:** “The Michelson-Morley experiment in 1887 was a large factor behind dispelling the notion that space consists of a medium, or 'aether' which we travel through, and is in fact vacuous.”

As with Case 2, these overlapping sentences are generally quite long and mostly neutral in sentiment.

### 8.4.3 Shared Entities

There were very few shared entities between the 5 participants and each of the 3 cases. Shared entities are, as you recall, the entities and activities found in a profile that share a similar relationship to entities and activities found in another profile. They are used to rank

the statements with the shared entities higher when ranking by relevance in order to create a summary profile.

Shared entities	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Total shared
Case 1	1	1	0	0	0	2
Case 2	3	4	4	4	3	18
Case 3	0	0	0	1	0	1
Total shared	4	5	4	5	3	21

Table 8g: shared entities between participants and different cases.

From this graph it would appear that most of the shared entities were between participants and the user from Case 2. For most of the participants, the words “may”, “experience”, and “point” showed up as shared entities for Case 2. While I am unsure if “may” is a result of an error in the dependency parser, “experience” and “point” most likely come from set phrases such as “in my experience” and “my point is”. In all likelihood, these words only became shared entities for this particular reason.

The remaining shared entities between *all* of the participants and some of the cases are “sony”, “spain”, “denmark”, and “japan”.

Profile sizes	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Mean
Statements	1655	5053	801	5147	2395	3010.2
Interesting	475	1206	222	1264	564	746.2

Table 8h: All statements and interesting statements of participants.

Profile sizes	Case 1	Case 2	Case 3	Mean
Statements	94	73	83	83.3
Interesting	28	25	29	27.3

Table 8i: All statements and interesting statements of cases.

From the profile sizes, it is apparent that the profiles of Case 1 through 3, which are based on the smaller 50 sentence data sets, also have significantly fewer “interesting” statements. Statements that are “interesting” are statements that have passed the system’s contextuality check, as explained in the previous section, [Interestingness](#).

While the average participant profile on contains 3010.2 statements where about 746.2 of them are considered “interesting”, the three cases only contain 83.3 statements and 27.3 interesting statements on average, a significantly lower amount. With this in mind, the small profile sizes of Case 1 through 3 are likely the primary reason for the system failing to find many shared entities and activities for this study.



## 8.5 Interview Results

All of the interviews were conducted in Danish and all of the interview quotes in this section have been translated from Danish. The original transcribed sentences are available in Appendix B at the marked time for the specified participant.

### 8.5.1 Partially Accurate Profiles

All 5 participants felt something was missing from their own 50-sentence profiles.

Participant 1 was able to recognise something in the profile given to them (4:07). The participant remarked that (4:46) there was *“an awful lot about video games”* and that (4:16) *“it is probably what I write the most about on Reddit, so of course it’s not so strange that it pops up when it takes up the most space”*.

Participant 2 was much more hesitant and stated that (5:41) *“For the most part, it covers subjects that I’ve been interested enough in to write a comment, but I don’t think it’s, like, highlights”*. The participant later said (8:16): *“Yeah, I think something is missing. Can I put a finger on what is missing? No”*.

Participant 3 expressed something similar to Participant 2 at 5:12: *“Well, it does give an impression of what I’m interested in, but it also shows that you – at least when it comes to Reddit – have these periods where something is more interesting to spend time on than other things”*. The participant elaborated (6:45): *“I think it might give a wrong impression of me. Because it’s not that I have problem discussing politics in everyday life, for instance, but I tend to avoid doing that on Reddit, since I know what kind of shit show it can turn into if you don’t happen to agree...”*

Participant 4 thought the profile seemed *“random”* (10:06), but when asked *“so you wouldn’t have picked this?”* the participant replied (10:22) *“No, but perhaps it’s a good way to... at least it’s honest, you could say, right? The fact that it’s just random. But that’s the first thing I think: ‘it’s random’.”*

Participant 5 expressed similar thoughts to the other participants, but using more positive sentiment (5:40): *“It gives, at least, a good impression of part of my personality”*. The participant continues (6:14): *“It’s a surprising the amount of stuff you get to know about me. It’s a bit scary. That thing about where I’ve worked. I’ve actually often considered whether I should create a new user, because it’s very easy to find out who I am based on that username. But there’s a lot of stuff here: [it] knows where I work, where I live, where I have worked, but I kinda knew that. I think this mostly gives insights into the part of my personality that likes to discuss things and that likes to tell others about the things that I think I know something about.”*

