Freien Universität Berlin
Department of Mathematics and Computer Science

Generating Trust in Collaborative Environments
Evaluating Design Parameters in Area of Semantic Annotations

by

Jamal Al Qundus

Submitted in partial fulfilment of the requirements for the degree of
doktor rerum naturalium (Dr. rer. nat.)

Berlin, January 2019

Supervisors

First Supervisor	Prof. Dr. Adrian Paschke
Second Supervisor	Prof. Dr. Shivam Gupta

Submission Date	19th November 2018
Disputation Date	28th January 2019

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen verwendet habe. Ich erkläre weiterhin, dass ich die vorliegende Arbeit oder deren Inhalt nicht in einem früheren Promotionsverfahren eingereicht habe.

Jamal Al Qundus
Berlin, 19 November "

Acknowledgements

Praise to Allah, who has guided us to this; and we would never have been guided if Allah had not guided us.

First, I would like to thank my advisor Prof. Adrian Paschke for the continued support of my doctoral thesis and related research, for his patience, motivation, and great knowledge. His leadership helped me throughout the time of researching and writing this work. I could not have imagined a better consultant and mentor for my doctoral thesis.

Prof. Shivam Gupta always gave me courage and hope. No matter where he was and how much work he still had, he always found time for me and supported me. I am proud that Prof. Shivam Gupta is my friend.

In addition to my advisors, I would also like to thank the rest of my thesis committee: Prof. Tim Landgraf and Dr. Abderrahmane Khiat for their insightful comments and encouragements, but also for the difficult question that led me to expand my research from different perspectives.

My heartfelt thanks also go to my family and above all to my dear wife Ramona. She has kept my options open, endured me in different phases and supported me unconditionally.

Abstract

Trust is the key feature for human interaction, including the consumption of information. Due to the increasing distance communication in a digital world and the multiplicity of mostly unknown sources of information on the web, it has become difficult to identify trusted information. Increasing numbers of users consider online communities to be a source of information that can be created by almost any other user or more than one. However, one main challenge in such online communities is how to verify the credibility of this information. Furthermore, the various platforms and the complexity of the subject (trust) make the development of mechanisms to identify trustworthy information more challenging. More precisely, this raises the research question of what are the socio-technical design parameters for building trust in collaborative annotation environments? To this end, this dissertation has examined the collaborative environments Genius and Stackoverflow in light of their real data. The goal is to understand user behavior in order to identify the information characteristics that make such information trustworthy through interaction. This work proposes a trust model that comprises the dimensions stability, credibility, and quality. It calculates a trust degree of short-text based on its characteristics and classifies it into a trust class (very-trusted, trusted, untrusted and very-untrusted). The information characteristics were considered from two perspectives: Metadata and content. The evaluation of the metadata is based on user preferences within a survey, while the content is verified for its text-embedded features using data mining techniques. The proposed trust model supports the identification of trusted information in collaborative environments. It can be used in various online communities that deliver the appropriate metadata of the information provided. The trust model helps to filter the information and thus reduces the information-overload shared on the web. Applications can integrate the trust model into their development in order to increase the likelihood of their use, as users are able to recognize trusted information easily. In contrast to existing works, this thesis proposes a trust model that combines the metadata and short text characteristics to produce a human-readable interpretation of the calculated trust degree.

Vertrauen generieren in kollaborativen Umgebungen Bewertung von Designparametern im Bereich der semantischen Annotationen

Zusammenfassung

Vertrauen ist das Schlüsselelement für menschliche Interaktion, das umfasst auch die Nutzung von Informationen. Durch die zunehmende Fernkommunikation in einer digitalen Welt und die Vielzahl von meist unbekannten Informationsquellen im Web ist es schwierig geworden, vertrauenswürdige Informationen zu identifizieren. Immer mehr Nutzer betrachten Online-Communities als eine Informationsquelle, die von fast jedem anderen Nutzer oder mehr als einem Nutzer erstellt werden kann. Eine der größten Herausforderungen in solchen Online-Communities ist die Glaubwürdigkeit dieser Informationen zu überprüfen. Darüber hinaus machen die verschiedenen Plattformen und die Komplexität des Themas (Vertrauen) die Entwicklung von Mechanismen zur Identifizierung vertrauenswürdiger Informationen schwieriger. Es stellt sich die Frage, was sind die sozio-technischen Designparameter für den Aufbau von Vertrauen in kollaborativen Annotationsumgebungen? Um diese Forschungsfrage zu beantworten untersucht diese Dissertation die kollaborativen Umgebungen Genius und Stackoverflow im Hinblick auf ihre realen Daten. Ziel ist es, das Nutzerverhalten zu verstehen, um die Merkmale zu identifizieren, die solche Informationen durch Interaktion vertrauenswürdig machen. Diese Arbeit schlägt ein Vertrauensmodell vor, das die Dimensionen Stabilität, Glaubwürdigkeit und Qualität umfasst. Es berechnet einen Vertrauensgrad des Kurztextes basierend auf seinen Merkmalen und klassifiziert ihn in eine der Vertrauensklassen sehr-vertrauenswürdig, vertrauenswürdig, nicht-vertrauenswürdig und sehr-nicht-vertrauenswürdig. Die Informationsmerkmale wurden aus zwei Perspektiven betrachtet: Metadaten und Inhalte. Mit Hilfe einer Umfrage wurden die Metadaten basierend auf den Präferenzen der Nutzer auswertet, während der Inhalt mittels Data-Mining-Techniken auf seine in Text eingebetteten Merkmale überprüft wurde. Das vorgeschlagene Vertrauensmodell unterstützt die Identifizierung vertrauenswürdiger Informationen in kollaborativen Umgebungen. Es kann in verschiedenen Online-Communities verwendet werden, die die entsprechenden Metadaten der bereitgestellten Informationen liefern. Das Vertrauensmodell hilft, die Informationen zu filtern und reduziert somit das Überangebot der im Web geteilten Informationen. Anwendungen können das Vertrauensmodell in ihre Entwicklung integrieren, um die Wahrscheinlichkeit ihrer Nutzung zu erhöhen, da Benutzer in der Lage sind, vertrauenswürdige Informationen besser zu erkennen. Im Gegensatz zu bestehenden Arbeiten, schlägt diese These ein Vertrauensmodell vor, das die Metadaten und Kurztextmerkmale kombiniert, um eine menschenlesbare Interpretation des berechneten Vertrauensgrades bereitzustellen.

Contents

Abstract **v**

Zusammenfassung **vi**

Table of Contents **vii**

List of Figures **ix**

List of Tables **xii**

1 Introduction **1**
1.1 Research Motivation 2
1.2 Research Gaps 3
1.3 Research Questions 4
1.4 Contribution of the Study 6
1.5 Research Design 6
1.6 Conclusion 7

2 Background **13**
2.1 Insights into Trust 13
2.2 Trust Definitions 15
2.3 Trust Measurement 16
2.4 Trust on the Web 16
2.4.1 Genius 17
2.4.2 Wikipedia 17
2.4.3 Twitter 18
2.4.4 Community Question-Answering (CQA) 19
2.4.5 Recommendation Systems 20
2.4.6 Semantic Web discussion 21

3 Related Work **27**
3.1 Trust Dimensions 27
3.1.1 Stability 27
3.1.2 Credibility 28
3.1.3 Quality 29
3.2 Trust Computation 31
3.2.1 Trust based on Metrics 31

3.2.2 Trust based on Text-embedded Features 36
3.3 Trust Modeling 37

4 Generating Trust in Collaborative Annotation Environments 48
4.1 Abstract 48
4.2 Background & Motivation 49
4.3 Research Goals 50
4.4 Research Methodology 51
4.5 Results 52
4.6 Conclusion 53

5 Technical Analysis of the Social Media Platform Genius 60
5.1 Abstract 60
5.2 Introduction 61
5.3 Social Structure 61
5.3.1 User-Generated Content (UGC) 61
5.3.2 Participation 62
5.4 Technical Structure 64
5.4.1 Developers 64
5.4.2 Firehose 64
5.5 Activity Study 65
5.5.1 Activity Types 69
5.5.2 Collaboration on Genius 71
5.6 Conclusion 73

6 Calculating Trust in Domain Analysis: Theoretical Trust Model 78
6.1 Abstract 78
6.2 Introduction 78
6.3 Related Work 81
6.4 Domain Analysis 83
6.4.1 Genius 83
6.4.2 Annotation 83
6.4.3 Member Roles 84
6.4.4 Role Permissions 84
6.4.5 Edit Types 85
6.4.6 Data Set 86
6.4.7 Empirical Cumulative Distribution Function 87
6.5 Trust 88
6.6 Trust Model Construction 89
6.7 Results 92
6.8 Discussion 93
6.9 Conclusion 96

7 Investigating the Effect of Attributes on User Trust in Social Media 107
7.1 Abstract . . . 107
7.2 Related Work . . . 109
7.3 Trust Model Construction . . . 109
7.4 Methodology . . . 114
7.5 Results and Discussion . . . 114
7.6 Conclusion and Future Work . . . 116

8 Application Natural Language Processing 120
8.1 Part-of-Speech and Readability Indexes of Short Text . . . 120
8.2 Experiment . . . 121
8.3 Discussion . . . 122
8.4 Conclusion . . . 122

9 Design Science Research: Exploring the effects of text complexity on its quality in social media 129
9.1 Abstract . . . 129
9.2 Introduction . . . 130
9.3 Related Work . . . 131
9.4 Background . . . 132
9.4.1 Trust model . . . 132
9.4.2 Random Forest (RF) . . . 133
9.5 Approach . . . 134
9.6 Results . . . 137
9.7 Discussion . . . 140
9.7.1 RF classifier using BoW Model . . . 140
9.7.2 Theoretical Contributions . . . 140
9.7.3 Information Manipulation Theory (IMT) . . . 141
9.7.4 Managerial Implications . . . 141
9.8 Conclusion, Limitation and Future Scope of Research . . . 141

10 Conclusion 145
10.1 Discussion of the Objects Explored . . . 145
10.2 Conclusion . . . 146
10.3 Limitation . . . 149
10.4 Recommendation . . . 149
10.5 Further Work . . . 151

Appendix 152
Appendix A . . . 152
Appendix B . . . 163
Appendix C . . . 166

List of Figures

1.1 The proposed Trust Model . . . 7
1.2 The General Design Theory . . . 8

3.1 The proposed 3S-model of information trust [33] . . . 38
3.2 The proposed trust model by Mayer et al. [35] . . . 39
3.3 extended model of Mayer et al. by Kelton et al. [26] . . . 39

4.1 LWPP and HWPP Interactions . . . 53

5.1 Annotation Example . . . 62
5.2 Annotation Activity JSON Object . . . 65
5.3 Activity on Firehose . . . 66
5.4 Activity Notifications Overview . . . 67
5.5 Statistic Elements of the Collected Activities Daily . . . 67
5.6 Active Users . . . 68
5.7 Statistic Elements of Activities and Users . . . 68
5.8 Page Edits . . . 69
5.9 Statistics of Edits on Pages . . . 69
5.10 Function Curve of the Edits . . . 70
5.11 Active Users . . . 70
5.12 Interaction States . . . 74

6.1 Annotation Example . . . 84
6.2 Users IQs Distribution based on ECDF . . . 87
6.3 Edits IQs Distribution based on ECDF . . . 88
6.4 Edits Distribution based on ECDF . . . 88
6.5 Trust Model . . . 93
6.6 Number of Users per Roles . . . 94
6.7 Roles and Activities . . . 95

7.1 Respondent Task . . . 113

8.1 The average number of sentences . . . 122
8.2 The average number of complex words . . . 124
8.3 The average number of words . . . 124
8.4 The average number of characters . . . 125
8.5 FOG index . . . 125

8.6 FLESCH index 126
8.7 KINCAID index 126
8.8 Automated Readability Index 127
8.9 SMOG index 127
8.10 COLEMAN_LIAU index 128

9.1 Knime Workflow 135
9.2 Part-of-Speech Analysis 138
9.3 Readability Indexes Analysis 138

List of Tables

1.1 Trust Classification in the Literature . . . 4

3.1 Extended table of analysis of information quality dimensions. Source: [18] . 32

4.1 Collaboration Interactions . . . 52
4.2 Generated Annotation Activities by Roles . . . 53

5.1 Technical Services . . . 64
5.2 Summary Collected Activities . . . 66
5.3 Abstraction of Activity Types . . . 71
5.4 Collaboration Interactions . . . 73

6.1 Activity and Earned IQs . . . 85
6.2 Classification of Activities on Genius [2] . . . 86
6.3 Data distribution based on ECDF and Trust Degree Translator . . . 89
6.4 Description of trust Dimension . . . 91
6.5 Generated Annotations by Users/Role [2] . . . 95
6.6 Attributes and Levels Design . . . 106

7.1 Trust Degree Translator . . . 110
7.2 Attributes and Levels Calculation Example . . . 113
7.3 Attributes Importance and Levels Utilities Results . . . 115

9.1 Genius and Stackoverflow Corpus Overview . . . 134
9.2 Pre-processing Phase of the Knime-workflow including the Nodes and their description. . . . 135
9.3 Binary Combination of the Trust Classes . . . 136
9.4 Classifier Performance of the Random Forest based on Bag-of-Words Model 139
9.5 Distribution of Accepted Answers over Trust Classes . . . 140

10.1 Activity Types . . . 152
10.2 Activity Descriptions . . . 154
10.3 Role Permissions . . . 162

Chapter 1

Introduction

Due to growing globalization during the past decade, the world is moving quickly and purposefully towards digitization. For example, the number of printed newspapers[1] is gradually decreasing. The flow of information is no longer one-to-many, but many-to-many. As a consequence, a lot of information has become available on the web and users can create and share information efficiently and easily. However, the evaluation of user-generated content regarding trust is becoming a vital issue.

Online communities reach many participants around the world; alone stack exchange [2] includes more than 130 Q&A communities. These communities have one thing in common: they must meet a certain level of quality for their content to be trusted. This raises the questions of how to deal with the quality of information regarding trust in online communities and how to determine it. Furthermore, users are able to generate content of variable quality using less time, their intentions and aims being unknown. All of these factors make identifying trust more challenging.

Interactions between users at online communities require having trust in that community's sub-objects. Trust is recognized by users' activities on provided information [34, 35], which encourages users to participate and make critical decisions. For instance, trust between entities has a strong impact on the interactions, which may produce high-quality content. As aforementioned, interactions on provided information indicate trust and usually begin with active/close reading. Active or close reading, as is generally known, combines critical thinking with learning [32] and implies annotating documents by highlighting, underlining or adding comments [25]. Scholars integrate notes, comments or footnotes into the digital media and also inspire other readers [15]. Especially when annotations are performed in collaborative environments, trust is a fundamental necessity for using, adding and extending annotations of others, as annotations might be ineffective without trust.

Assuming a theory of trust in online communities can be mapped to these annotations, this will keep users active-reading and helps them take the offer to build an opinion or open a new perspective on an issue. It is, therefore, necessary to develop models and strategies for user-generated content (in particular annotations) that offer an improved policy for user participation.

[1] The number of printed newspaper is in decline, e.g. the total U.S. daily newspaper in 2017 decreased 11% from the previous year. http://www.journalism.org/fact-sheet/newspapers/ accessed 05.07.2018

[2] http://stackexchange.com/about

Our literature review reveals that existing works support the differentiation between quality and trust [10]. Meanwhile, a quality statement is unnecessary without the trust that influences the perceived quality of exchange partner interactions [26]. Therefore, the issue of trust is a top priority and there is a need to clarify its factors before developing applications that could not be used. More specifically, factors that influence trust must be understood to ensure that the quality of user-generated content can be evaluated easily. For instance, credibility, quality and other sub-factors [30] are widely investigated and designed for several domains (e.g. Wikipedia, Twitter, Yahoo-Answers, etc.).

Our investigation covers also approaches that analyze the user-generated content (i.e., text) by extracting text-embedded factors; for instance, text-complexity using readability indexes (Automated Readability Index (ARI), The Coleman-Liau Index) and other metrics (lexical frequency). These particular factors help to identify high-quality content, which is usually error-free and follows good writing styles such as a clear information structure with simple sentences. These characteristics are undisputed and generally accepted.

Trust-interpretation is a vital aspect; assuming a trust value has been calculated based on these factors, we need to examine the ways in which it can be interpreted and used for the benefit of the end user and the application [39]. We need an appropriate trust classification and a mechanism to allow a meaningful interpretation of such a trust value.

In this thesis, we pay attention to annotations provided by humans in collaborative platforms. We conduct studies to find out which design parameters are fundamental to create a model of trust. We take into consideration the trust factors established in related works and combine them together with factors defined in our context in order to establish a preliminary unique model. This Model is then evaluated and refined.

This thesis implies material from five papers [ACM International Symposium on Open Collaboration Companion, refubium.fu-berlin.de FUB, International Journal of Information Management IJIM, Database and Expert Systems Applications Technologies for Information Retrieval DEXA TIR , French Journal of Management Information Systems SIM[3], and European Conference on Information Systems ECIS[3]]. Chapter 4 relies on reference [ACM]. Chapter 5 uses material from reference [FUB]. Chapter 6 is based on reference [IJIM], which is co-authored with A. Paschke, K. Sameer and S. Gupta. Chapter 7 uses material from reference [DEXA TIR], which is co-authored with A. Paschke. Chapter 9 is based on reference [SIM and ECIS], which is co-authored with A. Paschke, S. Gupta and M. Yousef.

1.1 Research Motivation

Factors that influence trust must be understood in order to develop successful applications, which inspire confidence and that users are willing to participate in. This requires the extension of the human-to-human dimension by a human-to-machine dimension for building a trust model. Annotations are the means for conveying information and creating associations between resources. Annotations identify a specific text in a document and contain additional attributes [7] and facilitate working with annotated documents [25]. Both contributions of an online community as a platform for collaborative environment as

[3](under review)

well as annotations in scholarship are common. Both of them obtain the same keywords write-/read and trust.

Online communities (for example, Wikipedia and Genius) have a commonality in that they must have certain quality of content. Numerous studies (see [12, 13, 29], etc.) were concerned with the quality of information in online communities and how this could be determined. Quality statements are meaningless without the trust that has been found to influence the perceived quality of exchange partner interactions [26]. Therefore, the trust question has top priority with the necessity of definition of its factors. But what does trust mean? Trust is defined as a willingness to rely on an exchange partner in whom one has confidence [26]. Trust is the willingness to rely on a specific other, based on the confidence that one's trust will lead to positive outcomes [6]. Moorman et al. define trust through describing cases, in which trust is required or not. If one does not care about content, or if a trustor can be controlled by confidence, in those cases trust is unnecessary [26].

This work purposes trust as a personal trait and social response, which calls up selective attention that motivates making a critical decision based on user-generated content. It is measured by the value or the willingness of decision making to interact and consume provided critical information. Trust is individual and varies according to the individual, context and situation [5, 23, 30, 31]. Trust and quality of a contribution are not identical; in order to save time and avoid misleading declarations this expression is to be justified.

1.2 Research Gaps

In the academic literature, trust is researched from different perspectives: 1) Trust as a quality [18, 22] in the content of peer production environment e.g. Wikipedia, including a reputation system for authors [3], where communication and rating between peers are not used. For trust algorithms, analytical tools [36] or frameworks [28] are developed to compute quantitative values [2]. This approach helps to predict vandalism more than trust itself. 2) Content is classified as more or less trustworthy and is measured by surveys [20, 21], in which the collected recognition is tightly bound to time and location and the participants' number can be controlled [19] or limited [8]. In this case, this should be used as complementary to or as a confirmation of findings. 3) Trust researched as a relationship between users with transitivity characteristic [11, 33] and selected knowledge [27]; can I trust c? Find b, whom I trust, and b trusts c, so, I can trust c (a trust b $\wedge$ b trust c $\rightarrow$ a trust c). b could be something that I share with c e.g. location [40]. b can always be found if each participant is either directly known, or known through a third participant. This is not really an open collaborative environment which is the criticism of the approach. 4) Trust as a dilemma game [4] (daytrader, Prisoner's dilemma) which assumes risk, or is viewed as trust in systems and machines [9]. This view is not relevant to a collaborative environment, as the whole system is judged based on a part of it. Trust in this case only takes the values $\{0\%, 100\%\}$. You trust your car and drive it if its brake system is fully functional, nothing else. 5) Trust is based on user's behaviors and content [17, 38] e.g. Q&A to address trust prediction. This approach demands experience and research by the trustor to judge content especially when a reference is missed.

The work of Abdul-Rahman and Hails [1] and Marsh [24] are most closely related to this study, which aims to infer trust in online communities by creating a model of trust and mapping it over annotations. This work aims to understand and identify the factors that attract the attention of simple users in order to motivate them to deal with others' user-generated content. Table 1.1 summarized a classification of trust in the literature and our comments for this work.

Table 1.1: Trust Classification in the Literature

Authors	Model Metrics	Comments
Latif and Jaffry [18]; Lucassen and Schraagen [22]	Trust as a quality	It helps to predict vandalism more than trust itself.
Levin, Cross, and Abrams [20]; Liu et al. [21]	Content classification based on surveys	It should be used as complementary to or a confirmation of findings.
Golbeck and Hendler [11]; Scissors, Gill, and Gergle [33]	Transitivity relationship between users	Its environment is not an open one; users should find a relationship between each other.
Bos et al. [4]; Friedman, Khan Jr, and Howe[9]	Trust as a dilemma game or trust in systems and machines	Trust values are just 0% or 100%.
Knowles et al. [17]; West [38]	Trust prediction e.g. Q&A	It demands experience and research by trustor.

Tab.1.1 illustrates an overview of the trust categories in the literature and comments regarding to the proposed research in this work.

1.3 Research Questions

Why Do You Trust What You Trust?

This research graduates in two scholarly pieces: Theoretical and practical foundation. In the first part we establish a theoretical model of factors that influence trust of exchange partners inferring from our study on social media and from related works in the literature. The goal is to create a model of trust to be a reference for developing applications oriented on collaborative annotation, since we also believe that trust plays a vital part as a bridge between information quality and information usage [6, 16]. Such model contains socio-technical design parameters inferred from online communities operated on collaborative content and the state-of-the-art of semantic annotation technologies. To our knowledge, this is the first research in this area that addresses this challenge.

The collaborative platform Genius[4], as a representative of social media, is chosen in this work to be researched, analyzed and compared with others (e.g. Stackoverflow[5]) for addressing the aspects for trust in knowledge creating processes. We also compare trust factors established in related works with our design parameter to set up a model for the present.

[4] http://genius.com
[5] http://stackoverflow.com

In order to find out which elements guide trust, this research focuses on the following overarching question:

What are the socio-technical design parameters for building trust in collaborative annotation environments?

From the main research question a set of sub-questions can be derived:

I How can behaviour in online communities be? How to classify human activities in online communities?

II What is the correspondence between trust and quality? Is quality an obtainable fact of trust or vice-versa? Is quality equal to trust or does there exist a logical topology (according to our mental recognition of trust)?

III What influence does trust have on annotation and on human interaction?

IV What are the factors for measuring trust on annotation? How to provide a template that expresses such factors?

For the effective reuse of other parties' annotations, collaborative platforms are examined to find out which design parameters are essential to create a model of trust. Collaborative platforms such as Genius enable the analysis of user behavior in the sense of trust. The Genius Annotation activity deals for example with many interactions like Upvote, Downvote, Suggestion, Reply, etc. as well as with different specific member roles in relation to different permissions. From this information, we can derive design parameters for a human-to-machine model of trust.