## 8.5.2 Reviewing the System Results

Participant 1 had this to say, when comparing his own selections to the system and random results for Case 1 (18:09): *“Mine is relatively focused on the fact that I was – it was his negative of writing that was most apparent, so that’s why you often... it’s, like, more focused. [The system and random results] are probably more spread out, but perhaps also more accurate.”*

Some participants even guessed the system result, despite the random result featuring one of their own selections. For example, Participant 1 guessed the system result right for Case 2 based on the fact that it featured similar selections to his own (41:36; 41:56) and despite the fact that the random result contained one of his own selections (42:08).

In the eyes of Participant 2, both the system result and the random result were not very good for Case 2 (27:36; 28:26). With Case 2, the participant thought that the random result was actually the system result. Participant 2 felt the same way about the results presented in Case 3 (41:03): *“All of these show something, so they might as well be randomly selected, so I’m gonna say it’s B”*. The participant also picked the random result rather than the system result in this case.

Participant 5 mainly based his guesses on the lack of context found in the random selections (21:53; 32:54; 44:21). With Case 3 he also gave the system result a positive review (43:37): *“I almost think that those are better picks than some of my own”*. The participant later elaborated on their guess (44:21): *“It was primarily due to the the things that were in B. They were just better than those in A. There was also a couple of those in A that didn’t make much sense by themselves.”* In all 3 cases, Participant 5 correctly guessed the system result.

In the system results for Case 2, Participant 5 found an example of a sentence lacking context which had nonetheless been selected by the system (33:34): *“One of them says ‘Denmark does since the end of WW2’. In that case I wouldn’t say it gives any impression of him, because I don’t know what it is he’s talking about”*.

Summing up, the participants preferred the system results to their own in some cases, but as was also evident from overview in section [8.4.1](#), the system results were not universally preferred over the random results.

## 8.5.3 Selecting Based on Non-neutrality

In the study, Participant 3 often selected non-neutral sentences for their own results. The participant also applied this strategy when guessing the system result.

Participant 3 guessed that the system result was indeed the system result for Case 1. The participant based this on the idea that (19:13): *“A (...) gives me a more neutral impression of that person than when I was sitting here with all of the comments, compared to B”*. The result marked as “B” was the system result. In this case, Participant 3 consciously chose the result that gave a more non-neutral impression of the user in Case 1 in line with her own selection strategy.

With Case 2, the same participant had the following reason for guessing that the random result was the system result (31:10): *“It better shows the impression I had of [the user] than A does, because [A] seems every random, where [B] it sort of goes back to emphasising some of the things that I was talking about before (...) they just seem a bit friendlier the comments here, rather than just sort of cold and indifferent”*. Participant 3 elaborates (32:18): *“[A] doesn’t really give me an impression of the person, while [B] kind of gives me an impression”*. It should be noted that the system result in this case contained a duplicate sentence which might have influenced the guess of Participant 3 for Case 2.

For Case 3, despite also having overlap with the system result, the participant again chose the random result (44:14): *“based on all of his comments, I felt that it was a slightly more positive person, who had some passion for – well, ok – cars and who seemed very friendly in the things he wrote and [B] here is... I think [A] better shows that than [B].”*

To sum up, in all 3 cases, the participant based many of their own selections and all of their 3 guesses at the system result on the presence of non-neutral sentiment.

#### 8.5.4 Topic Bias

In the study, Participant 4 selected several sentences with a social science theme (32:14): *“I’ve chosen according to, what shall we say, something that tells me about social conditions in the world and what he thinks about that. And international things, something about other countries. And then there is one where he is smearing someone else. And something about Black Lives Matter and Islam”*. With Case 2, the participant notes at 49:41: *“There are a couple of themes, but one of them is no doubt technology and the other is history. There are a lot of historical comments”*. With Case 3, the participant justified his choice of sentence 6 in the following way (73:26): *“Maybe half of these comments are about cars, so number 6 differentiated itself, since it’s just a kind of general observation (...) it’s sort of anthropological”*.