In addition, we have analysed existing research on semantic annotation platforms to find out how annotations are presented to users and what possibilities for collaboration already exist in this context. Trust models provided by related works are also considered to combine the trust characteristics used and merge with the model generated in this work to temporarily adapt to a unique model. From this study design parameters for a human-to-machine model of trust are considered. Based on both classes of design parameters (socio-technical) a classification is developed to allow a further definition of possible design parameters by inference.

The results were evaluated to improve the supposed trust model and to develop a recommendation for the socio-technical design parameters necessary to build trust in collaborative environments. Our research conducted a further investigation of the content of texts. The actual content, especially the text, plays an important role in a user's trust decision. This work uses existing techniques in the field of text analysis to explore these contents on syntactic and lexical levels.

The analysis of user behavior on Genius and existing research on semantic annotation platforms answer our research question. The evaluation reviewed and supplemented this answer in order to improve the exchange of information and cooperation through a model of trust.

1.4 Contribution of the Study

The aim of this work is to define trust and compare its factors that can be founded by related works with the proposed design parameters to establish a model for the present. This model means a recipe of trust in a written text provided by others and represents a reference to design annotation-oriented tools. The synthesis is that users will have more trust in systems considering the proposed model.

This research enjoys the following benefits:

I The type of trust this thesis investigates is intuitive, limited and its formalism using a tangible model.

II This thesis analyzed user real-data and makes conclusions to suggest a strategy describing how to provide information to support user making-decision to trust.

III This thesis aims to encourage valuable knowledge sharing by improving application development using the trust model proposed.

IV The trust model involves comprehensive dimensions and relies on evaluations and metrics derived from solid literature investigations, but due to the separation of the components of its mechanism, it is variable and flexible in its attitude to include further metrics that are relevant for integration in other but similar domains.

V Due to the complexity of the topic of trust, this nevertheless completed and documented research forms a part of an ongoing series of integrated scientific research in this field.

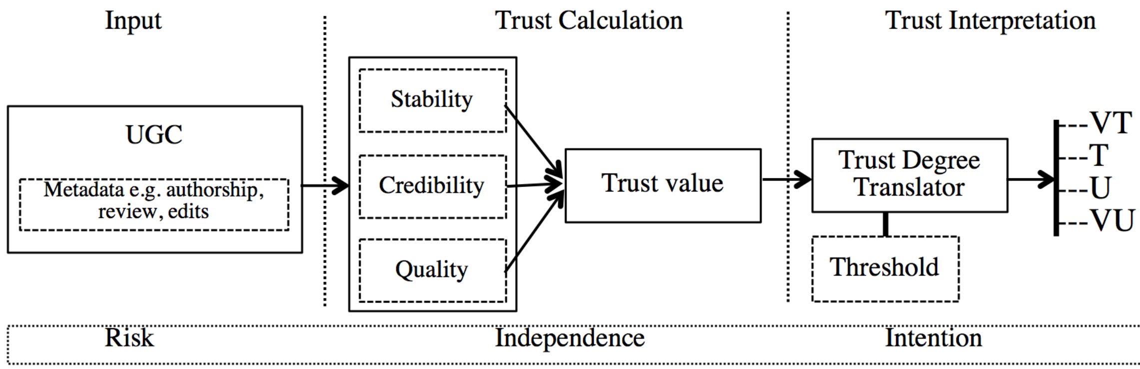
The trust model proposed in this thesis is illustrated in Figure 1.1. It consists of four stages: 1) Input that contains an annotation to be investigated and a set of metrics, which consists of metadata (e.g. rating, authorship etc.) and text-embedded features (e.g. syntactic analysis). These metrics are taken into account in the next stage; 2) Calculation that assesses a concrete value of trust using equations operating on the metrics; the calculated trust value is passed at the 3) Interpretation stage that applies a predefined threshold to classify the value and maps it to 4) the Output in one of the trust classes (Very Trusted, Trusted, Untrusted and Very Untrusted).

The following conditions shall help bounding trust: 1) including risk, a user (trustor) should be vulnerable in usage of information provided, that is, the information is important. 2) Independence, an information provider (trustee) cannot be controlled. If I can control the provider, so the trust question is senseless and 3) intention, the information provided can be incorrect, but it is not intentionally manipulated.

1.5 Research Design

This thesis follows the design cycle of the General Design Theory (GDT) [37] as illustrated in Figure 1.2. It consists of the following processes: 1) the *awareness of a problem*: to determine a new problem by comparing existing knowledge in the same issue context; 2) *suggestion*: to suggest approaches for solving that problem. This phase contributes to the

Figure 1.1: The proposed Trust Model



This is an abstract illustration of our trust model, which takes an annotation as an input and gives a human-readable interpretation of the trust degree calculated.

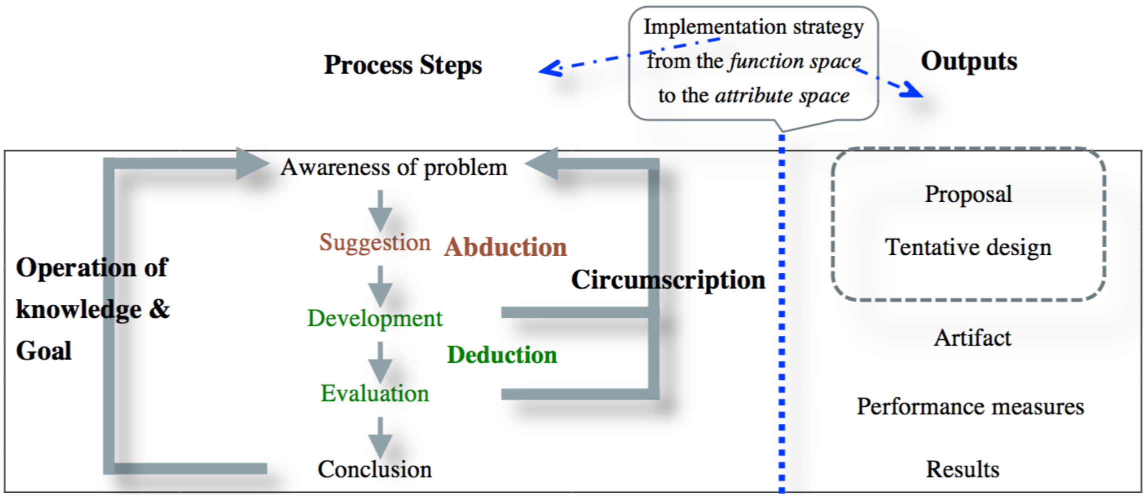
abduction stage; 3) *development*: to give concrete form to the solution candidates. This and the next phase contribute to the *deduction* stage; 4) *evaluation*: to accept, refine or refuse the whole or part of these candidates by use of various methods, such as surveys and case studies; 5) *conclusion*: to decide which candidates form the solution. These steps will be gradually modified (*development* and *evaluation* are conducted by *circumscription*, and *conclusion* is affected by *operation of knowledge & goal*) to complete the cycle of the research design towards the *awareness of the problem*. These process steps are illustrated on the left part of the figure. On the right side are the outputs that include 1) a *proposal* describing a *tentative design* that results from the *awareness of the problem* and *suggestion* of the process steps; based on the *suggestion* step 2) an *artifact* is translated from the *development* step and evaluated with regard to 3) *performance measures* in *evaluation* step; 4) *results* which are the product of the *conclusion* step.

The working hypothesis is based on the assumption that trust is fundamental for knowledge creation in collaborative annotation environments and for collaborative communication [14] in general. Based on this working hypothesis the following metrics can be derived: 1) By studying knowledge creation environments such as Genius, we can learn to which extent trust is needed for knowledge creation and what design parameters are required to create trust. 2) By studying existing research on trust modeling, we can learn how user-generated content are carried out in state-of-the-art contexts and which measures are used to predict trust. 3) These insights can be integrated into a model of trust in order to theoretically derive further parameters.

1.6 Conclusion

Particularly due to the increasing information and knowledge exchange in collaborative environments, factors influencing trust must be understood in order to develop successful systems with a confidence and willingness to use. There are several ways for providing information and one of those is annotation. Platforms such as Hypothes.is and the knowledge base Genius show the increasing importance of annotations in user-generated content. The

Figure 1.2: The General Design Theory



The General Design Theory (GDT) derived from (Takeda et al. 1990) including the cognitive model of design processes.

focus is on this area since information should be extended and connected. This research involves two scholarly stages: Theoretical and practical Foundation. In the first stage the work attempts to establish a theoretical model of factors that influence trust of exchange partners inferred from studying social media and related works in the literature. Related works from other disciplines are a possibility for additional investigation for this research. In the second stage this work aims to improve the adapted model to predict trust and argue related factors by means of simulation the decision-making to trust in users' real-life.

Bibliography

[1] A. Abdul-Rahman and S. Hailes, *Supporting trust in virtual communities*, in System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on, IEEE, 2000, pp. 9–pp.

[2] B. T. Adler, K. Chatterjee, L. De Alfaro, M. Faella, I. Pye, and V. Raman, *Assigning trust to wikipedia content*, in Proceedings of the 4th International Symposium on Wikis, ACM, 2008, p. 26.

[3] B. T. Adler and L. De Alfaro, *A content-driven reputation system for the wikipedia*, in Proceedings of the 16th international conference on World Wide Web, ACM, 2007, pp. 261–270.

[4] N. Bos, J. Olson, D. Gergle, G. Olson, and Z. Wright, *Effects of four computer-mediated communications channels on trust development*, in Proceedings of the SIGCHI conference on human factors in computing systems, ACM, 2002, pp. 135–140.

[5] X. Cheng, S. Fu, and G.-J. de Vreede, *Understanding trust influencing factors in social media communication: A qualitative study*, International Journal of Information Management, 37 (2017), pp. 25–35.

[6] K. Chopra, W. Wallace, et al., *Trust in electronic environments*, in System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on, IEEE, 2003, pp. 10–pp.

[7] L. Denoue and L. Vignollet, *An annotation tool for web browsers and its applications to information retrieval*, in Content-Based Multimedia Information Access-Volume 1, LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, 2000, pp. 180–195.

[8] F. J. Fowler Jr, *Survey research methods*, Sage publications, 2013.

[9] B. Friedman, P. H. Khan Jr, and D. C. Howe, *Trust online*, Communications of the ACM, 43 (2000), pp. 34–40.

[10] Y. Gil and D. Artz, *Towards content trust of web resources*, Web Semantics: Science, Services and Agents on the World Wide Web, 5 (2007), pp. 227–239.

[11] J. Golbeck and J. Hendler, *Inferring binary trust relationships in web-based social networks*, ACM Transactions on Internet Technology (TOIT), 6 (2006), pp. 497–529.

[12] F. M. Harper, D. Raban, S. Rafaeli, and J. A. Konstan, *Predictors of answer quality in online q&a sites*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2008, pp. 865–874.

[13] K. Hart and A. Sarma, *Perceptions of answer quality in an online technical question and answer forum*, in Proceedings of the 7th International Workshop on Cooperative and Human Aspects of Software Engineering, ACM, 2014, pp. 103–106.

[14] N. S. Hill, K. M. Bartol, P. E. Tesluk, and G. A. Langa, *Organizational context and face-to-face interaction: Influences on the development of trust and collaborative behaviors in computer-mediated groups*, Organizational Behavior and Human Decision Processes, 108 (2009), pp. 187–201.

[15] Z. Hood, N. Sahari, et al., *Researchers annotation collections and practices*, Procedia Technology, 11 (2013), pp. 354–358.

[16] K. Kelton, K. R. Fleischmann, and W. A. Wallace, *Trust in digital information*, Journal of the American Society for Information Science and Technology, 59 (2008), pp. 363–374.

[17] B. Knowles, M. Rouncefield, M. Harding, N. Davies, L. Blair, J. Hannon, J. Walden, and D. Wang, *Models and patterns of trust*, in Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, 2015, pp. 328–338.

[18] I. Latif and S. W. Jaffry, *Trust evaluation mechanisms for wikipedia*, in Sixth International Joint Conference on Natural Language Processing, October 2013, pp. 36–42.

[19] S. Lefever, M. Dal, and A. Matthiasdottir, *Online data collection in academic research: advantages and limitations*, British Journal of Educational Technology, 38 (2007), pp. 574–582.

[20] D. Z. Levin, R. Cross, and L. C. Abrams, *Why should i trust you? predictors of interpersonal trust in a knowledge transfer context*, Academy of Management, (2002).

[21] H. Liu, E.-P. Lim, H. W. Lauw, M.-T. Le, A. Sun, J. Srivastava, and Y. Kim, *Predicting trusts among users of online communities: an epinions case study*, in Proceedings of the 9th ACM conference on Electronic commerce, ACM, 2008, pp. 310–319.

[22] T. Lucassen and J. M. Schraagen, *Trust in wikipedia: how users trust information from an unknown source*, in Proceedings of the 4th workshop on Information credibility, ACM, 2010, pp. 19–26.

[23] P. Manuel, *A trust model of cloud computing based on quality of service*, Annals of Operations Research, 233 (2015), pp. 281–292.

[24] S. P. Marsh, *Formalising trust as a computational concept*, (1994).

[25] C. C. Marshall, *Toward an ecology of hypertext annotation*, in Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space—structure in hypermedia systems: links, objects, time and space—structure in hypermedia systems, ACM, 1998, pp. 40–49.

[26] C. Moorman, R. Deshpande, and G. Zaltman, *Factors affecting trust in market research relationships*, the Journal of Marketing, (1993), pp. 81–101.

[27] M. Mortensen and T. B. Neeley, *Reflected knowledge and trust in global collaboration*, Management Science, 58 (2012), pp. 2207–2224.

[28] S. T. Moturu and H. Liu, *Evaluating the trustworthiness of wikipedia articles through quality and credibility*, in Proceedings of the 5th International Symposium on Wikis and Open Collaboration, ACM, 2009, p. 28.

[29] R. Muilwijk, *Trust in online information-a comparison among high school students, college students and phd students with regard to trust in wikipedia*, (2012).

[30] J. R. Nurse, I. Agrafiotis, M. Goldsmith, S. Creese, and K. Lamberts, *Two sides of the coin: measuring and communicating the trustworthiness of online information*, Journal of Trust Management, 1 (2014), p. 5.

[31] J. R. Nurse, S. Creese, M. Goldsmith, and S. S. Rahman, *Supporting human decision-making online using information-trustworthiness metrics*, in International Conference on Human Aspects of Information Security, Privacy, and Trust, Springer, 2013, pp. 316–325.

[32] B. N. Schilit, G. Golovchinsky, and M. N. Price, *Beyond paper: supporting active reading with free form digital ink annotations*, in Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co., 1998, pp. 249–256.

[33] L. E. Scissors, A. J. Gill, and D. Gergle, *Linguistic mimicry and trust in text-based cmc*, in Proceedings of the 2008 ACM conference on Computer supported cooperative work, ACM, 2008, pp. 277–280.

[34] F. Skopik, D. Schall, and S. Dustdar, *The cycle of trust in mixed service-oriented systems*, in Software Engineering and Advanced Applications, 2009. SEAA'09. 35th Euromicro Conference on, IEEE, 2009, pp. 72–79.

[35] F. Skopik, D. Schall, and S. Dustdar, *Trustworthy interaction balancing in mixed service-oriented systems*, in Proceedings of the 2010 ACM Symposium on Applied Computing, ACM, 2010, pp. 799–806.

[36] S. Sousa, I. Shmorgun, D. Lamas, and A. Arakelyan, *A design space for trust-enabling interaction design*, in Proceedings of the 2014 Mulitmedia, Interaction, Design and Innovation International Conference on Multimedia, Interaction, Design and Innovation, ACM, 2014, pp. 1–8.

[37] H. Takeda, P. Veerkamp, and H. Yoshikawa, *Modeling design process*, AI magazine, 11 (1990), p. 37.

[38] A. G. West, *Calculating and Presenting Trust in Collaborative Content*, PhD thesis, Citeseer, 2010.

[39] A. G. West, J. Chang, K. K. Venkatasubramanian, and I. Lee, *Trust in collaborative web applications*, Future Generation Computer Systems, 28 (2012), pp. 1238–1251.

[40] J. Zheng, E. Veinott, N. Bos, J. S. Olson, and G. M. Olson, *Trust without touch: jumpstarting long distance trust with initial social activities*, in Proceedings of the SIGCHI conference on human factors in computing systems, ACM, 2002.

Chapter 2

Background

Objects can only be correctly understood in their environment. We need that process called sensemaking to help us to structure and understand retrieval information. What are you thinking about, when you hear a statement like "you can trust this doctor?" Can you trust his suggestion about a company share? What does trust mean? What is the domain of trust? According to [18] trust is needed on the Web as much as it is in the real world and reflects belief that a producer will create useful information, and willingness to invest some time in reading and identifying useful content; that includes social trust. "if users can identify the information producers they trust online, then they can spend their time more effectively by working with information from them" [27].

This chapter deals with the ambiguity of understanding the various terms related to trust from the perspective of the thesis and gives an overview of trust definitions and trust modeling in the context of computer science. How trust values are reflected and the mental trust model of a selection towards online communities (e.g. Genius, Wikipedia and Stackoverflow) are also presented in this chapter.

2.1 Insights into Trust

There are numerous terms related to trust, which we will introduce first. These terms build a border around the trust definition used in this work and are partially proposed by [12].

Untrust In particular situation, a trustor (T_{or}) knows that a trustee (T_{ee}) could not act satisfactory due to the limited competence. In such a situation, the T_{or} untrusts the T_{ee}, while in other situations, it is a different matter [31]. Untrust is also referred to as Mix-Trust [13].

Distrust is accompanied by the feeling of strong negative emotions, fearing harmful content due to lack of information about the content, therefore, a trustor is not willing to consume user-generated content provided by a trustee [25]. Distrust is *an active phenomenon*, in which T_{or} considered the situation, made a conclusion that T_{ee} has negative intentions towards him and distrusts T_{ee} [31].

Undistrust trustor cannot sufficiently make definite conclusions to or to not distrust a trustee [22].

Blind-trust In case the alternatives are worse and we cannot do otherwise, then we may have to trust blindly [16].

No-Trust In the context of trust modeling, the difference in meaning between untrust and no-trust is the existence of information about the trustee. If T_{or} knows T_{ee} cannot be trusted, we are talking about untrust. If T_{or} does not know if or how much T_{ee} can be trusted, we talk about no-trust [23].

Bounded trust Trust exists, but is tolerated by distrust and is under condition of verification by monitoring the actions carried out [7].

Trust in Agents Trust in a recommendation agent as an individual's confidence in an agent's competence and integrity [8]. Agents should be able to decide which other agents they could trust more and be able to allow the agents to adjust their understanding of another agent's subjective recommendations [1].

Mistrust Assuming misinformation is given in some form passively (intentionally or not), then we also assume that Mistrust is *misplaced* trust. In other words, after a decision in which there was a positive assessment of trust, and where the trustor was *betrayed* (in the case the trustees has bad intentions), we can say that the trustor mistrusted the trustee [31].

Dismistrust means erroneously rejecting good information. On the contrary, Mistrust means erroneously accepting bad information [33].

Trust in Trust Luhmann describes one's trust in another person's trust when one's trust has the ability to motivate others, e.g. to make decisions. This provides an additional motive for own trust. The author also describes the trust of others in own trust. This allows basing own action plans on the trust of others [30] -page 69-.

Pro-active trust is the highest level of trust, because it implies the knowledge that the subject is loyal and confirms that the subject to be analysed can be trusted [44].

Trust differs according to the mental model related to the different disciplines. In the context of this research, computer science, trust is considered from two aspects: 1) **Content** and 2) **user**.

1) **Content** provides the interaction types reading and sharing. Within a reading a user consumes content provided, while within sharing a user generates content. The users of these two interaction types take different roles regarding trust. A trustor, who is the user at the consuming activity, will have to make a decision to act on the content based on whether he trusts or not. Another role is trustee, who is the user at the sharing activity, will have to make a decision of trust toward the community he aims to generate content to. This kind of trust, which our thesis does not consider, could be build by inferring trust as trustor towards the content. However, this thesis considers the content aspect including the reading and sharing interaction types to compute a model of information trust.

Another aspect of trust-consideration is that of 2) **user** that is classified in trust between users (human-human)[1] and trust between user and object (human-system)[2].

In human-human trust, users are reflected using a network of nodes, which illustrate users, and edges, which present the relationship between the nodes connected. While a user-system trust takes the value {0%, 100%}, in case the system is represented as a physical object, because a physical object, e.g. a cash machine, can only be estimated as a unit and one trades with it if one trusts (100%) it or does not trade with it (0%).

Fake-news, **all trust terms introduced before**, **machine-generated content**

[1] Distance communication using computer is included.

[2] Applications are considered as a part of systems.

as well as **emotion and opinion** are separate research areas and are not considered in this work, which has a different focus. Fake-news is news that contains fabricated information with the aim to mislead readers intentionally [43] and are mostly distributed via social media. This thesis is based on the assumption that content is created with a clear conscience. Even if this content would be of low quality or even contains incorrect information, we do not assume that the user (contributor) intended it. The research area of machine-generated content i.e. annotations focuses on content extensions inferred from another data source e.g. data base. While emotion and opinion rely on reputation or word-of-mouth within trust decision-making.

2.2 Trust Definitions

Trust varies between individuals, context and time. The stability of a community depends at all times on the right balance between trust and mistrust [1]. Intuitively, trust embodies a risk-bearing, directed connection between entities to achieve a specific objective in a specific context at a specific time. Traditionally, these entities are two -trustor and trustee- that could be humans, physical or virtual objects or a mix of all.

According to Deutsch, trust behaviour occurs when a person made a decision to follow a given ambiguous path its result occurs positive or negative depending on another person [14]. This initial definition is generally applicable and is used in many disciplines (see [30, 31, 42] and others.).

Kelton et al. studied trust on four levels: 1) Individual, as a personality trait addressing the example statement "I trust". 2) Interpersonal, as a social tie directed from one actor to another extending the statement to "I trust you". 3) Relational, as an emergent property of a mutual relationship extending the statement further to "You and I trust each other" and 4) Societal, as a feature of a community as a whole extending the statement further more to "We all trust" [24].

Solhaug et al. define trust as the "subjective probability" of performing a given action as expected from the trustee depending on the "welfare" of the trustor; the authors define trustworthiness similarly, but as "objective probability" instead [45]. Obviously, incidents must be viewed objectively, i.e. described as observed. However, it is not always possible to completely ignore the personal point of view. The descriptions of an incident differ from person to person and are always somewhat subjective, since perception depends on the person and is usually inferred from given criteria related to the situation of the incident considered. In this definition and as it is mentioned in [30], the term "probability" includes a "threshold" to separate several scopes of trust. Despite trust being originally a human merit, which is relative, this motivates measuring trust to be reflected as a value (in a range).

Grandison and Solman state trust as a complex subject that is a composition of various attributes ("reliability, dependability, honesty, truthfulness, security, competence and timeliness"), which are environmentally dependent. From this perspective, they define trust as "the firm belief in the competence of any entity to act dependably, securely and reliably within a specified context" [21]. This definition best describes the observation we follow that a key correlation exists, on the one hand, between trust and a given situation and, on the other hand, between trust and trustee quality.

This thesis is in line with the stated conditions risk and interdependence of trust by Rousseau et al.; trust is required in a given situation if a probability exists to regret a decision made, which was needed to achieve the trustor interests. The author define trust as "the willingness to be vulnerable under conditions of risk and interdependence" [42].

Similarly to this definition of trust, in which trust decision leads to positive outcomes toward the trustor, Mayer et al. define trust as "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" [32]. Accordingly, Lu et al. describe trust in that situation, where an entity (trustor) accepts vulnerability based on positive assumptions about another entity's (trustee) intentions or behavior [28]. Further in the same spirit, Kim and Ahmad define trust as the measure of belief in the task competence of a content provider, on the assumption that the content provider generally and consistently produces satisfactory and high-quality content. Despite the possibility of risk, the content consumer is prepared to accept user-generated content. "This action is accompanied by feelings of security and strong positive emotions" [25].