Participant 4 also spent much longer than other participants explaining his reasoning for selecting various sentences within this theme. For example, while several participants picked the sentence 28 in Case 2 – the one about Stalin – Participant 4 spend the longest time of all

of the participants justifying this choice (51:24). The participant also noted that (52:44) *“he has probably read something about WW2. There are multiple comments mentioning WW2.”*

Participant 2 seemed to be biased in a different direction. This participant often selected sentences based on the presence of certain subject-specific keywords, for example with Case 1 that participant explains (15:22) that *“he’s using the MF abbreviation without any explanation, so he’s us showing that he’s not a novice. He uses these expressions without further consideration. A part of his technical terminology, so it should be a part of his profile”*.

While Participant 2 often stuck to this strategy, in some cases, the participant also more than others selected sentences with a science-related theme. For example, for Case 2, the participant selected sentence 23 because it was science-related. The participant also picked sentence 6 and provided the following explanation (25:15): *“He’s using science, talking about some physical phenomenon. One of the reasons that I wanted to exchange this for another is that he’s also talking about a game at one point (...) I’ve had to leave that out because there was [only room for] five”*. For Case 3 the participant – like some other participants – picked sentence 50 and – as the only participant out of the five – explained what the sentence was about (36:03): *“He’s referring to an experiment that disproves the existence of something called aether. It’s pretty specialised”*.

Participant 2 also picked sentence 42 for Case 3 (35:53) as the only participant: *“He has held a biology class. So he might be a biology teacher”*. This stood in contrast to Participant 5 who had the following opinion of the same sentence (44:21): *“It doesn’t say much about him except for the fact that he’s slightly funny or trying to be funny”*.

To sum up, both Participant 2 and Participant 4 selected statements that were related to very specific topics: hard sciences/teaching and social sciences/history, respectively.

### 8.5.5 The Presence of Tone

One of the major differences between the selections made by the participants and the system results was the inclusion of “tone”. By tone, I mean the way that the author is communicating with others as expressed through the sentence rather than the specific content of the sentence. The system does not consider tone or any kind of sentiment polarity at all, but many participants did so in their selections.

For example, Participant 1 selected sentence 16 from Case 1, which read

**16:** *“I can’t hear you over the sound of how retarded you are.”*

The participant reasoned that it was (11:20) *“a good sum-up of several of the other comments where he’s smearing other people”*. Participant 1 also selected sentence 35

because it (12:04) *“seems like it also fits the general theme of his opinions pretty well, which seems a bit aggressive”*. When comparing his selections to the system result and the random selections, the participant revealed that (18:09) *“it was his negative way of writing that was most apparent”*.

Participant 1 also selected sentence 22 from Case 3 based on its tone (38:12): *“I also think he generally seems like a very positive, happy person, so I included one with ‘Welcome to the club’. There are also several places where he’s thanking someone for posting something”*.

When comparing his selections with the system results, Participant 1 expressed the following (46:16): *“I can imagine that we humans can delve more into the tone and what lies behind”*.

The other participants made similar remarks about the users from Case 1 and 3 and all made selections based on tone. For example, Participant 2 selected sentence 16 from Case 1 too (13:28) *“because of the tone. It says something about how he behaves online”*.

Participant 3 selected a sentence from Case 3 that contained an emoticon because (25:33) *“this individual is sort of trying to straighten things out. So every time there was a smiley or emoji at the end, then it’s because he was making some – or she, I’m just assuming it’s a he again – some counterpoint”*.

Participant 4 also selected sentence 16 from Case 1. The participant said (24:29): *“It doesn’t really tell us much. It just tells us that he’s sitting there insulting some stranger on his computer. And that’s still something to know, I mean, that he doesn’t back out of a discussion that has turned into that.”*

Participant 5 picked sentence 25 from Case 1:

**25:** *“Maybe you would feel different if you missed an important meeting or flight, you condescending twat.”*

The participant explains that (16:26) he *“probably picked that because of his ‘you condescending twat’”*. Participant 5 also picked sentence 41 from Case 1, because it included the words *“the reality is”*. The participant stated that the reason for picking this sentence was that (19:56) *“it sounds like he’s correcting someone”*.

### 8.5.6 Summary

The interviews revealed that the Quality ranking of the system could produce profiles that were deemed at least partially correct by the participants, yet with the disclaimer that they primarily reflected their Reddit personas. The system generally performed well, even causing some participants to state that they preferred the system result to their own, yet did not manage to outperform the random result in every case. The study showed a great variance in the selection strategies of the different participants. One participant selected primarily based on non-neutrality, while others often selected sentences within very specific topics.

## 8.6 Discussion

In this section, I discuss the current challenges facing the system and how it might be improved upon, based on the results of the study.

### 8.6.1 Viability of Small Data Sets

From results presented in the [Shared Entities](#) section, it was apparent that the small data sets used in this study are not presently viable for the purpose of adjusting summary profiles by relevance. Only 4 proper shared entities were found in the entirety of the study and they did not have a big impact on the system results overall.

The small size of the data sets are, however, necessary, since spending a minute or more on creating a profile for a comment author is too long for any practical use case. I would argue that the process should take, at most, a few seconds and that this is only possible with a small data set such as the ones used for this study. This is due to the computational complexity of the system and its dependencies. The computational complexity is essentially a sunk cost, as much of it comes from building on top of the existing resource requirements of various annotators in the Stanford CoreNLP pipeline.

Having established the necessity of using a small data set, the question then becomes: is it possible to create a small data set with a higher likelihood of containing shared entities or activities?

Since downloading the needed comment data is not the bottleneck, the most obvious approach would be to construct the smaller data sets differently. Rather than picking the first and last sentences of a comment, comments should instead be selected for further processing based on the presence of the entity keywords found in the reader's Profile object. The entity keywords are easily extracted from the pre-computed profile of the Participant or any other user. Pre-selecting in this way would not introduce much overhead, but would likely increase the chances of finding shared entities while keeping the data set small.

In the [Shared Entities](#) section, I also showed that several of the shared entities found in this study were in fact just the result of participants and the user of Case 2 using certain set phrases. Adjusting by these entities adds a layer of noise to the results and it would be

desirable if this could be avoided. A future implementation would need to take into account the presence of these false positives – most likely by using an expanded list of stopwords – and then remove them from the Profile object.

It is interesting to note that Participant 2 often picked sentences with a science theme and even picked one sentence which revealed that the user might be a teacher. This is interesting because Participant 2 is actually a teacher. The same pattern can be seen with Participant 4, who seemed to select sentences with a social science or history theme. In this case, the participant has actually studied history in university. In my opinion, this topical bias can be used reaffirm the importance that relevance has for selecting statements.

### 8.6.2 System Efficacy

All participants felt that their personal profile, as represented by the 50 sentences that were selected by the system, was incomplete or only reflected their Reddit persona. However, none of the participants stated that their profiles outright mischaracterized them, so in that light the Quality ranking of the algorithm did have some efficacy.

The guesses, while they were made by a small sample of participants and possibly influenced by some mistakes made during the experiment, show that system is able to perform better than randomly selected sentences. In some cases the system result was even preferred to the participant's own. While still flawed in some key areas – particularly with regards to adjusting by Relevance – I would argue that the system's performance is otherwise quite promising, based on these results.

I suspect that the decision to have the system focus exclusively on selecting statements with proper context made up one of the major differences between the random results and the system results. However, this can be further improved by also not considering statements with the verb “do” in them for display. This was noted in one example by Participant 5. The verb “do” is *also* an anaphor and, without proper anaphora resolution, does not give a clue as to what action it actually covers.

### 8.6.3 Lessons Learned from Overlapping Sentences

It is clear from the overlapping sentence selections of the different participants, that many key sentences display one of two distinguishing features: they are either long, neutral in sentiment, and full of different information – or they are short and very controversial. This dichotomy is quite interesting, both because it allows us to see what a “human” version of the algorithm might look like, but also because it highlights similarities and differences in the results of the system and of the participants.