Noh et al. introduce trust based on the activity types describing the relationship of trust between users (trustor and trustee) as "implicit" or "explicit" and giving the examples of implicit as a trustor leaves a comment or review for a trustee, and a trustor follows a trustee for the case of explicit trust. The authors give four forms for this relationship between trustee and trustor, which are one-to-one, one-to-many, many-to-one or many-to-many structures [38]. This illustration has been inspired this thesis before. We hold the opinion that trust is indicated and can be only observed by interactions between trustor(s) and trustee(s), whereas each of them can appear as a single user or as a group of users.

2.3 Trust Measurement

Quantifying trust can take various forms as a result reflecting a trust level to the observer. In the following part, we introduce briefly such forms into two categories [23] 1) numeric i.e. continuous value, static value as an integer and rational number of a finite range area, and 2) textual e.g. trusted, very trusted etc.

The trust measure that occurs as numeric is represented as a number (fuzzy logic or a percentage) e.g. 24, 7, 80, as a level value e.g. {-1, 0, 1} or as a rang e.g. [0, 1] (e.g. [12, 25, 31, 46]). Generally, these representations are more suitable to express trust value in the context of systems, communities and, in case of high latency time, agents.

Trust depends on the assessment of the individual observer and, thus, cannot be objective but a "subjective probability" [16] and a "subjective degree of belief" [34] as text-interpretation degrees (e.g. [1, 11, 24]). However, this text representation requires a pre-calculation of trust metrics that are converted to a numerical value. Then, as a general rule, a threshold value determined on the basis of pre-observation and training is used to map this numerical value into a trust level class.

2.4 Trust on the Web

"strong communities are formed by

entities which trust each other" [2].

Our aim, in this section, is to review literature research about the mental model of trust in some subdomains of online communities: Genius[3], Wikipedia[4], Twitter[5] Stackoverflow[6] and an overview of investigations related to Recommendation Systems.

2.4.1 Genius

Genius, as social media platform for text interpretation in form of annotation, still young (it exists since 2009), but fast growing. To our knowledge, our research considering Genius as case study is novel. We have described the platform as part of our research. All aspects of Genius and its role as an important research area as well as activities including annotation, users and existing features are summarized in Chapter 5 and are located in more detail in our technical report in [4].

2.4.2 Wikipedia

Open Collaborative Authoring System (OCAS) e.g. Wikipedia is designed as an open editing model, therefore, trust on it's content still relevant. Wikipedia is a collaborative, free online encyclopedia platform that provides summaries of information on almost anything with the entire spectrum of knowledge from a wider range of sources. Anybody can use the wiki-tool directly in the web browser to contribute or edit the information provided in form of articles, that is what makes it important. However, this openness offers the possibility of manipulation and vandalism and presents us with the challenge that the quality of Wikipedia articles cannot be guaranteed [47]. Therefore, many researches have investigated and contributed to the solution of this issue.

Adler et al. propose a system that calculates quantitative trust values as a reliability indicator for the text in Wikipedia articles; this system or algorithm computes trust values of a word across the text in an article. The calculation is based on the information registered in the revision history of the article about the reputation of the original author and of the editors and so all changes to an article are reflected in sum of the computed trust values. The algorithm has been applied on thousands of articles of the English Wikipedia to compute and display trust of these articles. These trust values supply an evidence for reliability of the text. Additionally, it ensures that all changes to the text of an article are represented in the trust values and thus prevents any hidden content changes [3].

Moturu and Liu propose a framework to help one's by identifying relevant features for assessment trust in a Wikipedia's article. According to the authors, this can change the way information from social media is perceived and used [35]. They differ between quality and trust and define credibility as well as believe exactly according to [3] that content revision history and author information are essential for assessing trust. In this context, quality means a representation of an inherent characteristic or essential character and can be defined by content related predictors. They also describe credibility as the quality of

[3] http://genius.com
[4] https://de.wikipedia.org
[5] https://twitter.com
[6] http://stackoverflow.com

inspiring belief. The definitions of trust are similar: trust can be defined as the perception of the trustee about the degree to which the trustee would meet an expectation about a risky transaction.

Wikipedia contains countless articles whose true sources remain largely unknown. Lucassen and Schraagen, therefore, assume that other characteristics are necessary to assess the trustworthiness of articles. They manipulated quality and subject of articles from the same site and appear experimental that the main characteristics are textual characteristics, references and images [29]. Kelton et al. propose a model of trust in digital information based on the integration of research behaviour and social sciences with research quality and human-computer interaction. It is based on that trust is essential for research in the context of information science and show how a well-developed theory can develop a new understanding of evaluation and use of information [24]. In the same spirit, Kittur et al. [26] examine the domain of a mutable wiki system as Wikipedia and deal with the exploration of how users' notions of trustworthiness can be affected by the appearance of page-related information, the so-called Surfacing Trust, from which researchers and system designers can benefit. Based on the trust definition of Fogg and Tseng [15], Kittur et al. experimentally suggest that upcoming information that is relevant for the stability of the article and the behavior patterns of the editor influence the confidence of the user.

2.4.3 Twitter

Twitter is somewhat similar to our case study (Genius), especially the limited length of the message (tweet) and the likeness of the nature of their user-generated content. That is why we introduce a selection of relevant work on it.

According to Zhao et al. existing approaches for trustworthiness assessment in Twitter can be classified into feature-based trust ranking and social graph based trust ranking. The authors propose a trustworthiness estimation method that is based on two mechanisms: 1) topic-focused similarity-based trust evaluation that relies on keywords for specifying the tweet's domain and the tweet's event. For each, a weight sum is calculated based on the number of the keywords a tweet contains considering that keywords of a domain similar to. This applies to keywords of an event, too. 2) Trust propagation that is based on semantic or contextual relationship, authorship, or friendship [50].

In contrast, Basharat and Ahmad consider supervised machine learning in their work and introduce a framework based on the combination of features related to user (e.g. Friends_count, Followers, Favorite_count etc.) and tweet (Character_count, Word_count, Special_symbols etc.) on order to determine the reliability/trust of information provided. The authors used tweets related to well known events took place in 2016 and applied machine learning approaches (e.g. Naïve Bayes, Linear Regression etc.), which performed differently in determining trusted tweets spread in Twitter [6].

Bodnar et al. propose a trust model to investigate context-independent information on social media (Twitter). The model presented is based on the metadata of the users. The authors used a combination of natural language processing and machine learning algorithms to generate trust profiles to support event recognition. This is achieved by analyzing the differences between users who have predicted an event as (not) real and who have (not) accepted it [10].

Adali et al. investigate the relationship who-trusts-whom and introduce algorithms to calculate user trust on Twitter. This relationship is specified by a set of communication 3-tuples: Sender, receiver and time as input, and as output is a behavioural trust graph. The senders and receivers form the nodes and the behavioural trusts the weighted edges of this graph. The authors stated that users who trust each other are likely to have conversations and improve their relationship of trust. The number and duration of a conversation between two nodes and the development of such a conversation then measure trust by a third node (propagation) [2].

Ravikumar et al. propose a graph-based method consisting of users, tweets and web pages to rank tweets in terms of trustworthiness and relevance. The basic idea is that trustworthiness and popularity benefit from the implicit relationships (advertise) between tweets. However, the authors see a criticism of how Twitter recognizes the importance of popularity, which is evaluated by the number of re-tweets, and argue that, on the one hand, semantically similar tweets could occur as re-tweets, and on the other, re-tweeting is performed by users without verifying the content, and thus it can not indicate trust. To overcome these problems, the proposed method takes into account popularity and trustworthiness of tweets on the basis of their content by checking the consistency between a tweet and the page rank of the web page referenced in the tweet and the implicit links between the re-tweeters based on the follower-followee relationship [40].

2.4.4 Community Question-Answering (CQA)

In the research context of the question-answering community e.g. Stackoverflow, Yahoo! Answers etc., the term trust is rarely used. Instead, terms that are indicators of trust are examined e.g quality, credibility, reputation etc. In this way, qualitative and accepted answers are ranked high and evaluated by the users.

Neshati proposes a framework for recognizing high-quality content on Stackoverflow. This framework can simultaneously predict the quality of a question and the related answers shortly after submission. The author observed two patterns for this early detection: 1) Accepted answer effect, which indicates that the probability of obtaining an accepted answer to a high-quality question is higher and vice versa, and 2) answer competition effect, which indicates that the number of high-quality answers to a particular question is low [37].

Ginsca and Popescu suggest automatic method for quality assessment in Stackoverflow. The authors investigate the metadata of a user profile in relation to the quality of his contributions and observe that user characteristics are good indicators of quality detection of contributions. For example, users with full name as a user name are associated with high quality contributions; the same applied for users how use a personal avatar. Age and links to external platforms play a key role; users that are more comfortable in revealing their age tend to provide more valuable answers, this applies for users who provide links towards external online platforms as Twitter and Facebook [17].

Yao et al. propose algorithms for early detection of high-quality CQA. These algorithms support discovering a high-impact question and identify a useful answer that would earn positive feedback from users. The authors viewed the post quality from the perspective of the voting outcome and observed a strong correlation between voting score of an answer and that of its question [49].

Blooma et al. investigate the predictors of high-quality answers in a community-driven question answering service (Yahoo! Answers). The predictors used are organized in social- and content-features. The social-features are related to the community aspects and are selected from user interactions and feedbacks. Whereas the content features of answers to predict their quality e.g positive votes, completeness, accuracy, high frequency words, answer length etc [9].

Ponzanelli et al. introduce an approach to improve the review process used by Stackoverflow. They classified a public data dump of Stackoverflow into four classes, which are 1) very good questions that are associated with accepted answer, neither closed nor deleted and score > 7. 2) Good questions similar to very good questions, but with an interval score [1,6]. 3) Bad questions that are neither closed, nor deleted and with score < 0. 4) Very bad questions that are closed or deleted. Their approach applies metrics (e.g. readability metrics such as Automated Reading Index (ARI) and popularity metrics such as votes) to quantify the quality of posts. [39].

2.4.5 Recommendation Systems

Trust investigations play a key role in the area of Semantic Web as well, where trust is about the source of information. Artz and Gil introduce that humans make judgment about trust (source validation), especially, in case of multiple information sources. This will be not sufficient, since content will be represented in ontologies and axioms and agents and automated reasoners will be acting as consumer of information in addition to humans [5].

Su et al. propose a priority-based trust model (PBTrust model) that value the trustworthiness of a service provider based on its historical performance. Additionally to the known two types of agents (consumer and provider agents), the authors define the "referees" agents. This type is based on a "reference" generated by a consumer agent about the quality of service to other consumer agents. This is the main idea for the model calculation; namely, the consideration of a third party (referee) evaluation [46].

Yang et al. suggest a hybrid method (TrustMF) to improve the performance of collaborative filtering recommendations and the social trust network among users. The method combines both a trustor paradigm and trustee paradigm and map them into two spaces: 1) "truster space", where it is described that "to trust by reading ratings or review", and 2) "trustee space", where the case is that "to be trusted by generating ratings or reviews" [48].

Mui and Halberstadt propose computational model that can be integrated in a real system to value agents' trust and reputation. The model considers, additionally to reputation, reciprocity relationship, which is defined as "mutual exchange of deeds (such as favor or revenge)" and can be viewed either as a norm earned by agents in a society, or as a binary variable between two agents [36].

Xing et al. introduce in [46] a priority-based trust model for open and dynamic environments. It is based on a module that calculates reputation values taking into account the service provider's experience, the similarity of priorities between the reference and the service requested, the suitability of a service provider's potential performance and the temporal relevance to reduce the contribution of outdated references. The module uses a third-party reference to derive the reputation and suitability of a provider whose trustworthiness seems to be unclear, despite clarification of the provider's experience calculation.

In addition, old references are not necessarily of low quality or trustworthiness; these conclusions should be based on the user experience or reputation that result from interactions with such amounts.

2.4.6 Semantic Web discussion

In the Semantic Web, content is a series of statements that are not so easily represented for the user and therefore cannot be judged by appearance. Because the underlying philosophy of the Semantic Web is that the computer makes distributed statements about the same resource and aggregates them. According to Golbeck and Hendler, everyone is allowed to make a statement without demanding accuracy or truthfulness on the hypertext web. Human decisions are based on the appearance of the page and the source of the information [20]. Although it is relatively easy to maneuver information about the source, you can generate some of such information. Many research has focused on resource authentication, including working on digital signatures and public keys to define the word "trust" on the Semantic Web. This creates trust in the source or the author of a statement, which is very important, but trust in this sense ignores the question of credibility. Confirming the source of a statement has no explicit effect on the quality of the statement [19]. In the same sense, Richardson et al. discuss the philosophy behind the Semantic Web, which is the same as behind the World Wide Web - anyone can be a producer or a consumer of information from someone else. Most Semantic Web research has focused on defining standards for the communication of facts, rules and ontologies, etc. XML, RDF, RDF Schema, OWL and others form a necessary basis for building the Semantic Web. The question remains, however, how trustworthy each source of information is. One solution would be to make all information on the Semantic Web consistent and of high quality. But due to its sheer size and its variety of sources, that will be almost impossible. Instead, we should develop methods that work under the assumption that the information will be of very different quality [41].

We assert that the open and dynamic WWW initiated by Tim Berners Lee is not intended or even suitable for this dimension of today's development. Basic structural changes must be made to meet new requirements such as automatic verification of the information source and delivery of suitable and up-to-date information in context -known as sensemaking-. But one can also enrich newly generated information with metadata to have a better starting position for the problem. Old and already existing information in the WWW could be modified as far as possible or simply let it become outdated in order to take it out of circulation later by archiving.

Bibliography

[1] A. Abdul-Rahman and S. Hailes, *Supporting trust in virtual communities*, in System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on, IEEE, 2000, pp. 9–pp.

[2] S. Adali, R. Escriva, M. K. Goldberg, M. Hayvanovych, M. Magdon-Ismail, B. K. Szymanski, W. A. Wallace, and G. Williams, *Measuring behavioral trust in social networks*, in Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on, IEEE, 2010, pp. 150–152.

[3] B. T. Adler, K. Chatterjee, L. De Alfaro, M. Faella, I. Pye, and V. Raman, *Assigning trust to wikipedia content*, in Proceedings of the 4th International Symposium on Wikis, ACM, 2008, p. 26.

[4] J. Al Qundus, *Technical analysis of the social media platform genius*, tech. rep., Freie Universität Berlin, 03 2018.

[5] D. Artz and Y. Gil, *A survey of trust in computer science and the semantic web*, Web Semantics: Science, Services and Agents on the World Wide Web, 5 (2007), pp. 58–71.

[6] S. Basharat and M. Ahmad, *Inferring trust from message features using linear regression and support vector machines*, in International Conference on Next Generation Computing Technologies, Springer, 2017, pp. 577–598.

[7] J. Benamati, M. A. Serva, and M. A. Fuller, *Are trust and distrust distinct constructs? an empirical study of the effects of trust and distrust among online banking users*, in System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on, vol. 6, IEEE, 2006, pp. 121b–121b.

[8] I. Benbasat and W. Wang, *Trust in and adoption of online recommendation agents*, Journal of the association for information systems, 6 (2005), p. 4.

[9] M. J. Blooma, D. Hoe-Lian Goh, and A. Yeow-Kuan Chua, *Predictors of high-quality answers*, Online Information Review, 36 (2012), pp. 383–400.

[10] T. Bodnar, C. Tucker, K. Hopkinson, and S. G. Bilén, *Increasing the veracity of event detection on social media networks through user trust modeling*, in Big Data (Big Data), 2014 IEEE International Conference on, IEEE, 2014, pp. 636–643.

[11] C. Castelfranchi and R. Falcone, *Trust theory: A socio-cognitive and computational model*, vol. 18, John Wiley & Sons, 2010.

[12] J.-H. Cho, K. Chan, and S. Adali, *A survey on trust modeling*, ACM Computing Surveys (CSUR), 48 (2015), p. 28.

[13] P. Cofta, *Trust, complexity and control: confidence in a convergent world*, John Wiley & Sons, 2007.

[14] M. Deutsch, *Cooperation and trust: Some theoretical notes.*, (1962).

[15] B. Fogg and H. Tseng, *The elements of computer credibility*, in Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM, 1999, pp. 80–87.

[16] D. Gambetta et al., *Can we trust trust*, Trust: Making and breaking cooperative relations, 13 (2000), pp. 213–237.

[17] A. L. Ginsca and A. Popescu, *User profiling for answer quality assessment in q&a communities*, in Proceedings of the 2013 workshop on Data-driven user behavioral modelling and mining from social media, ACM, 2013, pp. 25–28.

[18] J. Golbeck, *Weaving a web of trust*, Science, 321 (2008), pp. 1640–1641.

[19] J. Golbeck and J. Hendler, *Accuracy of metrics for inferring trust and reputation in semantic web-based social networks*, in International Conference on Knowledge Engineering and Knowledge Management, Springer, 2004, pp. 116–131.

[20] J. Golbeck and J. Hendler, *Inferring reputation on the semantic web*, in Proceedings of the 13th International World Wide Web Conference, vol. 316, 2004.

[21] T. Grandison and M. Sloman, *A survey of trust in internet applications*, IEEE Communications Surveys & Tutorials, 3 (2000), pp. 2–16.

[22] N. Griffiths, *A fuzzy approach to reasoning with trust, distrust and insufficient trust*, in International Workshop on Cooperative Information Agents, Springer, 2006, pp. 360–374.

[23] R. He, J. Niu, and G. Zhang, *Cbtm: A trust model with uncertainty quantification and reasoning for pervasive computing*, in International Symposium on Parallel and Distributed Processing and Applications, Springer, 2005, pp. 541–552.

[24] K. Kelton, K. R. Fleischmann, and W. A. Wallace, *Trust in digital information*, Journal of the American Society for Information Science and Technology, 59 (2008), pp. 363–374.

[25] Y. A. Kim and M. A. Ahmad, *Trust, distrust and lack of confidence of users in online social media-sharing communities*, Knowledge-Based Systems, 37 (2013), pp. 438–450.

[26] A. Kittur, B. Suh, and E. H. Chi, *Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia*, in Proceedings of the 2008 ACM conference on Computer supported cooperative work, ACM, 2008, pp. 477–480.

[27] R. Kraut, M. L. Maher, J. Olson, T. W. Malone, P. Pirolli, and J. C. Thomas, *Scientific foundations: A case for technology-mediated social-participation theory*, (2010).

[28] S. C. Lu, D. T. Kong, D. L. Ferrin, and K. T. Dirks, *What are the determinants of interpersonal trust in dyadic negotiations? meta-analytic evidence and implications for future research*, Journal of Trust Research, 7 (2017), pp. 22–50.

[29] T. Lucassen and J. M. Schraagen, *Trust in wikipedia: how users trust information from an unknown source*, in Proceedings of the 4th workshop on Information credibility, ACM, 2010, pp. 19–26.

[30] N. Luhmann, *Trust and power/two works by niklas luhmann; with introduction by gianfranco poggi*, 1979.

[31] S. P. Marsh, *Formalising trust as a computational concept*, (1994).

[32] R. C. Mayer, J. H. Davis, and F. D. Schoorman, *An integrative model of organizational trust*, Academy of management review, 20 (1995), pp. 709–734.

[33] B. McGuinness and A. Leggatt, *Information trust and distrust in a sensemaking task*, in Command and Control Research and Technology Symposium, 2006.

[34] D. H. McKnight and N. L. Chervany, *The meanings of trust*, (1996).

[35] S. T. Moturu and H. Liu, *Evaluating the trustworthiness of wikipedia articles through quality and credibility*, in Proceedings of the 5th International Symposium on Wikis and Open Collaboration, ACM, 2009, p. 28.

[36] L. Mui, M. Mohtashemi, and A. Halberstadt, *A computational model of trust and reputation*, in System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on, IEEE, 2002, pp. 2431–2439.

[37] M. Neshati, *On early detection of high voted q&a on stack overflow*, Information Processing & Management, 53 (2017), pp. 780–798.

[38] G. Noh, H. Oh, and J. Lee, *Power users are not always powerful: the effect of social trust clusters in recommender systems*, Information Sciences, (2018).

[39] L. Ponzanelli, A. Mocci, A. Bacchelli, M. Lanza, and D. Fullerton, *Improving low quality stack overflow post detection*, in Software Maintenance and Evolution (ICSME), 2014 IEEE International Conference on, IEEE, 2014, pp. 541–544.

[40] S. Ravikumar, R. Balakrishnan, and S. Kambhampati, *Ranking tweets considering trust and relevance*, in Proceedings of the Ninth International Workshop on Information Integration on the Web, ACM, 2012, p. 4.

[41] M. Richardson, R. Agrawal, and P. Domingos, *Trust management for the semantic web*, in International semantic Web conference, Springer, 2003, pp. 351–368.

[42] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer, *Not so different after all: A cross-discipline view of trust*, Academy of management review, 23 (1998), pp. 393–404.

[43] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, *Fake news detection on social media: A data mining perspective*, ACM SIGKDD Explorations Newsletter, 19 (2017), pp. 22–36.

[44] H. Singal and S. Kohli, *Escalation of trust analysis in web*, in Proceedings of the 12th ACM International Conference on Computing Frontiers, ACM, 2015, p. 56.

[45] B. Solhaug, D. Elgesem, and K. Stolen, *Why trust is not proportional to risk*, in Availability, Reliability and Security, 2007. ARES 2007. The Second International Conference on, IEEE, 2007, pp. 11–18.

[46] X. Su, M. Zhang, Y. Mu, and Q. Bai, *A robust trust model for service-oriented systems*, Journal of Computer and System Sciences, 79 (2013), pp. 596–608.

[47] T. Wöhner and R. Peters, *Assessing the quality of wikipedia articles with lifecycle based metrics*, in Proceedings of the 5th International Symposium on Wikis and Open Collaboration, ACM, 2009, p. 16.

[48] B. Yang, Y. Lei, J. Liu, and W. Li, *Social collaborative filtering by trust*, IEEE transactions on pattern analysis and machine intelligence, 39 (2017), pp. 1633–1647.

[49] Y. Yao, H. Tong, T. Xie, L. Akoglu, F. Xu, and J. Lu, *Detecting high-quality posts in community question answering sites*, Information Sciences, 302 (2015), pp. 70–82.

[50] L. Zhao, T. Hua, C.-T. Lu, and R. Chen, *A topic-focused trust model for twitter*, Computer Communications, 76 (2016), pp. 1–11.

Chapter 3

Related Work

Eric Raymond says "Every good work of software starts by scratching a developer's personal itch". Our background in developing user-oriented applications means that we focus more on the quality of the application than on quantity. This determines how the user can be encouraged to deal with the information provided about the application. We could not find a satisfactory answer to this question in the literature; we are also upset by the numerous definitions of trust. Nevertheless, we found interesting works that motivated us to study and research this topic. This section provides an overview of these relevant works; however, due to the nature of this thesis as thesis-by-publication, chapters introduce their own relevant works in more detail.

3.1 Trust Dimensions

There are a number of researches and in particular in the domain of Wikipedia, which evaluated content exclusively on the basis of quality [3, 23, 30, 70]. Different metrics - article length, the total number of edits, unique editors, number of authors, internal/external linking, etc. - are examined to recognize, measure and/or predict the quality of content [55, 65]. Models, algorithms and different approaches have been developed on the topic of trust, but many of these studies consider or deal with quality and trust as equal. While other research projects make this distinction [19, 52]. We agree with this distinction, however we are also interested in these works, which do not mix quality, credibility and trust.

We know now the definition of trust (see Chapter 2), the entities (trustor and information by trustee), the preconditions (risk/vulnerability and independence) and the situation (acting with provided information). Such characteristics build the trust model proposed in this thesis. We call them dimensions that are stability, credibility and quality. In the following we introduce the related works that address various metrics to assess trust and present earlier studies on trust modeling.

3.1.1 Stability

Kittur et al. suggest that stability of the Wikipedia article has significant impact on user's trust and define stability as "measured by changed words in the last day, month, and year" [28]. Denning et al. suggest the article stability, which is illustrated by the number of changes since the last viewing, as a factor of risks involved in use Wikipedia [12] and examining vandalism [9].

Dondio el at. suggest in [13] a Wikipedia Trust Calculator (WTC) that consists of Data Retrieval module contains the needed data of an article, Factors Calculator module calculates and merges the trust factors into the macro-areas defined and Trust Evaluator module calculate a numeric trust value and judge it in a natural language explanation using constraints provided by a Logic Conditions module. The authors suggest the function:

$$N(t) : t \rightarrow \tau$$

that returns the number of edits done at time t, which is used by the stability function

$$E(t) = \sum_{t}^{p} N(t)$$

that calculates the number of edits done from a given time t to the present time p. Meanwhile, stability is defined as "only active and articles with good text can be considered stable". This work considers a part of edits' contributors as "n% top active users", which is calculated by the function they called "Users' Distribution/Leadership

$$P(n) = \sum_{U_a} E(u)$$

with U_a the set of n% top active users and "$E(u)$ as the number of edits for user u for a specific article". Part of our work takes over the stability function presented in this paper and is inspired by the idea to consider the n-top-active user.

3.1.2 Credibility

Credibility has been widely investigated as a concept related trustworthiness. [16] define credibility as the source or message believability and see a relationship between credibility several concepts, including trust and quality. Credibility is basically build on rating and is assessed based on that so-called trust circle consisting of a group of known people (e.g. friends, family members etc.) or people, who are known as highly credible [29, 45]. The members of such trust circle are called elite members, who tend to be more (double as much) trusted and whose reviews are considered as more significant [58, 61]. Meanwhile, Pranata and Susilo examine the trustworthiness of this elite in giving ratings. The results indicate that the ratings of popular users is not, however, the only decisive factor with a view to evaluating the trustworthiness, other factors, such as the total number of ratings and rating of all users available, must be considered as well [46].

Credibility literature has concentrated mainly on text information that is presented on websites. The online information assessment typically comprises five criteria that users should apply, including verification of the accuracy, authority, objectivity, currency, and coverage of the information and/or its source [36]. Accuracy is the degree to which a website or other source is error-free and whether the information is capable of being verified offline. A website's authority can be measured by who wrote the information, what the author's references and qualifications are, and whether the website is recommended by a trusted third party. Objectivity refers to identifying the author's intent to provide the information and determining whether the information is fact or opinion, including understanding whether there is a commercial intent or conflict of interest, and the nature of the relationships between linked information sources. The currency refers to the timeliness of the information, and coverage to the completeness or depth of the information presented [37].

Castillo et al. [7] assess credibility of tweets by classifying trending topics as credible or not credible using features relying on users' tweets and re-tweetts and externals sources, while, [21] apply for assessing credibility of events on twitter an approach called BasicCA which performs PageRank-like iterations. It relies on four feature-sets: 1) User features (i.e. number of friends, followers and status updates); 2) Tweet Features (i.e. formally written, external URLs, number of pronouns, event-relation, completeness and tweet-event sentiment); 3) Event features (e.g. number of tweets and re-tweett related, number of distinct URLs, users, hashtags etc.); and 4) Computation of features (the previous three feature-sets are computed once).

3.1.3 Quality

A large body of research investigates the correlation between credibility, quality and trust (recently e.g. [8]), however, the focus is on the concept of quality related to trust. Meanwhile, a lot of work has been done to describe quality according to several criteria. Although we believe that the goal differs from trust in many cases, it is not easy to distinguish and possibly separate this strong relationship between quality and trust.

Information quality is considered as "fitness for use" [57] or as "user satisfaction" [11]. Naumann [39] enumerates criteria for assessing the quality of information: Content, technical, intellectual and instantiate (data)-related. While the criteria by Rieh et al. [51] are: Source, content, format, presentation, currency, accuracy and speed of loading. Rieh et al. in another work [50] refined such criteria to: Characteristics of information (e.g. title, graphics, functionality etc.); characteristics of sources (e.g. organization, collaboration, authorship etc.); knowledge referring to user's aspects (e.g. experience, personal background etc.); information context; search-ranking and preconception (e.g. biased opinion). Tate [56] introduces criteria: Authority (e.g. author experience), accuracy (e.g. extensive know-how), objectivity (e.g. unbiased opinion), currency (e.g. up-to-date) and coverage (e.g. depth in meaning). Bizer and Cyganiak [2] propose quality assessment classes: Content-based, context-based and rating-based. According to Brando and Bucher [4], quality is connected with the user's trust in the content (a subjective term), which leads to a connection between the content-quality and the authority of the provider. Quality measurement can be based on metrics, such as the number of edits and the number of unique contributors [54, 66].

To evaluate the quality of articles in Wikipedia collaborative authoring, Hu et al. [23] propose three models for measuring article quality based on the interaction data between the articles and their contributors from the article editing history, that is, "the quality contribution to article content and the authority of contributors". 1) The Basic-Model is based on the interdependence between article quality and author authority. Accordingly, Q_i, the quality of each article a_i and A_j, the authority of each user u_j are calculated by

$$Q_j = \sum_j c_{ij} \times A_j$$

and

$$A_j = \sum_i c_{ij} \times Q_i$$

with c_{ij} denotes the amount of words u_j authored in a_i. 2) The PeerReView model introduces review behaviour into the measurement of article quality, which is assessed using the equations

$$q_{ik} = \sum_{\alpha \cup \beta} A_j$$

and

$$A_j = \sum_{\alpha \cup \beta} q_{ik}$$

with

$$\alpha = w_{ik} \xleftarrow{\mathrm{A}} u_j$$

word w_{ik} is authored by a user u_j,

$$\beta = w_{ik} \xleftarrow{\mathrm{R}} u_j$$

word w_{ik} is reviewed by a user u_j, and q_{ik} is a word quality that is summed up into the quality of the article

$$(Q_i = \sum_k q_{ik})$$

3) The ProbReview model is an extension of the PeerReview model by a partial reviewership of contributions while they edit different parts of the articles. This is calculated using the equations:

$$q_{ik} = \sum_j f(w_{ik}, u_j) A_j$$

and

$$A_j = \sum_{i,k} f(w_{ik}, u_j) q_{ik}$$

, where $f(w_{ik}, u_j)$ equals 1 if the word is authored or equals the probability of the word reviewed

$$(Prob(w_{ik} \xleftarrow{\mathrm{R}} u_j))$$

Gamble and Goble investigated the essence of "quality" by using three predominant characteristics of scientific data sets: 1) that data quality is generally objectively defined; 2) the provenance and origin plays a well understood role in its production; and 3) "fitness for use" is a definition of utility rather than quality or trust, the quality and trustworthiness of the data and the entities that have generated it determining its utility. A scientist's decision to use data is influenced by whether this is the case: "1) good when compared against norms and standards; 2) likely to be good given its provenance; and 3) a good fit to current needs". The authors pose an approach for assessment by modeling the causal relationships between the dimensions 1) quality that is defined as "a function of the artifact or process assessed against a quality standard independent of the consumer to provide a specific, objective measure of quality e.g. accuracy."; 2) trust that is defined as "a function of the artifact, producer, provider or process (along with perhaps the consumer) that can be assessed independent of a standard to provide a general prediction of quality e.g. reputation"; and 3) utility that is defined as "a function of the artifact and consumer to assess whether data are fit for purpose and meet the users subjective needs e.g. relevance" [18]. The authors provide an analysis of the dimensions of information quality in the literature in Table 3.1, from which we take the relevant part and comment on each dimension according to our research.

In contrast to the original table, Table 3.1 omits some dimensions: Currency, Objectivity, Accessibility/Availability, Relevance, Timeliness, Conciseness, Interpretability, Security, Applicability/Appropriateness and Cost. These are not directly taken into account in our approach. Due to the complexity degree of measuring such metrics, it is not an easy task to cover them satisfactory. In addition, the nature of such matrices makes them semantically very close to each other and the distinction between them is also a difficult subject. For example, for measuring Objectivity satisfactory, deep analysis of contributors is required, which could be a field of research, because we are dealing with subjective metrics that contain opinions. Another example: What is the difference between Currency and Timeliness or Relevance and Appropriateness? In which cases can we consider a contribution as Currency? A new contribution (high-Currency?) may be of interest to the user for the moment and activities could be generated on, however, an old contribution (low-Currency?) that was classified as not current could be again highlighted (high-Currency, isn't it?) based on a current event related to.

This thesis proposes a trust model that adopts the terms 1) stability proposed in [13] with the difference of considering the growth of the contributions rating instead of the difference of the article version history. 2) Credibility defined in [36, 37, 46] and called authority/reputation in [62]. 3) Quality introduced in [18, 62], which influences trustworthiness and can be estimated by user and content (goodness) [17]. However, till now there is no scientific evidence that ratings provided by users are always credible and trustworthy [46], but we prone to trust interactions of most popular users [32, 46, 58].

3.2 Trust Computation

Assessing trust in online information is not an easy task in research. The observer must analyze the characteristics of trust. Before doing that, one first has to determine the situation and the preconditions of trust. Who or what forms of entities are involved in the trust at all, what is needed to begin the trust process etc. At this point, if not already gone astray before, one is overturned by countless factors. It quickly becomes clear that a definition of trust is needed in order to know what is being talked about at all.

3.2.1 Trust based on Metrics

Singal and Kohli [53] measure trust on the basis of web metrics and deal with the question of which web content is trustworthy and to what extent? This should help users to decide whether or not they can trust the information they are viewing. They collect data from medical websites because it is believed to involve information that is specific and sensitive. Based on identified Key Performance Indicators (KPIs), which are a set of web metrics, i.e. Average Time on Website (ATW), Pages/Visit (PPV), Average Daily Visits (ADV) and Bounce Rate (BR), they provide a solution where it is suggested that pages/visits, average

Table 3.1: Extended table of analysis of information quality dimensions.
Source: [18]

Quality Dimension	**Claimed as indicator of**	**Our Comment to**
Completeness	Quality	Disagree; it indicates Credibility, since the users only can decide whether the information is complete or not, however, in case it's not complete, Stability is then applied; users may try to complete the information provided collaboratively.
Accuracy	Quality	Agree; mostely the elite (user with experience) generate accuracy-content
Consistency	Quality	Agree; (see Accuracy)
Reputation	Trust	Disagree; it indicates Quality; (see Accuracy)
Correctness	Quality	Agree; Quality; (see Accuracy)
Value-Added	Utility	Disagree; it indicates Quality; (see Accuracy)
Authority	Trust	Disagree; it indicates Quality; (see Accuracy)
Freedom form Errors	Quality	Agree; it indicates Quality; (see Accuracy)
Understandability	Utility	Disagree; it indicates Credibility, since the users only can decide whether the information is understandable or not.
Believablity	Trust	Disagree; it indicates Credibility; (see Understandability)
Usefullness	Utility	Disagree; it indicates Credibility; (see Understandability)
Usability	Utility	Disagree; it indicates Credibility; (see Understandability)
Recommendation	Trust	Disagree; it indicates Credibility; (see Understandability)
Amount of Data	Utility	Disagree; it indicates Stability; collaboratively work, the produces editing activities.
Stability/Volatility	Quality	Disagree; it indicates Stability; (see Amount of Data)
Trustworthiness	Trust	Agree; Trustworthiness is the property of a trustee resulted by Trust as a process or activity of a trustor.

time spent on a website and average daily visits have a positive impact on the trust level, while the bounce rate is negatively linked to the trust level. In addition, an equation has been proposed to help classify trust with an accuracy of 87.21%:

$$trust = -0.1ATW + 0.33BR + 0.45ADV + 0.07PPV$$

The factors contained were estimated using a survey. This equation is intended to adequately complement the trust evaluation of websites by summarizing the behaviour of users who have already used them and thus passing on their experiences directly or indirectly.

Nurse et al. in [42] provide an approach consisting of policies i.e. provenance, quality and other factors as competition, reputation, recency, etc., which are combined in their further work [41] for assessing trustworthiness to support users of social-media information in making informed decisions, in light of risk. The applied factors are an extension of metrics introduced in the work of Gil and Ratnakar, who developed a tool that enables users to express their trust in a source and the statements made by that source. The individual views are combined to form an overall evaluation of each information source. The decision whether the content of a Web resource can be trusted depends on its source [20]. Nurse et al. selected these factors based on their context (e.g. location or event and likelihood of a compromise). Each factor is given a weight that is correspondent with its importance, which is impacted by the user's own decision-making being more or less important. This approach assesses information-risk based on information infrastructure integrity (III), which is the complement of vulnerability in the information infrastructure (IIV) that is the generic measure of

$$Exposure = Threat \times Vulnerability$$

Exposure reflects the possibility that some of the information may have been damaged and is considered as the result of a motivated Threat associated with an exploitable Vulnerability. Threats are entities that commit attacks and deliberately attempt to corrupt information. In addition, the timeliness of information is considered to be the most fundamental trustworthiness factor, assuming that the closer the information is published at a given time, the more likely it is to be current and potentially accurate. To achieve a single trustworthiness value six methods are explored:
Conjugated Root-Mean-Squared

$$(1 - (\tfrac{1}{n}\sum_{i=1}^{n}(1-x_i)^2)^{\frac{1}{2}}))$$

, Arithmetic mean

$$(\tfrac{1}{n}\sum_{i=1}^{n}x_i)$$

, Geometric mean

$$((\prod_{i=1}^{n}x_i)^{\frac{1}{n}})$$

, Quadratic mean

$$((\tfrac{1}{n}\sum_{i=1}^{n}x_i)^{\frac{1}{2}})$$

, Harmonic mean

$$(n/\sum_{i=1}^{n}\frac{1}{x_i})$$

and Square-Mean-Root

$$((\frac{1}{n}\sum_{i=1}^{n}\sqrt{x_i})^2)$$

using a set of criteria and an evaluation. The authors reports that "all the means did quite well" and "all the formulae performed in line with expectations" and stated that "the choice of combination method applied should be the user's, as they may have their own perspectives or contexts.". Although description of a reference measure e.g. a threshold for their calculation is not clear, however, the method of calculation conducted (several kind of means) is of interest.

Gil and Artz investigate content trust in web and argue that the decision to trust the information provided by an entity is a complex process that is affected by many factors, and only one of them is the degree of trust. The authors have identified several factors that affect user trust in web information source content: 1) Topic considered, 2) Context and criticality of the need for information, 3) Popularity of the resource, 4) Recognized authority of associations, 5) Reputation by direct experience, 6) Referrals by other users, 7) Association by other trusted resources (e.g. citations), 8) Provenance and pedigree, 9) Expertise of the user, 10) Perceived bias of source, 11) Perceived incentive in providing accurate information, 12) Absence of other alternative resources, 13) Agreement with other resources, 14) Precise and specific content, 15) Likelihood of content being correct given what is known, 16) Time of creation of the content, 17) Professional appearance, 18) Likelihood of deceptive behavior and 19) Recency of factors under consideration. Some of these factors are correlated e.g. Topics and Criticality as well as Direct Experience and Recommendations, others are heuristic in nature e.g. Incentive and Likelihood. The authors suggest that the main factors are: authority (referring to Factor 4), related resources (referred to Factor 7), origin (referred to Factor 8) and bias (referred to Factors 10, 11 and 18). These are determined by examining the associations of the resource. They pose that associations are fundamental in determining content trust for any resource [19].

Moreover, the assertion that the provenance is the most important factor for assessing trust [10, 47, 68]. In the same sense, trust is investigated in three domains: Provenance, Quality and Infrastructure Integrity by Nurse et al. [40]. The authors consider a number of sub-factors for measuring users' trust in a piece of open information (e.g. Twitter, Facebook or a blog) and conduct a visualization experiment using a framework in which participants receive radar graphs and are asked to evaluate each graph within 10 seconds with a trust level between 0 and 100. The sub-factors of trust examined are: "Competence (Cm), the level of knowledge of a person or information source; Proximity (Pr), the geographical closeness of a source to an event of interest; Popularity (Po), how well-known is a source; Recency (Re), how recent or up-to-date is information to the event of interest; and Corroboration (Cr), how well supported the information is by a variety of different sources". They found that competence was the most dominant factor, followed by corroboration, recency, proximity and popularity, as shown in the following equation:

$$Trustworthiness = -5.425 + 0.176Re + 0.405Cm + 0.235Cr + 0.127Po + 0.141Pr$$

While trust comprises three dimensions: Trust -Originator, -Purpose and -Target by Mu and Yuan [38], who aim to maintain initial trust value of Direct Trust (e.g. entity A has a trust relationship with entity B) and Recommender Trust (e.g. entity A has an indirect trust relationship with entity C via entity B that recommends entity C to entity A). This granularity is similar to [43, 67]. In the mentioned examples, entity A represents the trust originator and entity B represents the trust recommender of the trust target or entity C. While trust purpose is defined as a class with a set of properties that vary in situations "for example, an employee is trusted to deal with financial a airs below a certain amount by a company, but is possibly not trusted to be a spokesman". The authors use a four-scale set T to represent trust: T3: Expressed "great distrust" subset; T2: Expressed "a little distrust" subset; T1: Expressed "a little trust" subset; and, T0: Expressed "great trust" subset, which are not exclusive, that is, the entity belonging to a specific trust subset cannot be precisely (true or false) defined.

Figueiredo et al. [15] examines the relative quality of texts in social media. Quality describes the potential effectiveness of a feature as supporting data for Information Retrieval (IR) services and depends on several aspects, including (1) whether it contains enough content to be useful, (2) provides a good description of the content of the object, and (3) can effectively distinguish objects into different predefined categories (e.g. for tasks such as object classification and directory organization) or into relevance levels (e.g. for searching). Examples of objects have been taken from four applications (i.e. YouTube, YahooVideo, LastFM, and CiteULike) into consideration. The collected examples contain the contents of four text characteristics, namely TITLE, TAGS, DESCRIPTION and COMMENTS, to assess the relative quality of the various text features that often occur in such different popular Web 2.0 applications. In order to extract indications for the quality, the data collected were characterized in terms of use, quantity and semantics of the content, descriptive and discriminatory force, as well as indications for content and diversity of information about characteristics associated with the same object. The results are: 1) all four features except TITLE are missing (e.g.B, without content) in a non-negligible proportion of the collected objects, especially in applications that allow the joint creation and editing of the feature's content; 2) collaborative features, if any, tend to carry larger amounts of content than non-collaborative features; 3) a significant amount of non-existent terms concern all four features in all four applications; 4) The TITLE and TAGS features generally have a higher descriptive and discriminatory effect, followed by DESCRIPTION and COMMENT; and 5) an object is associated with different parts of content and information of its features.

Zeng et al. have developed a revision history-based trust model for calculating and tracking the trustworthiness of documents in collaborative environments. The revision history-based trust model is based on the hypothesis that revision information has the potential to calculate the trustworthiness of an article: The trustworthiness of the revised version depends on the trustworthiness of the previous version, the author of the last revision and the amount of contents of the last revision. According to the authors, there are various aspects of the investigation that could be carried out: Article trust that refers to a version of an article; Fragment Trust that refers to a fragment of an author in a version of an article; Author trust, which refers to an author. The revision trust model proposed focuses on article trust and the trust of an original article depends on its author. For this purpose, the authors considered three manually classified groups of articles in Wikipedia: *featured articles*, these were checked by the Wikipedia community for style, prose, completeness, accuracy and neutrality and considered as trustworthy; *clean-up articles* that are considered untrustworthy by the Wikipedia community; and the rest as *normal articles*. Dynamic Bayesian networks and the Beta Distribution, which as a combination is usually applied to investigate trust in the context of Web Service networks [60, 64], however, here they are used to approximate author trust by the author's editing permissions, which refer to the author role as administrator, registered author, anonymous author or blocked author. The resulting statistics of the record show that featured articles have far more revisions than clean-up articles and normal articles, while clean-up articles have the lowest proportion of administrator authors. In addition to the number of revisions and the trustworthiness of the authors, which are the most important factors in determining the trustworthiness of an article, other factors such as the amount of changes and the sequence of revisions are also important. The authors found normal articles have the least number of revisions, which shows that both the presented articles and the cleanup articles receive more attention from Wikipedia authors [70]. This work supports the hypothesis that high quality content is mostly generated by experienced users, such as users of higher roles such as administrators in this case.

West et al. discuss trust in two categories: 1) trust calculation, focusing on algorithms for calculating trust values and their relative benefits; and 2) trust usage, examining how trust values can be passed on to end users or used internally to improve application security. From these two categories the authors have found that the combination of certain properties has enormous potential for building trustworthy collaborative Web applications. These properties, which are drawn from information-quality literature and are quite qualitative in nature, are: "1) Scope: The content should be scaled accordingly. 2) Accuracy: If the content is to be factual, then it should be in the truth without misinforming or deceiving the readers. 3) Source: If the content is to be factual, then claims should be referenced and verifiable via reliable and high-quality sources. 4) Volatility: The degree to which the content is stable. 5) Cohesiveness: Quality of writing and presentation style. 6) Comprehensive: The breadth and depth of the subject examination. 7) Timeliness: The currency of the content (i.e. "Is it up-to-date?"). 8) Neutrality: The degree of bias in the presentation" [63].

3.2.2 Trust based on Text-embedded Features

Natural Language Processing (NLP)-based work is of interest to our research, as our approach follows three phases: In the first phase only the meta data of the content is considered, whereby the content itself is not visible according to the black box principle, and in the second phase the properties of the content are examined, whereas here the meta data are ignored according to the white box principle. In the third phase, the results of the first two phases are combined, as long as both results can be linked together.

NLP is used to evaluate the degree of readability using several indices such as Fog scale, Coleman-Liau index, Automated Readability Index etc. as a measure of the text complexity that indicates its quality [49, 54]. The basic approach of NLP-based work is to study features embedded in text such as the number of words per sentence, characters per word, syllable count, etc. using machine learning techniques. For instance, Potthast et al. analyzed text with a number of features such as the proportion of uppercase letters, word length and pronounce frequency [44] and others derived from a NLP toolkit as in [49].

Wang and McKeown propose a novel Web-based shallow syntactic and semantic method that is linguistically motivated. The authors use the Web as corpus by retrieving search engine results to learn the topic-specific n-tag syntax and syntactic n-gram semantic models, then, they apply a machine-learning technique over a feature-set, which is considered by three NLP categories: 1) lexical features, for which they count vulgarism frequencies (bad words) and also introduce three new lexical features: Web slang (e.g. "haha", "LOL" and "OMG"), punctuation misuse (e.g. "!!!" and "???") and comment cue words (e.g. comments to changes such as "edit revised, page changed, item cleaned up") [59]; according to [63], for the 2) syntactic feature, Wang and McKeown carry out a n-gram analysis using only Part-of-Speech (PoS) tags. This means that the probability of all PoS sequences with length n is calculated using a general or topic-specific corpus. The probabilities of new PoS sequences are then calculated during editing. Unlikely PoS sequences indicate a damaging cut. 3) The semantic analysis also uses n-gram analysis, but uses unique words instead of PoS tags. This approach is not as practical and scalable, since their approach requires crawling a substantial number (150) of web pages to be applicable [22].