Two sentences displayed inter-participant overlap, but also overlapped between the system and the participants: sentence 28 from Case 1 and sentence 28 from Case 2. Both are short and controversial statements about Islam and Stalin, respectively. The system cannot distinguish controversial subjects from uncontroversial subjects, but it does favour certain sentence constructions over others by using Lexical Density as a baseline and favouring statements with proper nouns and emotional words in them, while removing most statements

that lack context. From the overlap with the participants' selections, it would appear that the system works well at finding and picking these specific sentences, despite featuring no sentiment analysis or any kind of commonsense reasoning.

One danger might of course be, that the sentences selected are so controversial as to create an even more negative impression of the user it is supposed to represent. In fact, if the goal of the system is to influence the comment reader to get a neutral impression of the comment author, then picking a sentence that denotes that "Islam is terrible" might be a very risky choice. This potential issue could perhaps be avoided by intentionally not preferring statements with certain negative words in them, unless the reader clearly shares the sentiment towards the same entity – a sort of precautionary principle with regards to words that display negative sentiment.

The longer sentences picked by the participants were generally not part of the system results. This can be explained by the fact that the object of analysis of the system is not sentences, but rather Statement objects which are a subset of a sentence. While participants may have had an interest in picking long and information-rich sentences to fit as much information as possible into their selections, the system did not consider sentence length at all. In fact, a secondary goal of the system in its final application form would be to limit the amount of words that are necessary to read, in order to make it as fast as possible to get a quick overview.

However, looking at some of the choices made by the participants it is easy to see why a sentence such as

**12:** "As a Mustang owner on European roads, who comes at the GT from a background of a Focus RS and an E90 M3, I have to say your impressions are probably down to the individual car you rented."

from Case 3 was universally picked by the participants. This sentence not only provides information about car ownership and geographical location of the author, but also reflects the overall tone and theme of the data set that the sentence was selected from. A future improvement to the algorithm should be able to find this kind of sentence too. The main complication lies with the fact that this sentence consists of multiple statements and not just a single statement. The system would have to be modified to handle this case.

#### 8.6.4 The Importance of Tone

One of the interesting results of this study, is the fact that the participants selected sentences solely based on tone. The system is obviously deficient in this regard as it doesn't consider something like sentiment polarity or the presence of emoticons when selecting statements, whereas the participants are keenly aware of these things.

Of course, it can be argued that displaying examples of rude behaviour should not be a feature of the system, as it could perhaps encourage uncivil replies towards the author of a comment. If taking this perspective, the system should learn to classify examples of rude



behaviour and use this information to not consider them for display. This is in line with overall concept of attempting to balance negatively worded comments with neutral or positive information about the author that I derived from the Hyperpersonal perspective in the [Theoretical Background](#) section of this thesis.

A counterpoint can be made, however. Perhaps certain discussions are simply always going to be hostile and negative in tone? Or maybe some things can never be agreed? If assuming this to be the case, then such a tone marker can be used as a warning to the reader that civil discussion is unlikely and that replies are best avoided.

#### 8.6.5 Concluding Remark

Summing up, the system showed some promise in this first evaluation of its efficacy, despite some mistakes made by me during the study. The study overall provided good feedback for the future direction of the system, especially in the area of tone.

## 9 Conclusion

This section is the conclusion of the thesis. In this section, I answer the research question written in section [1.2](#). I do this by first summarising the findings of the two studies, then I move on to the threats to the validity of the thesis and, finally, I outline future work, based on the challenges facing the system in its present state.

While the open-ended approach towards designing the concept and implementing the system software was challenging in itself, the biggest challenge of this thesis was simply coming up with valid methods to evaluate the complex psychological effects of such a system.

### 9.1 Summary

The thesis as a whole documents the technical challenges involved with the implementation of the concept. That being said, I should briefly mention the “restart” of the design and implementation phase exemplified in the [Second Approach](#). As outlined in the [First Approach](#), basing the implementation around the Socher et al sentiment analysis system, did not yield good results. The second approach was arguably more explorative than the first, as it was not inspired by an existing research paper. This also made the second approach a much bigger challenge to design and implement than the first, due to the much larger set of “unknowns” involved in the process. In general, designing and implementing software that deals with natural language is always going to be a challenge, as the rules of natural language do not conform to any agreed standard.