3.3 Trust Modeling

Many approaches investigate user-generated content in order to model information trust. According to Castelfranchi and Falcone in [5], the most relevant approaches to trust can be divided into the logical approaches, the computational approaches and the socio-cognitive approaches. The logical approaches include models that are based on mathematical logics for investigating trust relationships. The computational approaches aim to integrate trust models implementation into automatic systems disregarding the representational framework. The socio-cognitive approaches include models that consider, on the one hand, trust on the basis of direct experience (socio statistical) and, on the other hand, on the basis of a set of factors that are trustor- and environmental-features (cognitive).

Yan et al. examine the factors that have an impact on trust in human-computer interaction, i.e. the concept of human-computer trust interaction (HCTI). The factors investigated provide positively influence on HCTI. These are combined into three core constructs to comprise a model of trust. 1) Intention to interact that includes personality, social factors, personal motivation etc. 2) Computer system trust that consists of system quality, information quality, user interface/ease of use and system trust solutions and 3) Communication trust that contains perceived privacy, perceived identity and communication context. The authors plan to verify the model using a survey. These factors are derived from theoretical studies in the literature [69]. The idea being pursued in this work inspires our work; namely, investigating related factors influencing trust in the literature, classifying and combining them into a model and verifying it using a survey.

Lucassen and Schraagen propose a model they call the 3S-model in which they take into account user judgments about trust in information. There are three user characteristics: Source experience, professional competence and information literacy. Their work is based on the assumption that these characteristics lead to different features, i.e. source semantic and surface features, of the information that is used in trust judgments by different users and in different contexts. Using Wikipedia as a case study, an experiment was conducted with the help of domain experts and novices to examine their trust judgments and evaluate the 3S-model. In Figure 3.1 the 3S-model characteristics of information and user are illustrated including examples of the features applied. The authors suggest that users can alternatively rely on their previous experience (e.g. authority, Website) with a particular source rather than actively evaluating various features (e.g. accuracy, completeness or length, references etc.) of information to judge confidence [33].

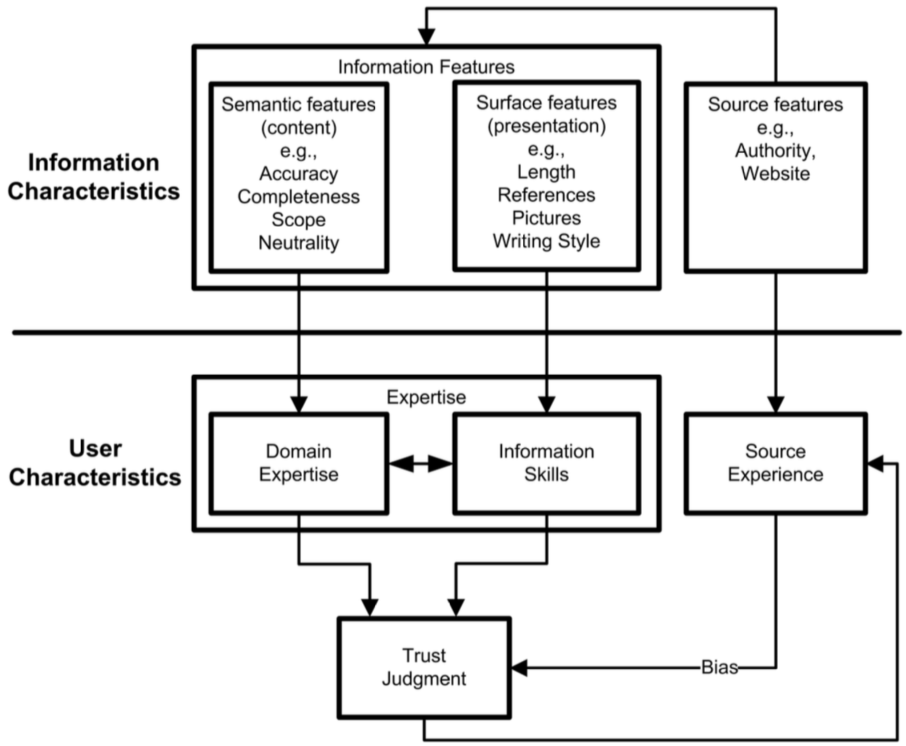


Figure 3.1: The proposed 3S-model of information trust [33]

Mayer et al. propose a model of trust that distinguishes clearly between trustor and trustee and contains components related to both. The trustor's trait is referred to the trustor's propensity to trust. Despite a trait is relative, since people differ in their inherent tendency to trust, it is proposed in the model to be stable. The characteristics of the trustee are represented in the model by the concept of trustworthiness, which consists of factors that are 1) "ability" e.g. skills or competence, 2) "benevolence" means having specific attachment to the trustor, and 3) "integrity" means following some set of principles that the trustor finds acceptable. The authors assert that these attributes of trustee are the reason why a trustor has more or less much trust for a trustee. Trust for a trustee will be a function of these attributes and together with the trustor's propensity will help to create the basis for the development of trust, but a decision to trust is not yet been made. According to trust definition, the model lacks the important aspect that is risk. You do not have to risk anything to trust, but you have to take a risk to be able to engage in trusting action. This sequence of components (factors of perceived trustworthiness, trustor's propensity and risk taking in relationship) leads to outcomes, which in turn are used as arguments for the trustee attributes to support them positively or negatively [35].

Kelton et al. extend the model proposed by Mayer et al. [35] to consider additional aspects i.e. the necessary preconditions for trust, the influence of context and social trust, and the role of trust development processes. The preconditions are "uncertainty" and "vulnerability"; i.e. when the trustor faces a risk and when there is a status of "dependence", which concerns two matters between trustor and trustee: the first has a special need to meet and the last has the potential to satisfy this need [26]. Figures 3.2 and 3.3 show a comparison of both models. Remarkable are the requirements of Kelton's et al. model, which are an extension of Mayer's et al. model.

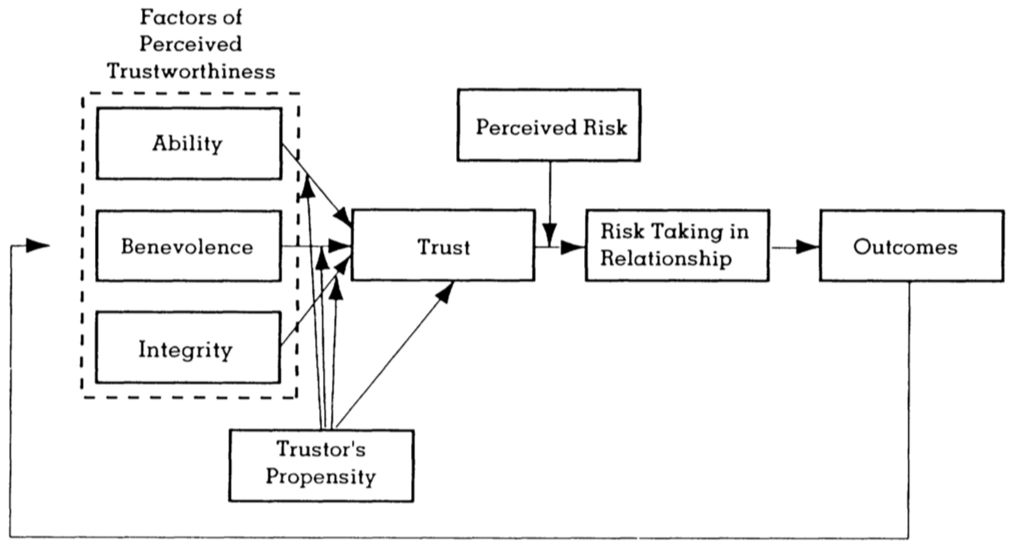


Figure 3.2: The proposed trust model by Mayer et al. [35]

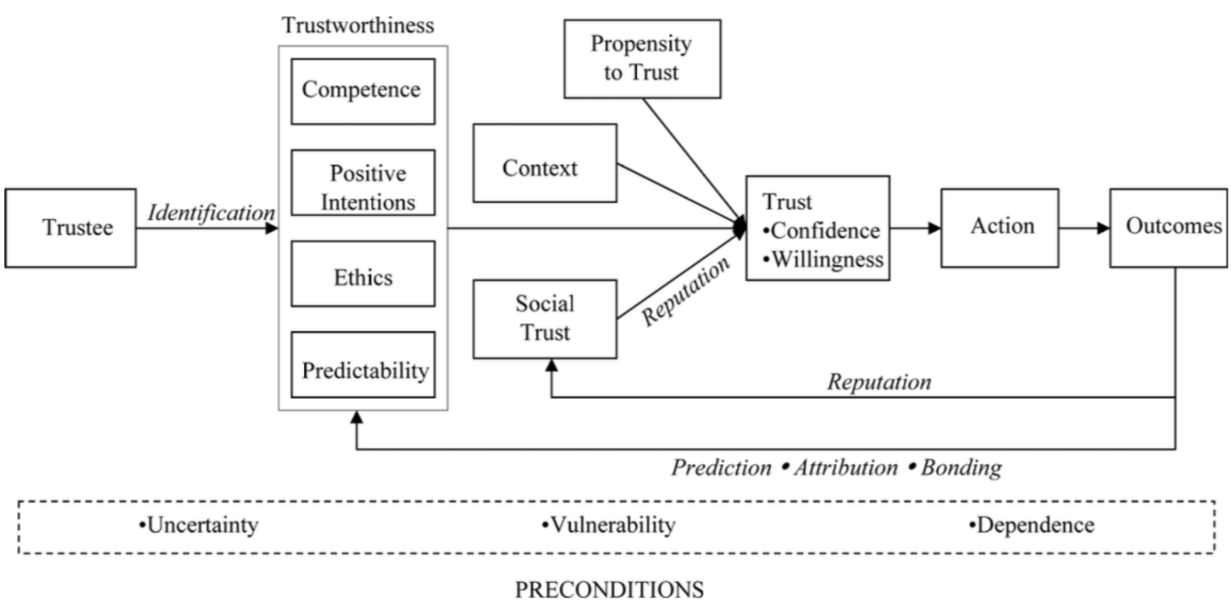


Figure 3.3: extended model of Mayer et al. by Kelton et al. [26]

Trust model of Abdul-Rahman and Hailes in [1] is based on Marsh' model [34]. It deals with sociological characteristics and trust beliefs between agents based on experience (of trustor self) and reputation (comes from recommended agent), those are combined to build trust opinion to make a decision for the interaction with the provided information. Both models are not simple to understand, as the authors tried to handle all aspects of trust. As trust is complex, this makes such models complex as well. However, the Abdul-Rahman and Hailes model has inspired our work through the trusted translator, which we adopt and modify in relation to our research.

The work of Ruth et al. designed a model of social trust for users, which is similar to the model introduced by Abdul-Rahman. The authors focus on users that are agents to determine the validation probability of a given data. The proposed model considers the categories "very trustworthy", "trustworthy", "untrustworthy" and "very untrustworthy" a user or an agent assigned to. This classification is derived from Abdul-Rahman's work and relies on first hand experience of a user; e.g. "if user A has 4 very trustworthy experience and 5 trustworthy experience with user B, the A applies the category trustworthy to B". This example considers past recommendations and the resulting experience in terms to find the semantic differences between them. This semantic difference is applied in future recommendation; if two recommendations about a user are inconsistent, the user's grades will be adjusted in favour of the worse recommendation. For example, entity A is in the category "trustworthy" based on a given recommendation. However, there is another recommendation that classifies A in the category "not trustworthy". In this case, the semantic difference requires reducing the category from entity A to the category "not trustworthy" [25] -page 183-.

Castelfranchi et al. suggest a socio-cognitive trust model using fuzzy logic with the aim to analyse the different nature of the belief sources and their trustworthiness. The model consists of four bottom-up layers: The first layer includes "beliefs sources" e.g. "who/what the source is -to evaluate-", "direct experience" e.g. "In my experience", "reputation" e.g. "A friend says that ..", "categorization" e.g. "usually doctors ..." and "reasoning" e.g. "I can infer that ...". The second layer includes internal factors i.e. ability, accessibility and harmfulness. The third layer includes external factors i.e. opportunity and danger. The last layer combines the internal and external factors into one component that influences trustfulness. The model introduces a degree of trust relied on the credibility of the trust beliefs, and its implementation allows changing the components according to the situation and agent personality [6].

Ur Rahman et al. introduce a hybrid reputation approach that consists of the combination of user expertise e.g. qualification level and research contribution, and user willingness to participate in collaborative activities reflected by a user active status over a period of time. The qualification (academic) levels are assigned numerical values, which reflect the expected experience of a user within a level. Similarly, the individual activities of a user are categorized based on their type in terms to assign numerical values [48].

Javanmardi and Videira present in [24] a Wiki Trust Model (WTM) based on Hidden Markov Model (HMM) for platforms based on wiki technology such as Wikipedia. WTM should be used for identifying vandals as well as contributors of high quality content. It can be integrated into a wiki for improving information reliability and for automatic detection of vandals to restrict their access. Figueiredo et al. propose in [14] a software architecture framework to support a trust model of members of a community and the construction of vocabulary with the focus on online communities. The calculated members' trustworthiness is composed of artifacts and parameters specified by the community itself. Trust factors can be quantitative (number of contributions) or qualitative, depending on their definition by the community. The problem of predicting trust is to infer unknown aspects from known aspects; for example, mapping trust of a known community over another unknown one. Liu et al. propose in [31] a classification approach to address the trust prediction problem. They argue that trust relationship between users bases on their behavior in a community and found out that such trust can be predicted using pre-trained classifiers. Dondio el at. suggest in [13] a Wikipedia Trust Calculator (WTC) that consists of a Data Retrieval module which contains the needed data of an article, a Factors Calculator module which calculates and merges the trust factors into the macro-areas defined and a Trust Evaluator module that calculates a numeric trust value and judges it in a natural language explanation using constraints provided by a Logic Conditions module.

Kim and Ahmad propose a framework for measuring web trust (and distrust) that calculates weighted combination of private and public reputation through reliability. Private reputation arises from the direct interactions of a trustor with a trustee, that is, it is measured between two users. While public reputation is based on witnesses' opinion about a trustee, that is, it is measured for each single content provider [27]. From our view of point, this framework embodies a promising approach, but it is based solely on authorship metrics.

In contrast to these related studies, the focus of this work is on the presentation of information based on annotation in collaborative platforms. The aim is to make it easier for users to make decisions about user-generated content by providing the relevant metadata e.g authorship, user review and embedded text-features. This thesis suggests trust as the key for dealing with such content. In addition, our proposed model offers an unique set of dimensions, but the metrics they contain are consistent with the literature and the user's mental model of trust.

Bibliography

[1] A. Abdul-Rahman and S. Hailes, *Supporting trust in virtual communities*, in System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on, IEEE, 2000, pp. 9–pp.

[2] C. Bizer and R. Cyganiak, *Quality-driven information filtering using the wiqa policy framework*, Web Semantics: Science, Services and Agents on the World Wide Web, 7 (2009), pp. 1–10.

[3] J. E. Blumenstock, *Size matters: word count as a measure of quality on wikipedia*, in Proceedings of the 17th international conference on World Wide Web, ACM, 2008, pp. 1095–1096.

[4] C. Brando and B. Bucher, *Quality in user generated spatial content: A matter of specifications*, in Proceedings of the 13th AGILE international conference on geographic information science, Springer Verlag: Guimar aes, Portugal, 2010, pp. 11–14.

[5] C. Castelfranchi and R. Falcone, *Trust theory: A socio-cognitive and computational model*, vol. 18, John Wiley & Sons, 2010.

[6] C. Castelfranchi, R. Falcone, and G. Pezzulo, *Trust in information sources as a source for trust: a fuzzy approach*, in Proceedings of the second international joint conference on Autonomous agents and multiagent systems, ACM, 2003, pp. 89–96.

[7] C. Castillo, M. Mendoza, and B. Poblete, *Information credibility on twitter*, in Proceedings of the 20th international conference on World wide web, ACM, 2011, pp. 675–684.

[8] D. Ceolin, V. Maccatrozzo, L. Aroyo, and T. De-Nies, *Linking trust to data quality*, in 4th International Workshop on Methods for Establishing Trust of (Open) Data, 2015.

[9] S.-C. Chin, W. N. Street, P. Srinivasan, and D. Eichmann, *Detecting wikipedia vandalism with active learning and statistical language models*, in Proceedings of the 4th workshop on Information credibility, ACM, 2010, pp. 3–10.

[10] C. Dai, D. Lin, E. Bertino, and M. Kantarcioglu, *An approach to evaluate data trustworthiness based on data provenance*, in Workshop on Secure Data Management, Springer, 2008, pp. 82–98.

[11] W. H. DeLone and E. R. McLean, *Information systems success: The quest for the dependent variable*, Information systems research, 3 (1992), pp. 60–95.

[12] P. Denning, J. Horning, D. Parnas, and L. Weinstein, *Wikipedia risks*, Communications of the ACM, 48 (2005), pp. 152–152.

[13] P. Dondio, S. Barrett, S. Weber, and J. M. Seigneur, *Extracting trust from domain analysis: A case study on the wikipedia project*, in International Conference on Autonomic and Trusted Computing, Springer, 2006, pp. 362–373.

[14] A. Figueiredo, O. Nabuco, T. Al-Chueyr, and M. Rodrigues, *Framework proposal to evaluate trustworthiness in an online community*, in Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 03, IEEE Computer Society, 2009, pp. 579–582.

[15] F. Figueiredo, H. Pinto, F. BeléM, J. Almeida, M. GonçAlves, D. Fernandes, and E. Moura, *Assessing the quality of textual features in social media*, Information Processing & Management, 49 (2013), pp. 222–247.

[16] A. J. Flanagin and M. J. Metzger, *Digital media and youth: Unparalleled opportunity and unprecedented responsibility*, Digital media, youth, and credibility, (2008), pp. 5–27.

[17] B. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, et al., *What makes web sites credible?: a report on a large quantitative study*, in Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, 2001, pp. 61–68.

[18] M. Gamble and C. Goble, *Quality, trust, and utility of scientific data on the web: Towards a joint model*, in Proceedings of the 3rd international web science conference, ACM, 2011, p. 15.

[19] Y. Gil and D. Artz, *Towards content trust of web resources*, Web Semantics: Science, Services and Agents on the World Wide Web, 5 (2007), pp. 227–239.

[20] Y. Gil and V. Ratnakar, *Trusting information sources one citizen at a time*, in International Semantic Web Conference, Springer, 2002, pp. 162–176.

[21] M. Gupta, P. Zhao, and J. Han, *Evaluating event credibility on twitter*, in Proceedings of the 2012 SIAM International Conference on Data Mining, SIAM, 2012, pp. 153–164.

[22] M. Harpalani, M. Hart, S. Singh, R. Johnson, and Y. Choi, *Language of vandalism: Improving wikipedia vandalism detection via stylometric analysis*, in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, Association for Computational Linguistics, 2011, pp. 83–88.

[23] M. Hu, E.-P. Lim, A. Sun, H. W. Lauw, and B.-Q. Vuong, *Measuring article quality in wikipedia: models and evaluation*, in Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, ACM, 2007, pp. 243–252.

[24] S. Javanmardi and C. V. Lopes, *Modeling trust in collaborative information systems*, in Collaborative Computing: Networking, Applications and Worksharing, 2007. CollaborateCom 2007. International Conference on, IEEE, 2007, pp. 299–302.

[25] C. Jensen, S. Poslad, and T. Dimitrakos, *Trust Management: Second International Conference, iTrust 2004, Oxford, UK, March 29-April 1, 2004, Proceedings*, vol. 2995, Springer, 2004.

[26] K. Kelton, K. R. Fleischmann, and W. A. Wallace, *Trust in digital information*, Journal of the American Society for Information Science and Technology, 59 (2008), pp. 363–374.

[27] Y. A. Kim and M. A. Ahmad, *Trust, distrust and lack of confidence of users in online social media-sharing communities*, Knowledge-Based Systems, 37 (2013), pp. 438–450.

[28] A. Kittur, B. Suh, and E. H. Chi, *Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia*, in Proceedings of the 2008 ACM conference on Computer supported cooperative work, ACM, 2008, pp. 477–480.

[29] D. Lauterbach, H. Truong, T. Shah, and L. Adamic, *Surfing a web of trust: Reputation and reciprocity on couchsurfing. com*, in Computational science and engineering, 2009. CSE'09. International conference on, vol. 4, IEEE, 2009, pp. 346–353.

[30] A. Lih, *Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource*, Nature, 3 (2004).

[31] H. Liu, E.-P. Lim, H. W. Lauw, M.-T. Le, A. Sun, J. Srivastava, and Y. Kim, *Predicting trusts among users of online communities: an epinions case study*, in Proceedings of the 9th ACM conference on Electronic commerce, ACM, 2008, pp. 310–319.

[32] M. Luca, *Reviews, reputation, and revenue: The case of yelp. com*, (2016).

[33] T. Lucassen and J. M. Schraagen, *Factual accuracy and trust in information: The role of expertise*, Journal of the American Society for Information Science and Technology, 62 (2011), pp. 1232–1242.

[34] S. P. Marsh, *Formalising trust as a computational concept*, (1994).

[35] R. C. Mayer, J. H. Davis, and F. D. Schoorman, *An integrative model of organizational trust*, Academy of management review, 20 (1995), pp. 709–734.

[36] M. J. Metzger, *Making sense of credibility on the web: Models for evaluating online information and recommendations for future research*, Journal of the American Society for Information Science and Technology, 58 (2007), pp. 2078–2091.

[37] M. J. Metzger and A. J. Flanagin, *Credibility and trust of information in online environments: The use of cognitive heuristics*, Journal of Pragmatics, 59 (2013), pp. 210–220.

[38] B. Mu and S. Yuan, *A method for evaluating initial trust value of direct trust and recommender trust*, in Computer Design and Applications (ICCDA), 2010 International Conference on, vol. 2, IEEE, 2010, pp. V2–185.

[39] F. Naumann, *Quality-driven query answering for integrated information systems*, vol. 2261, Springer, 2003.

[40] J. R. Nurse, I. Agrafiotis, M. Goldsmith, S. Creese, and K. Lamberts, *Two sides of the coin: measuring and communicating the trustworthiness of online information*, Journal of Trust Management, 1 (2014), p. 5.

[41] J. R. Nurse, S. Creese, M. Goldsmith, and S. S. Rahman, *Supporting human decision-making online using information-trustworthiness metrics*, in International Conference on Human Aspects of Information Security, Privacy, and Trust, Springer, 2013, pp. 316–325.

[42] J. R. Nurse, S. S. Rahman, S. Creese, M. Goldsmith, and K. Lamberts, *Information quality and trustworthiness: a topical state- of- the- art review*, in The International Conference on Computer Applications and Network Security (ICCANS) 2011, IEEE, 2011, pp. 492–500.

[43] W. Ou, X. Wang, W. Han, and Y. Wang, *Research on trust evaluation model based on tpm*, in Frontier of Computer Science and Technology, 2009. FCST'09. Fourth International Conference on, IEEE, 2009, pp. 593–597.

[44] M. Potthast, B. Stein, and R. Gerling, *Automatic vandalism detection in wikipedia*, in European conference on information retrieval, Springer, 2008, pp. 663–668.

[45] I. Pranata, R. Athauda, and G. Skinner, *Modeling decentralized reputation-based trust for initial transactions in digital environments*, ACM Transactions on Internet Technology (TOIT), 12 (2013), p. 8.

[46] I. Pranata and W. Susilo, *Are the most popular users always trustworthy? the case of yelp*, Electronic Commerce Research and Applications, 20 (2016), pp. 30–41.

[47] S. S. Rahman, S. Creese, and M. Goldsmith, *Accepting information with a pinch of salt: handling untrusted information sources*, in International Workshop on Security and Trust Management, Springer, 2011, pp. 223–238.

[48] T. U. Rahman, S. Khusro, I. Ullah, and Z. Ali, *Exploiting user expertise and willingness of participation in building reputation model for scholarly community-based question and answering (cqa) platforms*, in Computer Science On-line Conference, Springer, 2018, pp. 436–444.

[49] L. Rassbach, T. Pincock, and B. Mingus, *Exploring the feasibility of automatically rating online article quality*, in Proceedings of the 2007 International Wikimedia Conference (WikiMania), Taipei, Taiwan, 2007, p. 66.

[50] S. Y. Rieh and N. Belkin, *Interaction on the web: Scholars' judgement of information quality and cognitive authority*, in Proceedings of the annual meeting-american society for information science, vol. 37, Information Today; 1998, 2000, pp. 25–38.

[51] S. Y. Rieh and N. J. Belkin, *Understanding judgment of information quality and cognitive authority in the www*, in Proceedings of the 61st annual meeting of the american society for information science, vol. 35, 1998, pp. 279–289.

[52] S. Y. Rieh and D. R. Danielson, *Credibility: A multidisciplinary framework*, Annual review of information science and technology, 41 (2007), pp. 307–364.

[53] H. Singal and S. Kohli, *Escalation of trust analysis in web*, in Proceedings of the 12th ACM International Conference on Computing Frontiers, ACM, 2015, p. 56.

[54] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser, *Assessing information quality of a community-based encyclopedia.*, in IQ, 2005.

[55] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser, *Information quality work organization in wikipedia*, Journal of the American society for information science and technology, 59 (2008), pp. 983–1001.

[56] M. A. Tate, *Web wisdom: How to evaluate and create information quality on the Web*, CRC Press, 2009.

[57] G. K. Tayi and D. P. Ballou, *Examining data quality*, Communications of the ACM, 41 (1998), pp. 54–57.

[58] T. Tucker, *Online word of mouth: characteristics of yelp. com reviews*, Elon Journal of Undergraduate Research in Communications, 2 (2011), pp. 37–42.

[59] W. Y. Wang and K. R. McKeown, *Got you!: automatic vandalism detection in wikipedia with web-based shallow syntactic-semantic modeling*, in Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, 2010, pp. 1146–1154.

[60] X. Wang, L. Liu, and J. Su, *Rlm: A general model for trust representation and aggregation*, IEEE Transactions on Services Computing, 5 (2012), pp. 131–143.

[61] Z. Wang, *Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews*, The BE Journal of Economic Analysis & Policy, 10 (2010).

[62] M. Warncke-Wang, D. Cosley, and J. Riedl, *Tell me more: an actionable quality model for wikipedia*, in Proceedings of the 9th International Symposium on Open Collaboration, ACM, 2013, p. 8.

[63] A. G. West, J. Chang, K. K. Venkatasubramanian, and I. Lee, *Trust in collaborative web applications*, Future Generation Computer Systems, 28 (2012), pp. 1238–1251.

[64] A. Whitby, A. Jøsang, and J. Indulska, *Filtering out unfair ratings in bayesian reputation systems*, in Proc. 7th Int. Workshop on Trust in Agent Societies, vol. 6, 2004, pp. 106–117.

[65] D. M. Wilkinson and B. A. Huberman, *Assessing the value of coooperation in wikipedia*, arXiv preprint cs/0702140, (2007).

[66] D. M. Wilkinson and B. A. Huberman, *Cooperation and quality in wikipedia*, in Proceedings of the 2007 international symposium on Wikis, ACM, 2007, pp. 157–164.

[67] L. Xin, S. Leyi, W. Yao, X. Zhaojun, and F. Wenjing, *A dynamic trust conference algorithm for social network*, in P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2013 Eighth International Conference on, IEEE, 2013, pp. 340–346.

[68] S. Xu, R. Sandhu, and E. Bertino, *Tiupam: A framework for trustworthiness-centric information sharing*, in IFIP International Conference on Trust Management, Springer, 2009, pp. 164–175.

[69] Z. Yan, R. Kantola, and P. Zhang, *A research model for human-computer trust interaction*, in Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on, IEEE, 2011, pp. 274–281.

[70] H. Zeng, M. A. Alhossaini, L. Ding, R. Fikes, and D. L. McGuinness, *Computing trust from revision history*, tech. rep., Stanford Univ Ca Knowledge Systems LAB, 2006.

Chapter 4

Generating Trust in Collaborative Annotation Environments

[1]

4.1 Abstract

The main goal of this work is to create a model of trust, which can be considered as a reference for developing applications oriented on collaborative annotation. Such a model includes design parameters inferred from online communities operated on collaborative content. This study aims to create a static model, but it could be dynamic or more than one model depending on the context of an application. An analysis on Genius as a peer production community was done to understand user behaviors. This study characterizes user interactions based on the differentiation between Lightweight Peer Production (LWPP) and Heavyweight Peer Production (HWPP). It was found that more LWPP- interactions take place in the lower levels of this system. As the level in the role system increases, there will be more HWPP- interactions. This can be explained, as LWPP-interactions are straightforward, while HWPP-interactions demand more agility by the user. These provide more opportunities and therefore attract other users for further interactions.

Keywords Collaboration, Trust, Annotation, Genius, User Generated Content, Lightweight Peer Production, Heavyweight Peer Production.

We refer the reader to the publication:
Title=Generating trust in collaborative annotation environments,
author=Al Qundus, Jamal,
booktitle=Proceedings of the 12th International Symposium on Open Collaboration Companion,
pages=3,
year=2016,
organization=ACM
https://doi.org/10.1145/2962132.2962136.

[1] The content of this chapter has been published in [5]

Overview

This thesis is the first study to consider Genius as a study case. It may read like a technical report in some places, nevertheless, it is important to get an overview of this platform to understand the preliminary analysis of the trust model proposed in this dissertation. For example, member roles, types of activity, privileges, etc. are compared against several aspects of the literature and brought together to develop the trust model. Therefore, the following chapter provides an extract of the technical and social analysis of the social media genius as well as a description of the data collected. For more details, please refer to the technical report published in [2] on which the following sections are based.

Chapter 5

Technical Analysis of the Social Media Platform Genius

[1]

5.1 Abstract

Genius members have six different roles that are closely tied to authorizations sequentially: Whitehat, Editor, Moderator, Verified Artist, Mediator and Staff. This study monitored Genius activities on firehose for five weeks, collected 1.3 million activities, 762 thousand of them are annotation activities[2]. 57 thousand unique users were observed and it was found, that users generate on average 13.33 annotation activities in this period of analysis, which is 0.36 annotation activities per day. The distribution over user groups displays the roles: Moderator, Staff, Artist, Mediator and Whitehat. Whitehat embody the most registered user, but when it comes to drive Genius ahead, then those roles are presented in the following sequence: Staff, Mediator, Moderator, Editor, Artist and at the end is the role Whitehat.

Intelligence Quotient (IQ) can be earned by the most of activities and indicate experience of a member. Although a count of IQs is required to do certain activities, for instance to post into forums, but it does not promote automatically a member's role to become a higher member level.

High-quality[3] annotations and decision maker such as an Editor establish nomination criteria. Earning IQs implies to edit pages; a page is edited on average 295 times, which varies greatly from the median of 195 times. This indicates that some pages attract users more than others.

For developers Genius provides API, documentation and support forum. There are subdomains in different countries and languages. This study attempts to discover members' collaboration on editing Genius pages and clarifies the social, technical and participation architecture of Genius, such as member's permissions as well as options, activity types and distribution of page edits.

[1]The content of this chapter has been published in [2]

[2]Those are a part of the collected activities that refer to annotations and we call them annotation activities.

[3]well written, without errors in grammar and contains knowledge that adds meaning depth

This chapter is structured as follows: Section 5.3 introduces the social structure of Genius (annotation and member roles). Section 5.4 continues with the technical options for developers to bind Genius services in applications and firehose as notifications process. Section 5.5 describes the member activity types and collaboration on Genius. Finally, section 5.6 presents our conclusions.

5.2 Introduction

In the 1990s web content was commonly generated by a small amount of publishers and the far bigger rest of users were consumers. Only a decade later another type of content became available on the web: User Generated Content (UGC), in which more and more users participate in content generation. UGC's domain is Social Media (SM) that additionally includes a social networking platform [1] for user collaboration such as Genius. With the trend of social networking web- sites (e.g., 2003 MySpace, 2004 Facebook, 2004 Flickr, 2006 Twitter, Instagram 2010 etc.) SM has become an additional channel of content sharing variety that enables annotation of UGC.

Genius as a part of SM follows it's modern way strategy that allows user to create and collaboratively modify UGC to support annotating, which makes Genius an online platform for annotations [13][12], that *breaks down text with line-by-line annotations* [9] and provides interpretation of any form of text [15].

5.3 Social Structure

Genius as a collaborative annotation platform with its UGC builds a social media, in which everyone can participate to communicate indirectly with other members to interchange and value interpretations provided in form of annotation to clarify meanings of a piece of text. After registration the member is assigned a role, which is Whitehat and can be promoted by earning IQs to extend permissions. How to earn IQ, which other roles and permission are possible at Genius will be clarified in the next subsections. First, we want to clarify the used terms in this technical report:

Genius annotation A frame that includes the interpretation of a piece of text and options. Additionally to a piece of text, an annotation can be referred to by description and suggestion.

Annotation activity A member activity as upvote, downvote, suggest, reply etc., which deals with an annotation. Such activity we call annotation-activity.

5.3.1 User-Generated Content (UGC)

UGC in Genius comes in the form of annotation that is the essential module in order to build up member interactions. Exclusively by means of annotation a text can be interpreted, which is the main aim of Genius. Description and suggestion can be annotated, too. By highlighting any piece of text appears a pop-up field that enables creation of annotation. An annotation that was generated by a Whitehat remains un-reviewed waiting to be accepted by authorized member to get published.

Figure 5.1 represents an annotation (on the right side) including an interpretation of a piece of text (yellow marked) of a lyric (on the left side). At the top of the annotation a set of meta data (number, names, roles, IQ's and attributions of the contributors) is displayed. We can see who originally created the annotation and who accepted it. The edits can be viewed on the page. Below the interpretation (annotation body)is a set of meta data (e.g. annotation IQ count) and activity options. These options are voting, sharing, following, and creating a suggestion or edit to improve the annotation that a member with the appropriate permissions can perform.

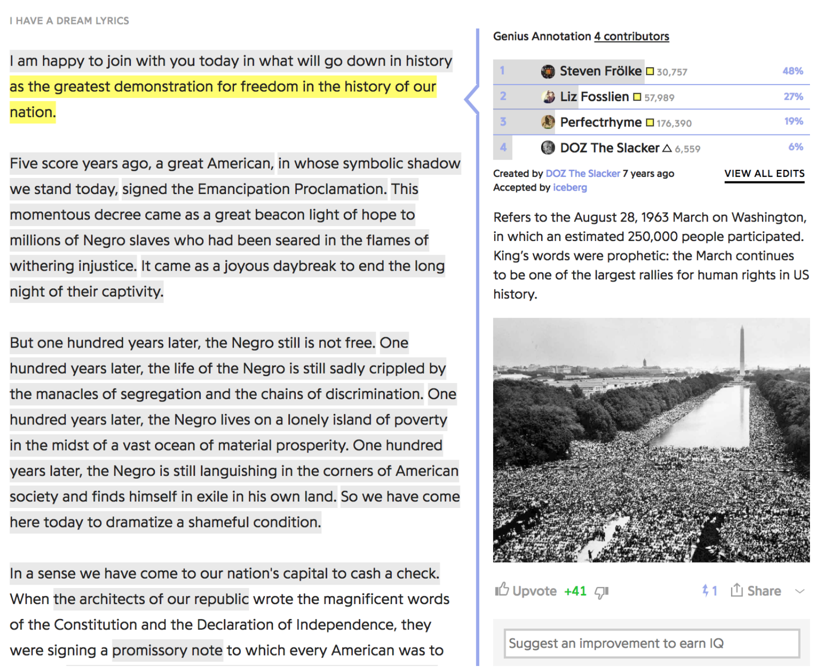


Figure 5.1: Annotation Example

Screenshot:https://genius.com/544987 accessed: 2018-08-12 at 19.51.06
This figure shows a annotation on a lyric at Genius

5.3.2 Participation

Each member has different roles that bear different permissions. Certain activities can be carried out with certain rights. Mostly the IQs score characterizes rights; the more rights a member has, the more central is his role in the community. Members take one of the following roles: *Whitehat*, *Artist*, *Editor*, *Mediator*, *Moderator* or *Staff*. Genius describes generally Editor's main task by correcting content and deciding about contributions of *Whitehat*, who is a normal user [6] and usually is new member. Moderator has more management activities such as verify Artist, who is an owner of a lyric. Mediator assists new members. Only Moderator's commune or Staff can promote both Moderator and Mediator [10, 11]. In detail Genius specifies the roles with following particulars:

Whitehat is usually new member and is at the low level of the community and has accordingly little permissions.

Editor is a contributor with high-quality annotations, which are formatted, consistently well written, without errors in grammar, spelling or punctuation and contain knowledge that adds depth to the meaning of the referent. Whitehat is selected by Editors/Moderators to become an Editor.

Moderator is an Editor who has proven , that he can coach[4] other contributors to write consistently high-quality annotations. Other tasks are to resolve conflicts and to *de-editor*[5] members, who break community guidelines or participate consistently with contribution of poor quality.

Verified artist has written own lyrics with full permissions on it and has the ability to annotate his own works, as another way to connect with fans.

Mediator role is designed for leaders, who are welcoming and assisting everyone, strive for a positive environment and either uninterested in or unable to do the work of Editors. Moderator commune selects mediators.

Staff is a designer, an engineer and another employee; those are very few in number, 12 up to now [8]. Responsible for curating the site and the community, they have exclusive powers and abilities.

Nomination process is based on the rule: Not Quantity but Quality. Through high quality of contributions in spelling, sources, images and citations members are promoted to gain more abilities independently of their IQ score range [6]. That is, there are members, who have more IQs count, but "minor" role level, which means less permission.

Table 10.3 in the appendix shows that each role is associated with a color, which is used for the detection of a role, while the IQ counters do not matter. The rights differ between the roles, but these also differ among users within a role. Some users, due to their IQ number own more rights than their equivalents within a role, for example see the role Mediator in the Table 10.3, which is discussed in section action below. In this Table, we classify member permissions according to the type of their operations and effect on Genius. Those categories are:

- *access* Specific areas are available only for certain users. It means read only access without write permission. Whitehats and Artists have mostly no access, because they have no administrative duties, as we can see in the Table 10.3 in the appendix.

- *action* Extreme roles such as Whitehat and staff are clear. Whitehat has hardly any permissions, while Staff holds full permissions. In this category the difference compared the other roles is in evidence especially Mediator, Editor and Moderator. It can be seen that Moderator and Staff are at the same level, and then comes Mediator followed by Editor. Certain Mediators with specific properties (600+) have more permissions than Editors or other Mediators, for example *lock / unlock pages* permission has an Editor but no Mediator with less than 600 IQ, while Mediator with 600+ IQ is comparable with Moderator and Staff.

- *create* is divided in create into song page (annotation, song page text, vote, album, tracklist and profile picture) and in create into forum (topic, post, postlets). Create into song page is usually open to all, but into forum the access is very limited.

[4] Inspire and train other contributors

[5] Set down a contributor's role from Editor to Whitehat

- *edit* included *delete* rights have mostly Moderators and Staffs, an Artist has almost the same rights but only over his own pages.
- *Promotion* shows again the privilege of the individual roles. The permissions *verifying artist* and *de-editor* contributor can only Staffs, while promoting a Mediator or a Moderator is a Staff's or Moderator commune responsibility. An Editor can only promote a Whitehat to an Editor.

5.4 Technical Structure

5.4.1 Developers

Genius provides API, documentation and support forum. Developers can sign up, create an API client and get access token to export own annotation into their application or website. A request with the annotation *id*[6] and developer's *key* to Genius server[7] responses a JSON object[8] in Figure 5.2, that contains all meta data like author, body, vote, time stamp etc. of the requested annotation. The technical services provided by Genius are listed in Table5.1 below.

Table 5.1: Technical Services

API URL	https://api.genius.com/
Supporting	http://genius.com/api-support
API client	http://genius.com/developers
Topic	annotation, voting, information
Request protocol	http, https
Response format	JSON, REST
Authentication protocol	OAuth2
Contact	code@genius.com
Documentation	https://docs.genius.com

According to Genius documentation
This table showing an overview of provided technical services

5.4.2 Firehose

Firehose pushes notifications about members' activities. It is the starting point of our study for data collection on Genius. This mechanism documents and records action of all members. Firehose includes filters to select specific notification languages and topics. As shown in Figure 5.3 an activity at Firehose consists of contributor's name, type, subject, symbol of type and time stamp as overview and by clicking on the activity its details are shown.

[6] can be found at the end of url in browser by clicking the annotation

[7] *http://api.genius.com/annotations/(annotation id)?access_token=(developer's key)*

[8] a map of its structure is provided in the appendix and more explained in Activity Study section 5.5 of this work

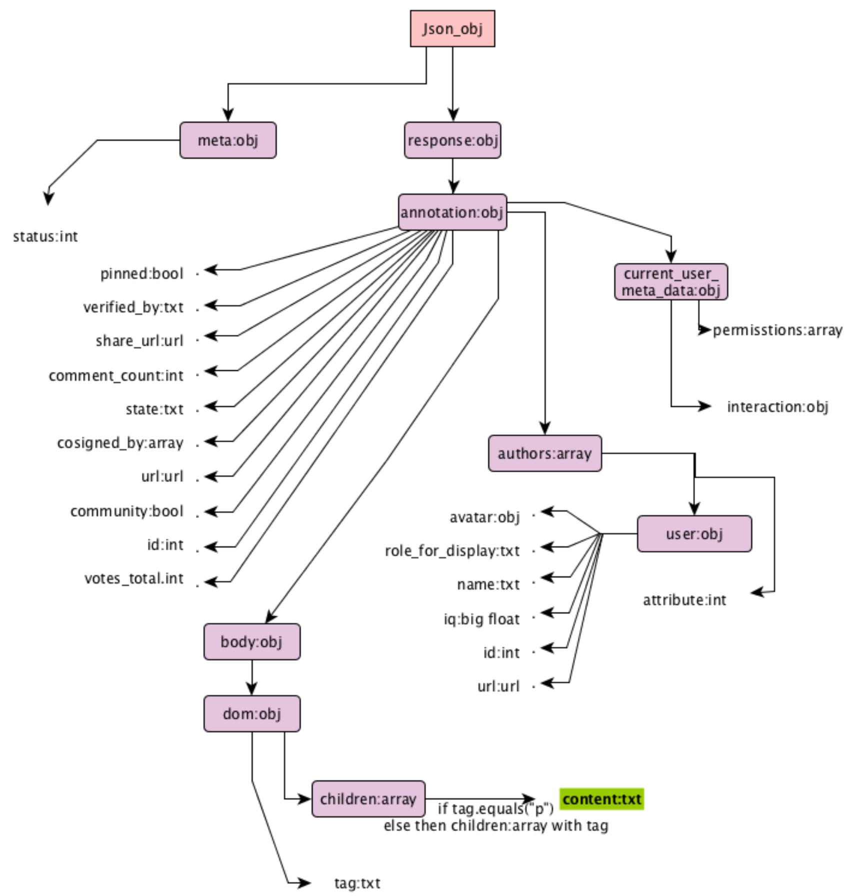


Figure 5.2: Annotation Activity JSON Object

Created: 25/08/2015
This figure shows a JSON object overview of annotation activity

5.5 Activity Study

Firehose is used as channel to be notified about user activities on Genius. An approach model is developed for that, it holds a couple operation steps: (a) getting activity notification (b) extracting available links (c) fetching JSON objects (d) identifying and classifying the information (i) forwarding into data base (see Figure 5.2).

Each notification has a sequence of characters that embodies a regular expression: {subject predicate object} as presented in the appendix Table 10.2. The column *Regex* describes the regular expression of every notification. For example the notification *username upvoted annotation* can be subdivided into username as subject, upvoted as predicate and annotation as object. Table 5.2 represents an overview for the collected data over the observation time span.

Figure 5.4 shows the distribution of the collected activities daily, which we began to collect form 30.09.2015 16:43 to 06.11.2015 23:27. On the first day the begin was at 16:43 PM, therefore it is taken out of the calculation. As shown in Figure 5.7 mean value (34,867)

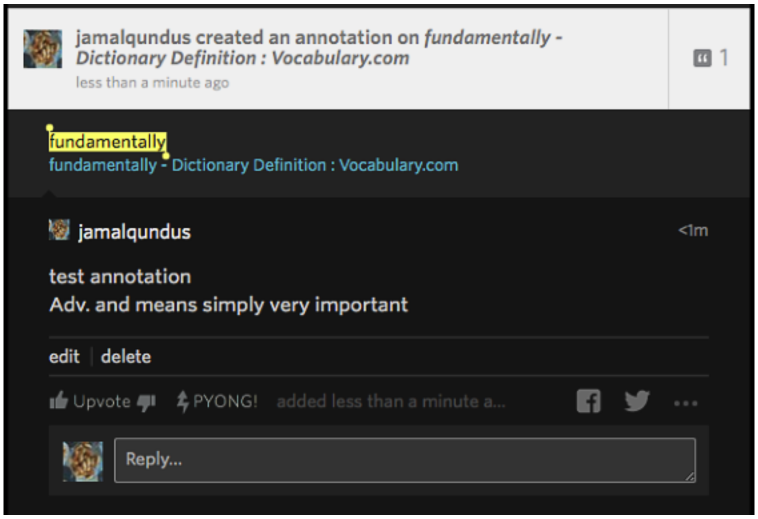


Figure 5.3: Activity on Firehose

Screenshot:16/09/2015
This figure shows an activity of creating annotation

Table 5.2: Summary Collected Activities

Observation period	2015-09-30 to 2015-11-06
Total notifications	1,306,560
Total annotations	762,853
Unique annotations	240,060
Unique annotators*	57,222

This table showing an overview of collected notifications (activities) over five weeks
*members, who generate annotation.

and the median (34,074) are close together, which indicates a balanced distribution and it does not matter what day it is, it's almost the same amount. The standard deviation is relative low, which confirms that statement.

The plot in Figure 5.6 for the active users daily has similar properties as those in Figure 5.4. Mean 5,863 and median 5,309 are close together in Figure 5.4. Both representations Figure 5.4 and 5.6 are build on annotation activities.

A user generated on mean 13.33 annotation activities in 5 weeks, which is 0.36 annotation activity per day. All annotation activities are based on 240,060 unique annotation activities, that is, a user carries out 4.19 different annotation activities on average.

Figure 5.8 shows the distribution of edits over the pages. Interestingly, the edits are uniform distributed, which means that almost each page has edits, equally, which is unusual observation. On mean a page is edited 295 times, while the median is 195 and the standard deviation is 461 as shown in Figure 5.9, which means that the distribution contains extreme values as presented in Figure 5.10.