The [Theoretical Background](#) of the thesis was based on social science theory. The implication of that was that the design of the two studies had to measure a psychological effect in order to evaluate the concept and the system. With the time constraints in mind, a methodology based around qualitative studies with semi-structured interviews seemed to be the most optimal solution. The interdisciplinary nature of the work also required reading a great deal of research papers from areas of study which I am less familiar with.

Unfortunately, I did not have time to build and evaluate a prototype of the entire concept within the thesis period, so the evaluation was limited to the two present studies, neither of which sought out to test the concept in full. The [Initial Feasibility Study](#) showed that displaying a summary profile of a comment author had the potential to positively modify the impressions of the author in the mind of the reader. The [Efficacy Study](#) showed that the system had the potential to generate similarly effective summary profiles, but also showed that the current implementation of the system had issues selecting personally relevant content – one of its intended purposes – and did not consider certain aspects of the comments, such as tone, that were important to the participants. These challenges must be addressed in any future version of the system.

The most ineffectual part of the current implementation of the system is the Relevance ranking. I have outlined in more detail why this is the case in the [Discussion](#) section of the

second study. To sum up, the system results were not particularly affected by the Relevance scores, as the system did not find any significant number of shared entities or activities between participants and either of the three cases in the second study. These shared entities and activities are required for the statement Relevance ranking to be different from the statement Quality ranking. The most likely cause of this lies with the combined effects of a reduced size data set and the lack of a targeted way to assemble this smaller data set. Without a proper Relevance ranking, the resulting information in the info box is less likely to affect the reader. I base this hypothesis on several cases from both the first and second study, where the efficacy of certain pieces of information hinged on this information's perceived relevance from the perspective of the individual participant. This, I argue, is due to the [Similarity Effect](#) of interpersonal attraction.

Another current challenge of the system is the inability to consider tone when selecting statements. The system does not currently have a way of distinguishing tone and it was not considered in the design of the system. This was especially evident in the second study, where all of the participants intentionally included examples of negative or positive tone in their selections, while any occurrence of tone in the system results was completely random. This was clearly an area of consideration to the participants, so it follows that it should be an area of consideration for the system too.

## 9.2 Threats to Validity

One threat to the validity of this study lies with the fact that I have not included a proper study of the [Relevance](#) ranking using normal-sized data sets. While I have mentioned that this ranking was affected by the small size of the data sets in the second study and this is the likely cause of the resulting inefficacy of this ranking, I have not demonstrated anywhere that the ranking produces satisfying results using full data sets. I did perform several informal tests of the Relevance ranking using full data sets and was personally satisfied with the results. However, in order to obtain scientifically valid results, I would have had to conduct a separate study or to have modified the second study to also measure the difference between results based on the [Quality](#) ranking and results based on the Relevance ranking using full data sets only.

Another currently unexamined area is the hypothesised link between modified impressions and modified replies. I chose to focus my two studies on the potential of the concept and of the system to modify impressions. As such, I did not actually measure the effect of the modified impression on the reply that was written by the participants in the first study. While it would have been possible to summarise and compare the aggregate replies of the control group to the replies of the experimental group, I chose not to do this, since I suspected that the interpersonal variance of the participants and their reply styles would likely be greater than the effect that could be produced by the system. It would be possible to do this in a future study with a much greater number of participants, provided that the replies could be categorised in a standardised way, e.g. by classifying replies on a civility scale based on certain features.

Finally, the mistakes made during the efficacy study provide the most serious threat to the validity of the thesis by likely influencing the results of some of the participants.

## 9.3 Future Work

A future implementation of the system should include a new method for constructing small data sets. This method should be based on discriminating author comments based on the presence or lack of any of the entity keywords of the reader's Profile in order to maximise the chances of entity or activity overlap between author and reader. A higher chance of entity or activity overlap also increases the chance that the Relevance ranking is significantly different from the Quality ranking. Another way to increase the chance of overlap and improve the pool of selectable statements is to implement anaphora resolution to provide more context to statements. This would increase the ratio of interesting statements to total statements.

Future versions of the system should also be able to measure the tone of the statement and possibly prefer statements with a more positive tone. This could be handled in a number of different ways, for example by considering certain lexical features, such as negative/positive words as well as emoticons, or by using a less domain-dependent sentiment analysis system than the Socher et al (2013) annotator used in the [First Approach](#). From an API design standpoint, it would make sense to simply allow [Statement Patterns](#) to match against the tone of a Statement in the same way that they currently can match other syntactic and lexical features.