This is traceable, some pages are more interesting (maximum value :7140 edits) than

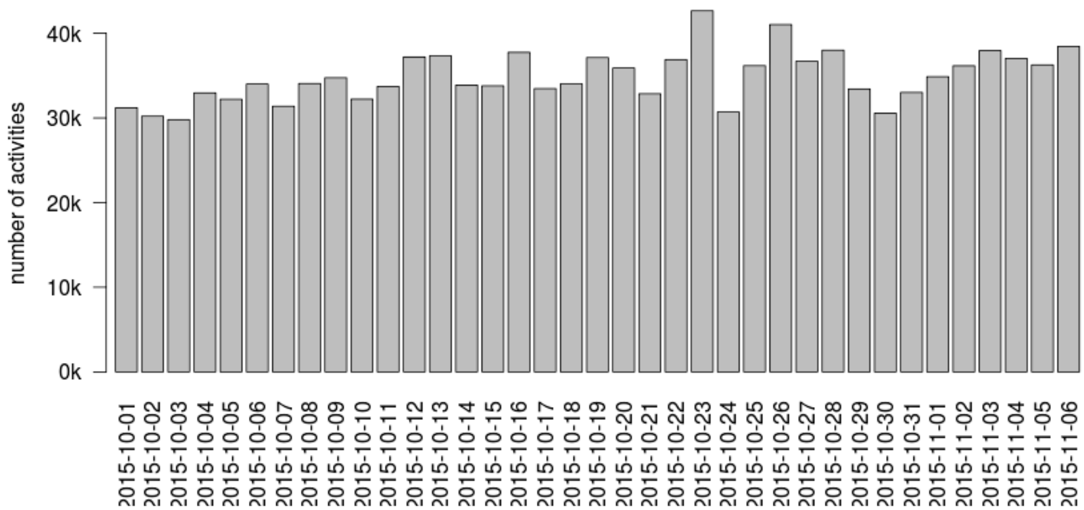


Figure 5.4: Activity Notifications Overview

This figure shows how much activities are generated daily

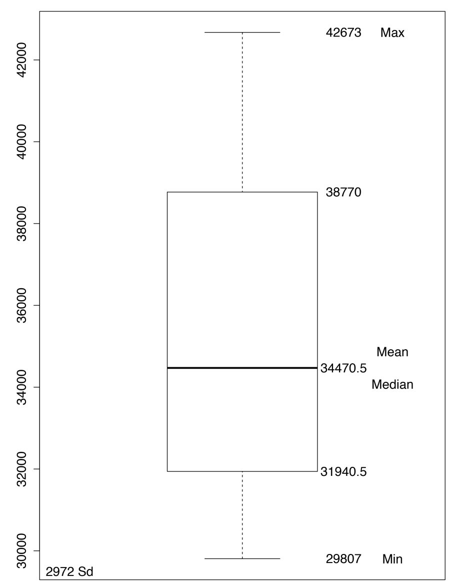


Figure 5.5: Statistic Elements of the Collected Activities Daily

This figure shows minimum (Min), mean, median, standard deviation (Sd) and maximum (Max) of the collected activities distribution

others (minimum value: 1 edits), therefore, those are more visited and more edited. Nevertheless, it stills not a measure of whether a page is interesting for users or not, as long as the creation date of pages is not taken into consideration, young pages have not the same opportunity to get edits as older pages.

According to the period of analysis the distribution in Figure 5.11 shows how many annotation activities are generated by how many users. Additionally, it shows interestingly a remarkable amount (7,396) as inner bolter of users who generate an annotation activities

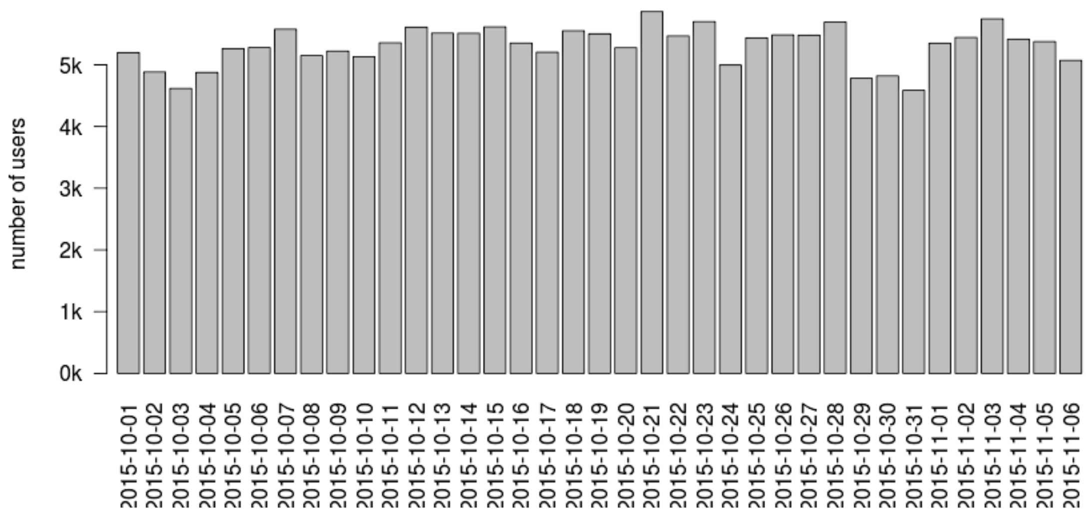


Figure 5.6: Active Users

This figure shows active users daily

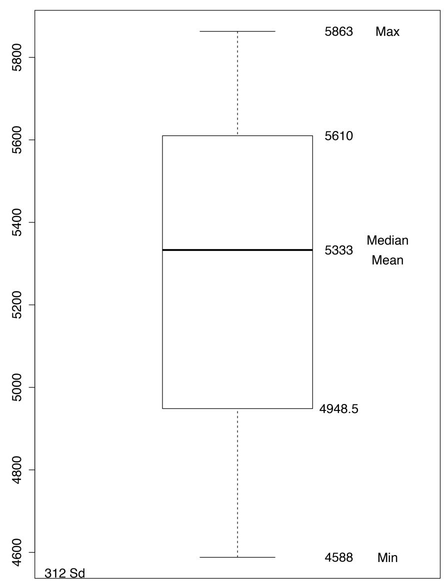


Figure 5.7: Statistic Elements of Activities and Users

This figure shows minimum (Min), mean, median, standard deviation (Sd) and maximum (Max) of the distribution users to activities

count between 10 and 50.

If we take both extreme values (1 and > 5k) out and look at the next big user group (16.5% of total users 57,222), which generate precisely only two activities, it is clear that the group of this inner bolter, which builds 12.9% of total users, is a large group. To recap here, this observation refers to the observation period and only to annotation activities.

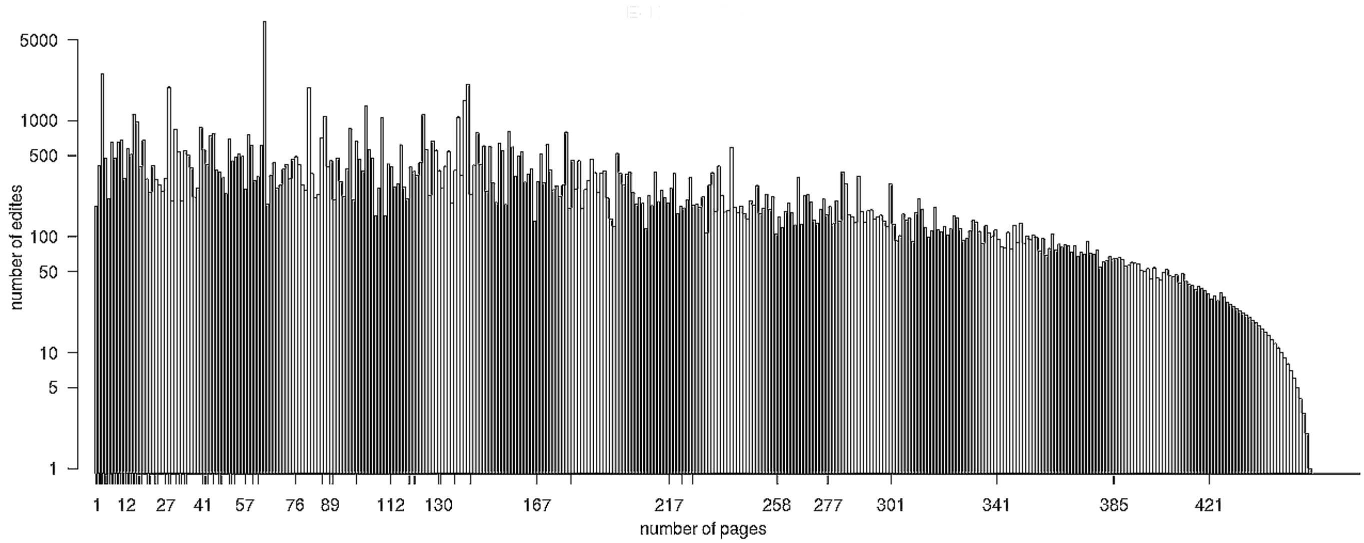


Figure 5.8: Page Edits

This figure shows edits overview on pages

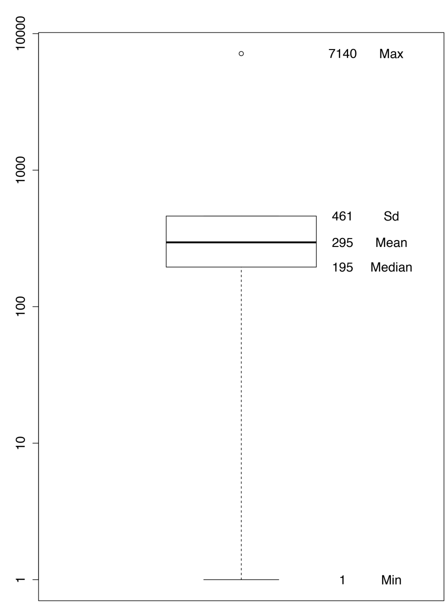


Figure 5.9: Statistics of Edits on Pages

This figure shows statistic values minimum (Min), mean, median, standard deviation (Sd) and maximum (Max) of edits on pages

5.5.1 Activity Types

In the observation time span 78 types of activities have been identified as presented in the appendix in the Table 10.1, from which 52 types are referred to annotation activities. These assume a certain pattern, which can be described with the regular expressions as

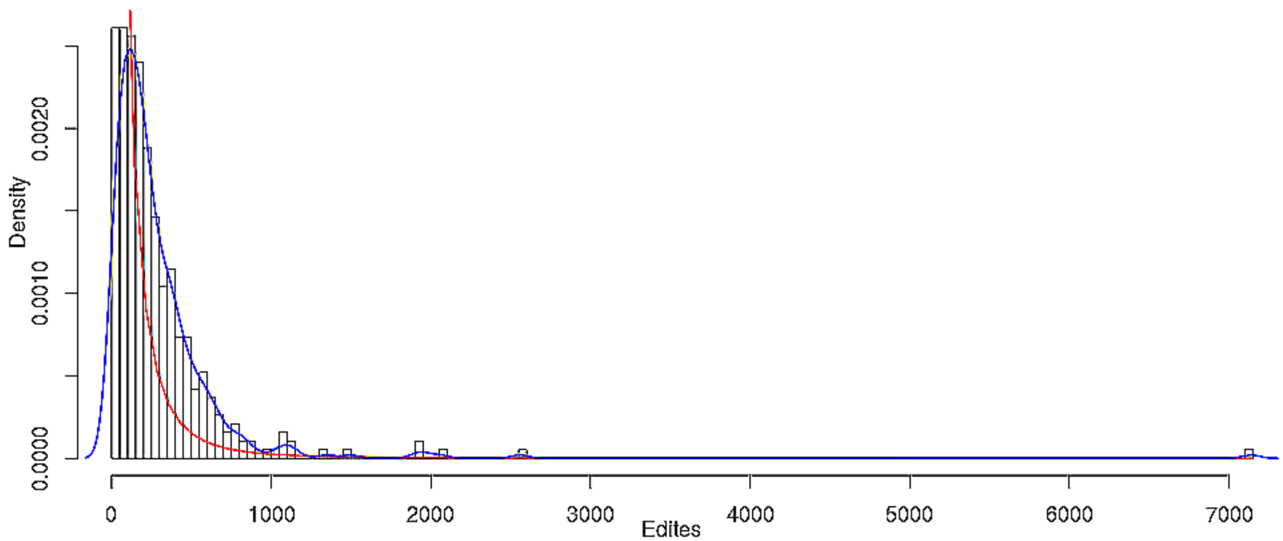


Figure 5.10: Function Curve of the Edits

This figure shows the distribution function of the edits over pages

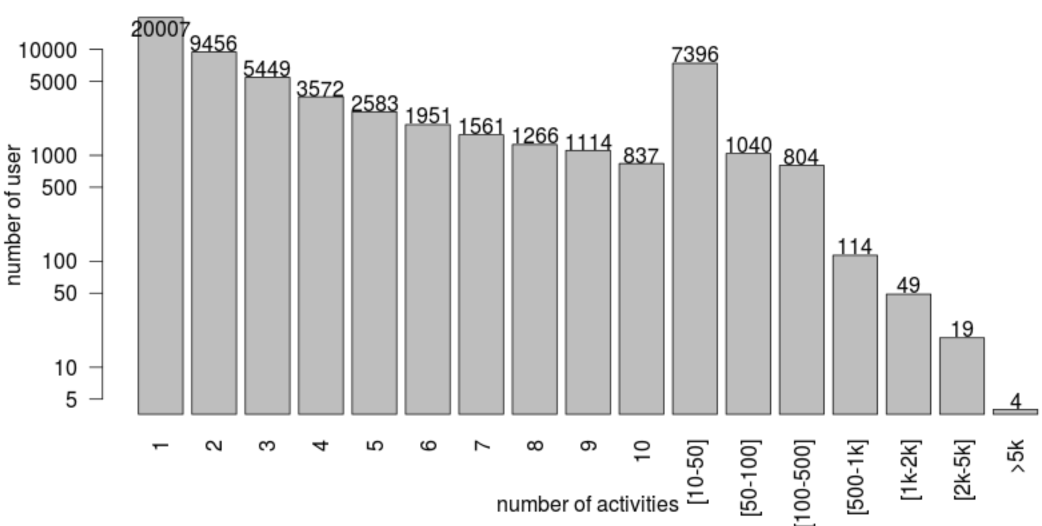


Figure 5.11: Active Users

This figure shows the distribution of the active users.

follows: With:

$$\sum_{subject} = \{members\},$$
$$\sum_{predicate} = \{activity_t ypes\}, \quad (5.1)$$
$$\sum_{object} = \{\epsilon, sub - activities\}$$

Regex:

$$\left\{ abc | a \in \sum_{subject}, b \in \sum_{predicate}, c \in \sum_{object} \right\} \tag{5.2}$$

Table 5.3: Abstraction of Activity Types

Category	Predicate	Description
augment	upvoted posted created added replied proposed	expand or add Genius with content
detach	rejected deleted archived downvoted	disconnect and disassemble content from Genius
manage	edited accepted marked integrated merged moved incorporated un-/pinned un-/locked verified	manage content. Activities that affect a change on Genius content
movement	followed mentioned registration pyonged cosigned gave access	interact with members and their contents. The influence here is cosmetic and defeat no change

This table showing an abstraction of activities types

Activities were classified into types according to their predicates; definitively there are many other activity types that did not occur in the observation time span, for example made Moderator / Staff, remixes annotation, text correction, embedding tweets / Facebook, create / edit / manage postlets, clear votes, Penalty Box, et cetera. But we focus on those that occurred and which of them we could find meta data about.

5.5.2 Collaboration on Genius

Benkler and Nissenbaum define *peer production* as a socio-economic system of production, which occurs in the digitally networked environment and involves collaboration among peers, who cooperate effectively to produce knowledge [3, 4] and “Goods" developed and shared according to community-defined rules [7]. Terveen and Frey et al. introduce collaboration as a process involving at least two entities working together to achieve shared goals [5, 14].

We adapt these definitions and extend them in the context of Genius to the amount of participant interactions on a Song Page to achieve the goal to interpret text. Our extension builds on the differentiation *lightweight peer production* (LWPP) and *heavyweight peer production* (HWPP) presented by Haythornthwaite, which are used to refer to participant contributions. LWPP involves interactions, which are targeted to simple and independent contribution without initiation relationships among participants. Its power is its simplicity that allows numerousness of participation, in contrast to HWPP that implies extensive and time consuming contributions and involves also more information about contribution and contributor. Thus, its power is to allow analysis based on such information. In these collaborative forms the user's participation occurs based on the complexity and dependency of their interaction, which are characterized by “weak-tie attachments" and “strong-tie attachments". Weak-ties are simple enough for participation, while strong-ties require agility and more experience by participants [7]. LWPP is independent on other contributions and straightforward and is described by weak-ties, while strong-ties identify HWPP, which is dependent and more complex.

Complexity refers to contributions -length, consuming-time and -divisibility as well as contributor's agility. This definitions (peer production and collaboration) and differentiation are suitable for the contributions of Genius, that are based on voluntary participation of peers, and for their properties, that can be distinguished into light and heavy consuming effort by a contributor.

The collaboration design of Genius is presented by users' interactions, which we classify in the dimensions LWPP and HWPP based on Haythornthwaite's approach as illustrated in the Table 5.4. We use predicates as representatives for the interactions. LWPP-predicates are atomistic and independent, therefore, there is no need to manage a history of contribution, but a quantitative recognition and measure are of certain interest. HWPP interactions include predicates matched in strong-ties. They are connected and revised; therefore, a history of contribution is important as well as qualitative recognition is relevant [7]. For instance, the predicates *down-/upvoted* are atomistic, independent, quantitative and done by one click (LWPP). While the predicates *created an annotation* or *proposed an edit* require more agility by contributor and are time-consuming, therefore, they are classified into HWPP. Descriptions of all predicates are in the appendix Table 10.2.

For another possibility to classify the predicates we consider the Song Page as the object of the collaboration on Genius and its life cycle as shown in Figure 5.12, where (A) illustrates HWPP-predicats, (B) Song Page states during the collaboration and (C) shows LWPP-predicates. A Song Page is permanently in one of these states (B). The state initialization (start state init) of a Song Page is the first step for the whole. Interaction (inter in (B)) provides collaboration ways for users and it is the core of such collaboration. *inter* includes both designs of collaboration; for instance a user votes (LWPP (B)) an annotation or a user replies (HWPP (A)) an annotation of another user.

Table 5.4: Collaboration Interactions

	Predicate	Object	Number of Activities
	upvoted	annotation	393,209
	downvoted	suggestion	33,424
	accepted	description	17,591
	marked	comment	11,168
	rejected	Song Page	10,752
	deleted	user	5,711
	archived		5,133
Lightweight	cosigned		1,886
	incorporated		369
	followed	Song Page	138,783
		user	
	pyonged	description	20,036
		annotation	
		Song Page	
	created	annotation	149,000
	edited	description	45,456 (154,505)
	mentioned	Song Page	66,642
	merged	meta data	2,670
	integrated		4,254
Heavyweight	replied		5,146
	added	suggestion	77,428
	proposed	reply	6,838
		edit	
		comment	

This table is an extension of [1] and illustrates the predicates of collaboration design on Genius, which are classified into LWPP and HWPP. Each predicate can form an activity with each object from its group. Groups are separated by a horizontal line. For example: The predicate *followed* (LWPP from the second group) may be combined with the objects *Song Page* or *user*, but not with the object *comment* (LWPP from the first group).

5.6 Conclusion

This chapter described Genius as SM, introduced the opportunities to use it and characterized user's activities. Users can upload text, annotate it and earn IQs by voting by other users and get new roles with more permission for more interactions.

This study analyzed Genius and presented concrete figures on the distribution of activities created over a time span. These figures have shown that Genius is still young, but it is growing very rapidly, which has confirmed the number of the Whitehats as mostly new comers to the number of the other roles. Different user roles have with different responsibilities and privileges. Artists and Staffs are the roles with the most generated content, while

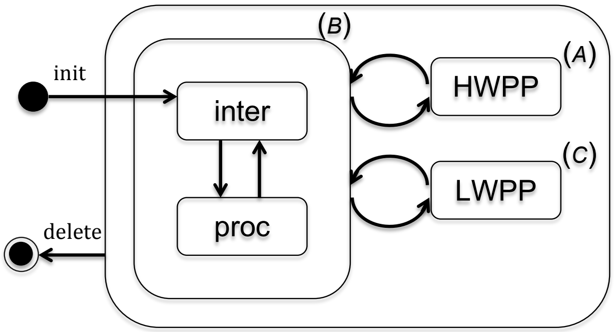


Figure 5.12: Interaction States

This state chart diagram illustrates interaction states of a Song Page

Editors take over the task of correcting the content and Moderators have user coaching and management tasks.

Genius is a stage for providing interpretations of texts, which builds together with the offered options a form for collaborations among users. Such collaboration is embodied by editing pages, knowledge generating as well as sharing and exchange of opinions about texts that will be confirmed or refuted by other users.

Bibliography

[1] J. Al Qundus, *Generating trust in collaborative annotation environments*, in Proceedings of the 12th International Symposium on Open Collaboration Companion, ACM, 2016, p. 3.

[2] J. Al Qundus, *Technical analysis of the social media platform genius*, tech. rep., Freie Universität Berlin, 03 2018.

[3] Y. Benkler, *Coase's penguin, or, linux and" the nature of the firm"*, Yale Law Journal, (2002), pp. 369–446.

[4] Y. Benkler and H. Nissenbaum, *Commons-based peer production and virtue**, Journal of Political Philosophy, 14 (2006), pp. 394–419.

[5] B. B. Frey, J. H. Lohmeier, S. W. Lee, and N. Tollefson, *Measuring collaboration among grant partners*, American Journal of Evaluation, 27 (2006), pp. 383–392.

[6] M. G. I. Genius, *about editor.* http://genius.com/Genius-what-is-an-editor-annotated, 2015. [28/09/2015].

[7] C. Haythornthwaite, *Crowds and communities: Light and heavyweight models of peer production*, in System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on, IEEE, 2009, pp. 1–10.

[8] G. M. G. I., *Moderator and staffer contact listing.* http://genius.com/Genius-moderator-and-staffer-contact-listing-annotated, 2015. [09/10/2015].

[9] G. Media Group Incorporated, *about genius.* http://genius.com/Genius-about-genius-annotated, 2015. [28/09/2015].

[10] G. Media Group-Incorporated, *about mediator.* http://genius.com/Genius-what-is-a-mediator-annotated, 2015. [28/09/2015].

[11] G. Media-Group Incorporated, *about moderator.* http://genius.com/Genius-what-is-a-moderator-annotated, 2015. [28/09/2015].

[12] M. Moghadam, 2015. [28/09/2015].

[13] Technically-Media-Inc., *Genius (website)*, 2015. [28/09/2015].

[14] L. G. Terveen, *Overview of human-computer collaboration*, Knowledge-Based Systems, 8 (1995), pp. 67–81.

[15] Wikipedia, *Genius (website) — wikipedia, the free encyclopedia.* https://en.wikipedia.org/w/index.php?title=Genius_(website)&oldid=682813176, 2015. [28/09/2015].

Overview

The previous chapter gives an overview of our Genius case study, which offers a variety of activities and features. An annotation is a place of text interpretation and contains a series of metrics that we have explored in order to experience their impact on users in dealing with the information provided. This type of interaction suggests that the user makes a decision to trust the information as discussed and implemented so far.

In the Genius study we observed about 80 activities (see Table 10.1 and 10.2), which influence the metadata of an annotation directly (e.g. vote) or indirectly (e.g. author role). The metrics (i.e. metadata elements) were extracted and compared with metrics from the literature on trust. The comparison could give us an indication of the influence of each individual indicator on trust. However, we have applied the Empirical Cumulative Distribution Function (ECDF) to all annotation metrics to examine the distribution changes in the database based on these metrics. Only a part of them has influenced the distribution behaviour of the database, while most of them have not demonstrated such an influence.

The identified relevant metrics were combined into dimensions, whereby the metrics were taken into account in the literature on the subject of trust. These dimensions form the trust model; its creation and calculation are described in detail in the next chapter.