I would also like to see a study of the fully implemented concept, including a full client-server setup, a web browser extension, and some method of collecting statistics from participants over time. This third study would be longitudinal and would primarily use quantitative methods. The purpose of the study would be to measure the effect of the full system over time on a much larger group of participants. This could perhaps be done by measuring the frequency of certain keywords that are deemed uncivil or civil within the comments, taking cues from the research of Reddit discussion dynamics done by Tan et al. (2016) – or if tone matching is implemented, the change in tone of replies over time. To corroborate the quantitative results, the study would also include interviews with a selection of the participants.

# Bibliography

Analyze My Writing (2016) *Lexical density*. Available at:

[http://www.analyzeemywriting.com/lexical\\_density.html](http://www.analyzeemywriting.com/lexical_density.html) (Accessed: 30 August 2016).

Andor, D., Alberti, C., Weiss, D., Severyn, A., Presta, A., Ganchev, K., Petrov, S. and Collins, M. (2016) 'Title: Globally normalized transition-based neural networks', .

Angeli, G., Premkumar, M.J.J. and Manning, C.D. (2015) 'Leveraging linguistic structure for open domain information extraction', *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, . doi: 10.3115/v1/p15-1034.

Arndt, R.Z. (2012) *ToneCheck test drive: Spots 'unprofessional' Email, knows Dick Cheney sounds 'sad,' 'angry'*. Available at:

<http://www.fastcompany.com/1672974/tonecheck-test-drive-spots-unprofessional-email-knows-dick-cheney-sounds-sad-angry> (Accessed: 30 August 2016).

Biyani, P., Caragea, C. and Bhamidipati, N. (2015) 'Title: Entity-specific sentiment classification of Yahoo news comments', *CoRR*, abs/1506.03775.

Burgoon, J.K., Fischer, J., Bonito, J.A., Kam, K., Ramirez, A. and Dunbar, N.E. (2002) 'Testing the Interactivity principle: Effects of mediation, Proximity, and verbal and nonverbal Modalities in interpersonal interaction', *Journal of Communication*, 52(3), pp. 657–677. doi: <http://dx.doi.org/10.1093/joc/52.3.657>.

Byrne, D. (1961) 'Interpersonal attraction and attitude similarity', *The Journal of Abnormal and Social Psychology*, 62(3), pp. 713–715. doi: 10.1037/h0044721.

Cambria, E. and Hussain, A. (2015) *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*. Available at:

<http://www.springer.com/gp/book/9783319236537?referer=springer.com> (Accessed: 17 March 2016).

Cambria, E., Poria, S., Bisio, F., Bajpai, R. and Chaturvedi, I. (2015) 'The CLSA model: A novel framework for concept-level sentiment analysis', *Computational Linguistics and Intelligent Text Processing*, 9042, pp. 3–22. doi: 10.1007/978-3-319-18117-2\_1.

Chang, A.X. and Manning, C.D. (2014) 'TokenssRegex: Defining cascaded regular expressions over tokens', *Stanford University Technical Report*, CSTR 2014-02.

Chen, D. and Manning, C.D. (2014) 'A fast and accurate dependency Parser using neural networks', *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, . doi: 10.3115/v1/d14-1082.

Chiticariu, L., Li, Y. and Reiss, F.R. (2013) 'Rule-based information extraction is dead! Long Live rule-based information extraction systems!', *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, , pp. 827–832.

De Marneffe, M.-C. and Manning, C.D. (2008) 'Stanford typed dependencies manual', <http://nlp.stanford.edu/software/stanford-dependencies.shtml>, .

Debusmann, R. (2000) 'An introduction to dependency grammar', *Hausarbeit für das Hauptseminar Dependenzgrammatik SoSe*, 99.

Feldman, R. (2013a) 'Techniques and applications for sentiment analysis', *Communications of the ACM*, 56(4), p. 82. doi: 10.1145/2436256.2436274.

Feldman, R. (2013b) 'Techniques and applications for sentiment analysis', *Communications of the ACM*, 56(4), p. 82. doi: 10.1145/2436256.2436274.

Finkel, J.R., Grenager, T. and Manning, C. (2005) 'Incorporating non-local information into information extraction systems by Gibbs sampling', *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05*, . doi: 10.3115/1219840.1219885.

Google (2016a) 'Announcing SyntaxNet: The world's most accurate Parser goes open source', 12 May. Available at: <https://research.googleblog.com/2016/05/announcing-syntaxnet-worlds-most.html> (Accessed: 29 August 2016).

Google (2016b) *Parsey's Cousins*. Available at: <https://github.com/tensorflow/models/blob/master/syntaxnet/universal.md> (Accessed: 28 August 2016).

Hmielowski, J.D., Hutchens, M.J. and Cicchirillo, V.J. (2014) 'Living in an age of online incivility: Examining the conditional indirect effects of online discussion on political flaming', *Information, Communication & Society*, . doi: 10.1080/1369118X.2014.899609.

kemitche (2012) *The reddit blog*. Available at: <http://www.redditblog.com/2012/07/on-reddiquette.html> (Accessed: 11 March 2016).

Krebs, D. (1975) 'Empathy and altruism', *Journal of Personality and Social Psychology*, 32(6), pp. 1134–1146. doi: 10.1037/0022-3514.32.6.1134.

Liu, B. (2015) *Sentiment analysis: Mining opinions, sentiments, and emotions*. United Kingdom: Cambridge University Press.

MacKenzie, S.I. (2012) *Human-computer interaction: An empirical research perspective*. Amsterdam: Morgan Kaufmann, an imprint of Elsevier.

Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J. and McClosky, D. (2014) 'The Stanford CoreNLP natural language processing Toolkit', *Proceedings of the 52nd*

*Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, , pp. 55–60.

Marneffe, M.-C. de and Manning, C.D. (2008) *Stanford typed dependencies manual*. Available at: [http://nlp.stanford.edu/software/dependencies\\_manual.pdf](http://nlp.stanford.edu/software/dependencies_manual.pdf) .revised version from 2015-04

Montoya, M.R., Horton, R.S. and Kirchner, J. (2008) 'Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity', *Journal of Social and Personal Relationships*, 25(6), pp. 889–922. doi: 10.1177/0265407508096700.

Murray, G. and Carenini, G. (2010) 'Subjectivity detection in spoken and written conversations', *Natural Language Engineering*, 17(03), pp. 397–418. doi: 10.1017/s1351324910000264.

Novielli, N., Calefato, F. and Lanubile, F. (2015) 'The challenges of sentiment detection in the social programmer ecosystem', *SSE 2015 Proceedings of the 7th International Workshop on Social Software Engineering*, , pp. 33–40. doi: 10.1145/2804381.2804387.

Singh, R. and Ho, S.Y. (2000) 'Attitudes and attraction: A new test of the attraction, repulsion and similarity-dissimilarity asymmetry hypotheses', *British Journal of Social Psychology*, 39(2), pp. 197–211. doi: 10.1348/014466600164426.

Stanford University (2015) *Class GenderAnnotator [CoreNLP Javadoc]*. Available at: <http://nlp.stanford.edu/nlp/javadoc/javanlp/edu/stanford/nlp/pipeline/GenderAnnotator.html> (Accessed: 28 August 2016).

Stürmer, S., Snyder, M. and Omoto, A.M. (2005) 'Prosocial emotions and helping: The moderating role of group membership', *Journal of Personality and Social Psychology*, 88(3), pp. 532–546. doi: 10.1037/0022-3514.88.3.532.

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C. and Lee, L. (2016) 'Title: Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions', *Proceedings of WWW 2016*, . doi: 10.1145/2872427.2883081.

Toma, C.L. (2010) 'Perceptions of trustworthiness online', *Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW '10*, . doi: 10.1145/1718918.1718923.

Toutanova, K., Klein, D., Manning, C.D. and Singer, Y. (2003) 'Feature-rich part-of-speech tagging with a cyclic dependency network', *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL '03*, . doi: 10.3115/1073445.1073478.

Walther, J.B. (1996) 'Computer-mediated communication: Impersonal, interpersonal, and Hyperpersonal interaction', *Communication Research*, 23(1), pp. 3–43. doi: 10.1177/009365096023001001.