Chapter 6

Calculating Trust in Domain Analysis: Theoretical Trust Model

[1]

6.1 Abstract

In recent decades, information has become increasingly available on the Web. Every user can actively participate in the generation and exchange of information. Investigating the quality of user-generated content (UGC) has therefore become a necessity and an ever-increasing challenge. In collaborative environments where users collect, share and build a knowledge base, trust is an important factor. If, for example, we as users trust UGC on the Web, this influences our interaction with this content. The aim of our research is to propose a model for the evaluation of trust in UGC. Based on the available research results, we define a model for measuring trust in collaborative environments. Our approach is based on three dimensions: stability, credibility and quality. We have combined these three concerns to create a trust-translator. We use a real-word data set of the social annotation platform Genius to calculate the value of our trust in an annotation. Based on this case study, we show which insights can be gained by calculating the trust in such an environment. When information has specific qualities, our approach will enable the user to better determine which information offers the highest level of trust.

Keywords Trust, Quality, Credibility; Stability, Annotation, Social Media

6.2 Introduction

In today's increasingly information-driven world, factors that influence trust in collaborative environments must be understood to ensure the value of user-generated content [38]. When we are building a knowledgebase collaboratively or when sharing information, we need to have trust in the information provided by others in order to build upon or share the existing knowledge [8, 14]. Often our decisions rely on information that is influenced by

[1]The content of this chapter is under review in International Journal of Information Management - Elsevier (IJIM), which is coauthored with A. Paschke, K. Sameer and S. Gupta

the Internet [31]. Metzger and Flanagin [50] and Rieh and Danielson [61] have addressed the evaluation of information in digital media. Fritch and Cromwell [26] and Metzger [49] have showcased in their works that there is lack of authority in online information since, according to Taraborelli [64], we as users are not applying much effort. Lack of trust or negative trust [52, 58] is also discussed, as well as the risk of it [39] and established risk as a fundamental condition for applying trust. That is, the willingness to take risks must be included in the decision-making process of trust. All of these discussions gives us an indication of how important trust is in regard to online information, where the source of information is hardly known to us.

Statistics[2] of social media usage show that North America holds the highest social network usage rate worldwide with 70% compared to the worldwide average of 42% in 2018. Within the United States, 77% of the population had a social media profile and in 2017 with around 209 million users. The range of the users of social networks is progressively and by 2022 the expected number in the United States is around 221 million social network users. On the other side, the number of printed news[3] is gradually decreasing. For example, the total United States daily newspaper in 2017 decreased 11% from 2016. This indicates clearly that news consumption has moved to social media, which are relatively new sources (foundation: Facebook in 2004, Twitter in 2006, Genius 2009 etc.), but with tremendous impact on the way we get information. As a consequence, evaluation of this new kind of information becomes urgent. We need developing approaches to support user in making decision to trust information and use it. Unfortunately, the concept of trust is too complex to put it in one universal definition and a generally accepted definition does not exist [12], despite almost everyone has an initial definition of what trust is, according to his mental model. Literature review (e.g. [22, 28, 42, 57, 59, 60]) shows that trust is explored towards aspects as stability, quality, credibility, believability, reliability, dependability, security, readability etc., entities as users, agents, both etc., relationships as a formula, a graph, one-to-one, one-to-many etc., related terms as untrust, distrust, mistrust, blind-trust etc. and the goal investigated as vandalism, fake-news, event (topic relevance) etc.. We agree that all of these are more or less related to trust. However, we need to break it down to make it possible to investigate. The following conditions shall help bounding trust: 1) including risk, a user (trustor) should be vulnerable in usage of information provided, that is, the information is important. 2) Independence, an information provider (trustee) cannot be controlled. If I can control the provider, so the trust question is senseless and 3) intention, the information provided can be incorrect, but it is not intentionally manipulated. Is the number of edits (stability) sufficient to improve the content to draw conclusions about trust? Is it the quality indicated by the authorship (e.g. reputation and community role) and the representation (error-free in grammar, style, depth of meaning, etc.)? Is it based on the reader rating (credibility)? Are they the degrees of believability, readability, truthfulness, etc.? To our knowledge, there is no approach that combines these concerns under the conditions mentioned to appreciate trust in social media, which is the novelty of our approach.

According to Mayer et al. [48] and Cheng et al. [16], we make a decision whether or not to react to such information provided, and only truly consume information that

[2] https://www.statista.com/topics/3196/social-media-usage-in-the-united-states/ accessed 28.10.2018

[3] http://www.journalism.org/fact-sheet/newspapers/ accessed 05.07.2018

we trust. Scientists acknowledge the importance of trust Deutsch (1958)[21] and many research studies show the importance of trust in everyday life [10, 63].

In the context of the web, trust is linked back to mechanisms for verifying the source of information [9]. Moreover, Berners-Lee et al. [13] and Kirs and Bagchi [36] see trust as essential prerequisite for gathering information from the web. Therefore, the measurement of trust is of interest and, based on the research results presented, it can be assumed that trust measurement can be made by evaluating willingness of users to make decisions and take action on the basis of information provided [29].

Trust is an inter-personal attribute that varies from person to person and depends on context and situation [16, 45]. Based on trust, a user builds an opinion or creates a new perspective on a topic [7]. The source of information is important, especially when it comes to making a decision to trust a person or an organization [65]. Otherwise, we may not accept high quality information (so-called mistrust) due to the unknown source of information.

The Social Media Genius has developed into an important medium for the exchange of content. We focus on this area since information should be extended and connected [44]. Genius has a broad audience that is aimed at users with different backgrounds and experiences. Genius has great potential for exploring merits based on user behavior and interactions. More information about Genius can be found in Section Domain Analysis.

Annotations identify a specific text in a document and contain additional attributes [20] and facilitate working with annotated documents [47]. Metadata that contains references to the document and to the author are examples of such additional attributes.

This work sees trust as a personal trait and a social response, which calls upon selective attention that motivates decision-making based on other proposed information. Our research objective is to establish an automatic mechanism to evaluate trust when relevant information is provided on the web. More precisely, the main contribution of this study is to assess UGC in terms of trust. Existing research in this context is limited in addressing trust. Instead, the focus is on quality, readability, credibility, etc., usually single-handedly. Moreover, these concerns are more or less linked to trust, however in many cases they are mixed together. There is hardly any clear dividing line between them or explanation of how these metrics interact in respect of trust. This work aims to identify and clarify this unclear relationship by examining such concerns and providing a trust model that combines relevant aspects. This phase of our work in progress focuses on the metadata of UGC and not on the content itself, which requires specific techniques for text mining. Therefore, the platform under investigation should have a higher frequency of user interaction than platforms dealing with contributions of relatively long texts (e.g. Wikipedia). Platform on which it is known that the content generated contains "bad" information that was intentionally made. This kind of content, known as fake-news, is another area of research. Any kind of investigation of objectives and personality of a contributor is difficult and requires different approaches. For these reasons, we have chosen to consider Genius as a case study in which users have less interest in intentionally manipulating information. Nevertheless, our proposed model should be able to be integrated into these and other communities that provide the necessary input.

The rest of the paper is structured as follows: It begins with a brief overview of Related Work section. To describe the approach in more detail, a Domain Analysis section is presented, which uses the Genius platform and a data set to build annotation-based knowledge.

The Trust section includes a definition of trust. The following Trust Model Construction section introduces the trust dimensions used in our approach and their backgrounds and derivations. The results are presented in Results section and Discussion section, and finally our conclusions and suggestions for further work are offered in the Conclusion section.

6.3 Related Work

There are several studies on trust that consider users' behavior on the retrieval of information. Using various approaches, researchers have proposed methods to identify vandals in some cases, or to predict trust in others [41, 53] by verifying the history of the generated content and reputation. These aspects are relevant to value content; however, the aim of these works is to investigate trust as an opposite of vandalism. Javanmardi and Lopes [32] present a Wiki Trust Model (WTM) based on the Hidden Markov Model (HMM) for platforms based on wiki technology (i.e., Wikipedia). WTM can be used for identifying vandals, as well as contributors with high quality. It can be integrated to a wiki to improve information reliability, and for automatic detection and restriction of vandals. Figueiredo et al. [23] propose a software architecture framework to support a trust model of community members and the construction of vocabulary with the focus on online communities. The calculated members' trust is composed of artifacts and parameters specified by the community itself. A trust factor can be quantitative (number of contributions) or qualitative defined by the community. When predicting trust, one must infer unknown property from known property, for example mapping trust of a known community over an unknown one. The mentioned works put mainly the contributor in light and show the importance of considering his reputation by valuating content. However, user trust differs from information trust that this work focuses on. Since, user trust as discussed neglects the fact of context influence, that so called untrust, which means an entity could be trusted in a specific situation to perform a specific action not in another.

Dondio et al. [22] propose a Wikipedia Trust Calculator (WTC) that consists of Data Retrieval module contains the needed data of an article, Factors Calculator module calculates and merges the trust factors into the macro-areas defined and Trust Evaluator module calculate a numeric trust value and judge it in a natural language explanation using constraints provided by a Logic Conditions module. The authors suggest the function:

$$N(t) : t \to \tau$$

that returns the number of edits done at time t, which is used by the stability function

$$E(t) = \sum_{t}^{p} N(t)$$

that calculates the number of edits done from a given time t to the present time p. Meanwhile, stability is defined as "only active and articles with good text can be considered stable". This work considers a part of edits' contributors as "n% top active users", which is calculated by the function they called "Users' Distribution/Leadership

$$P(n) = \sum_{U_a} E(u)$$

with U_a the set of n% top active users and "E(u) as the number of edits for user u for a specific article". Part of our work takes over the stability function presented in this paper and is inspired by the idea to consider the n-top-active user. Lucassen and Schraagen [43] propose a model they call the 3S-model that considers user judgments about trust in information. There are three user characteristics: Source experience, professional competence and information literacy. Their work is based on the assumption that these characteristics lead to different features, i.e. source semantic and surface features, of the information that is used in trust judgments by different users and in different contexts. The authors suggest that users can alternatively rely on their previous experience (e.g. authority, Website) with a particular source rather than actively evaluating various features (e.g. accuracy, completeness or length, references etc.) of information to judge confidence. Mayer et al. [48] propose a model of trust that distinguishes clearly between trustor and trustee and contains components related to both. The trustor's trait is referred to the trustor's propensity to trust. Despite a trait is relative, since people differ in their inherent tendency to trust, it is proposed in the model to be stable. The characteristics of the trustee are represented in the model by the concept of trustworthiness, which consists of factors that are 1) "ability" e.g. skills or competence, 2) "benevolence" means having specific attachment to the trustor, and 3) "integrity" means following some set of principles that the trustor finds acceptable. The authors assert that these attributes of trustee are the reason why a trustor has more or less much trust for a trustee. Trust for a trustee will be a function of these attributes and together with the trustor's propensity will help to create the basis for the development of trust, but a decision to trust is not yet been made. According to trust definitions, the model lacks the important aspect that is risk. Kelton et al. [35] extend the model proposed by Mayer et al. to consider additional aspects i.e. the necessary preconditions for trust, the influence of context and social trust, and the role of trust development processes. The preconditions are "uncertainty" and "vulnerability"; i.e. when the trustor faces a risk and when there is a status of "dependence", which concerns two matters between trustor and trustee: the first has a special need to meet and the last has the potential to satisfy this need. Trust model of Abdul-Rahman and Hailes [1] is based on Marsh' model [46]. It deals with sociological characteristics and trust beliefs between agents based on experience (of trustor self) and reputation (comes from recommended agent). These are combined to build trust opinion to make a decision for the interaction with the provided information. This work deals with agents, our work was inspired by the idea of trust degree translator, which we adopt and extend by a threshold Table 6.3. The work of Ruth et al. [33] -page 183- designed a model of social trust for users, which is similar to the model introduced by Abdul-Rahman and Hailes. The authors focus on users that are agents in terms to determine the validation probability of a given data. The proposed model considers the categories "very trustworthy", "trustworthy", "untrustworthy" and "very untrustworthy" a user or an agent assigned to. This classification is derived from Abdul-Rahman's and Hailes' work and relies on first hand experience of a user; e.g. "if user A has 4 very trustworthy experience and 5 trustworthy experience with user B, the A applies the category trustworthy to B". This example considers past recommendations and the resulting experience in terms to find the semantic differences between them. Castelfranchi et al. [15] suggest a socio-cognitive trust model using fuzzy logic with the aim to analyse the different nature of the belief sources and their trustworthiness. The model consists of four bottom-up layers: 1) The first layer includes "beliefs sources" e.g. "who/what the source is -to evaluate-", "direct experience" e.g. "In my experience", "reputation" e.g. "A friend says that ..", "categorization" e.g. "usually doctors ..." and "reasoning" e.g. "I can infer that ...". 2) The second layer contains ability, accessibility, harmfulness, opportunities and danger as the relevant basis beliefs. 3) The third layer includes internal factors i.e. ability, accessibility and harmfulness and external factors i.e. opportunity and danger. 4) The fourth layer combines the internal and external factors into one component that influences trustfulness. The model introduces a degree of trust relied on the credibility of the trust beliefs, and its implementation allows changing the components according to the situation and agent personality.

6.4 Domain Analysis

Domain analysis is used to identify records and to develop instances of a system (this is Genius in our case) to be used in the application or family domain (these are social media in our case) [18]).

The instances to be examined are those that should be related to trust. Trust has positive effects on knowledge sharing behaviour, on enhanced relationship [29, 38, 62], particularly in social media that increase user inter-activities [34, 56] and influence building opinions [5, 6, 40, 55] and are the most important information sources [69]. The social media Genius supports the creation of knowledge-based annotation. The collaboration of Genius members in text editing needs to be discovered. To this end, we will continue to clarify, at least in part[4], the architecture of Genius' social participation, such as annotation and membership privileges. This section describes Genius, which is our case study.

6.4.1 Genius

Genius is a part of Social Media (SM) and enables users to publish their own or publicly available texts, on which users generate interpretations in form of annotations. Evaluations of an annotation are done using voting and editing mainly. Activities are rewarded with points calculated on a user's credit account. This credit is referred to as a user's Intelligence Quotient (IQ). In addition, this corresponds to the content created and its evaluation by other users. While the IQ counter of an annotation represents the sum of the votes up and down for this annotation and reflects its degree of user acceptance.

6.4.2 Annotation

Annotations are placeholder of interpretations and provide a set of metadata and options as illustrated in Figure 6.1. On the right side of the figure an annotation takes place, while the left side shows the original text (lyric) containing a piece of text (highlighted in yellow), to which the annotation is related. We can see the metadata of the annotation such as the number of contributors, who they are, their role (the symbols right to each contributor name), their IQ count, their proportion (percentages), by whom it is created and accepted as well as the number of edits (view all edits). At the bottom and in addition to the annotation's IQ count (+41), a set of options (i.e. Up-/down vote, follow, share, create a suggestion to improve the annotation) a registered user can carry out.

[4] A detailed explanation can be found in our Genius technical report.

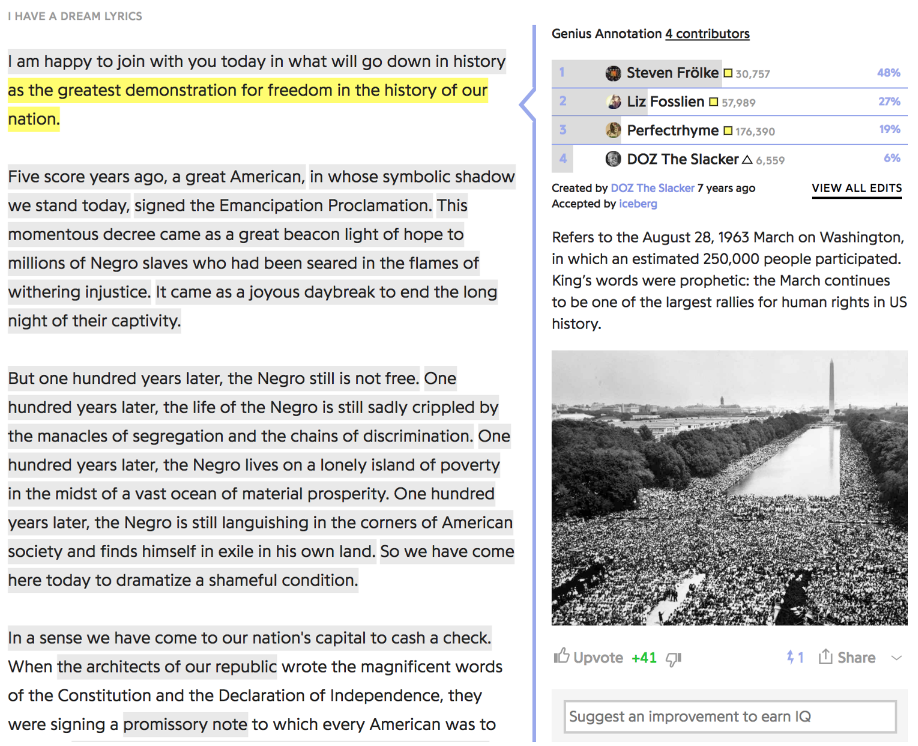


Figure 6.1: Annotation Example

Screenshot:https://genius.com/544987 accessed: 2018-08-12 at 19.51.06
This figure represents an example of an annotation (right side) of a lyric (left side) on Genius

Annotations represent extended information to the user-generated content in the form of additional contribution. According to Genius a high-quality annotation is a contribution that is error free in grammar and contains solid knowledge.

6.4.3 Member Roles

Genius members have six different roles that are closely linked to authorizations sequentially: Whitehat, Verified Artist, Editor, Moderator, Mediator and Staff. Intelligence Quotient (IQ) can be earned through most activities and indicates a member's experience. A count of IQs is required to perform certain activities, such as accepting an annotation. However, it does not automatically promote the role of a member to reach a higher membership level. High-quality annotations and a decision-maker such as an Editor establish nomination criteria for such role promotion.

6.4.4 Role Permissions

Each role is assigned a color that identifies the role, and the IQ count does not only provide a decisive factor. Thus, it is possible for a user to have a higher IQ count than the IQ count of another user of a higher role. However, the permissions differ between roles, but also between users within a role. Some users, due to their IQ, have more permissions than their role equivalents. Based on our analysis, we found that Staff has all rights 38/38 (100%), Mediator has many similar rights 36/38 (94.73%), Editor has 25/38 (65.78%), Mediator has 18/38 (47.36%), Artist has 4/38 (10:52%) and Whitehat has 3/38 (7.89%) [2]. This distribution of rights determines the control of the Staffs and Moderators and alludes to the fact that Whitehats, despite their large number, have little permission and therefore hardly influence on Genius.

6.4.5 Edit Types

Equivalent to the differentiation between different users, a distinction is to be made according to the types of edits on an annotation. Thus, we use the classification in [2] lightweight peer production (LWPP) and heavyweight peer production (HWPP) proposed by [30], which is used to differentiate participants' contributions (see Table 6.2). LWPP embodies simple and independent contribution interactions and its simplicity enables a high number of attendees. In contrast to HWPP, this is more extensive and time-consuming, but at the same time it provides more information about content and user. However, this information can be used for further analysis, which is an advantage of this approach. In these collaborative forms, the user's participation is based on the complexity and dependency of his interaction, which is characterized by weak-tie attachments and strong-tie attachments. Weak-ties are simple enough for participation, while strong-ties require agility and more experience from the participants that usually generates credible and high-quality content [30]. This is confirmed by Genius and can be seen in the distinction of the IQ count towards activities (edit types) as shown in Table 6.1.

Table 6.1: Activity and Earned IQs

activity	earned IQs
First profile picture added	+100
Annotation upvoted	+2[5]
Suggestion upvoted	+0.5
Suggestion downvoted	-0.5
Annotation downvoted	-1
Forum post upvoted	+0.5
Forum post downvoted	-0.5
Transcribing a song	0
Creating a description	+5
Write an annotation	+5
Annotation accepted	+10
Annotation rejected	-7[6]
Suggestion integrated	+2
Suggestion archived	0

Table 6.1 shows the quantities of received IQs (positive and negative) of several activities. The activities are described in the Genius technical report [3].

By consolidation of Table 6.1 and Table 6.2, a numeric factor between LWPP- and HWPP edits is specified. For example, 0.5 IQ is given for upvoting (LWPP) an annotation, while 5 IQs are earned for writing (HWPP) an annotation. The factor is 1 to 10 for HWPP compared to LWPP.

Table 6.2: Classification of Activities on Genius [2]

	Predicate	Object
Lightweight	upvoted downvoted accepted marked rejected deleted archived cosigned incorporated	annotation suggestion description comment Song Page user
	followed	Song Page, user
	pyonged	description, annotation, Song Page
Heavyweight	created edited mentioned merged integrated replied	annotation description Song Page metadata
	added proposed	suggestion, reply edit, comment

This table illustrates the predicates of the Genius interaction design, which are classified into LWPP and HWPP [2] and grouped separately by a horizontal line. For example: The predicate *followed* (LWPP from the second group) may be combined with the objects *Song Page* or *user*, but not with the object *comment* (LWPP from the first group).

6.4.6 Data Set

In this study, we use data observed on Genius Firehose. Firehose is a subpage of Genius, which can only be accessed by registered users. All activities on Genius can be tracked in real time. Since the user interactions of the community at this point are already in the correct chronological succession and in an approximately machine-readable form (XML). This functionality is the central starting point for data collection. Firehose activity consists of contributors' name, activity type, subject, symbol of type, and time stamp. Clicking on an activity displays its details. We extract such details and follow links provided to get further information about the activity and the user. The data collection comprises 1,516,829 activities, of which 1,257,555 annotation activities are performed on 419,048 individual annotations generated by 499,338 users, consisting of 166,417 individual users. Our focus in this study is on the users and the editing of annotations.

6.4.7 Empirical Cumulative Distribution Function

The trust distribution was analysed on various dimensions on the basis of the Empirical Cumulative Distribution Function (ECDF) , and is shown in the Figure 6.2, Figure 6.3 and Figure 6.4. We take into account the user IQs based on the user role and the attribution in addition to the IQ count, the edits IQs correlated with the edit types and the number of the edits on the annotations. Figure 6.2 shows 87.5% active users with up to 1000 IQs and 43.7% with less than 0 IQs. As reminder, these numbers of IQs are refined by the roles and the attributions. Though, in Figure 6.3 are 37.5% edits having less than 0 IQs. As a reminder, these numbers of IQs are refined by the edit types. With 87.5% of edits, IQs will rise by about 35. In Figure 6.4 there are 37.5% of the annotations with less than 0 edits, which means, they were deleted or rejected. At 87.5% of the annotation, the number of edits is higher than 5 and at each point where the values raise, we can use the trust degree translator in Table 6.3 to classify it as very trustworthy (vt). These values include edits number 5, edits IQ 35 and user IQ 1000. At the point where the edits IQ is less than 0 (37.5%), the annotations have gained less than 2 edits (deleted or rejected), which were generated by users having IQ count less than 0 (about -100). This is the area of annotations that is classified as very untrustworthy (vu). In the relatively small area (6.25%) of untrustworthy (ut) users have IQ count less than 0 but higher than -100, edits IQs less than 5 and higher than 2 edits. The highest percentage hold the trustworthy (t) with 43.75%, in which users have IQ count of less than 1000, edits IQs count of less than 35 and less than 5 edits. Table 6.5 shows and summarizes the groups of the annotations observed on Genius over a certain time period. Each group is illustrated by: its trust degree, its percentage to the whole, how many edits it has, how many IQs it earned and what is the IQs count of its users or rather contributors.

Figure 6.2: Users IQs Distribution based on ECDF

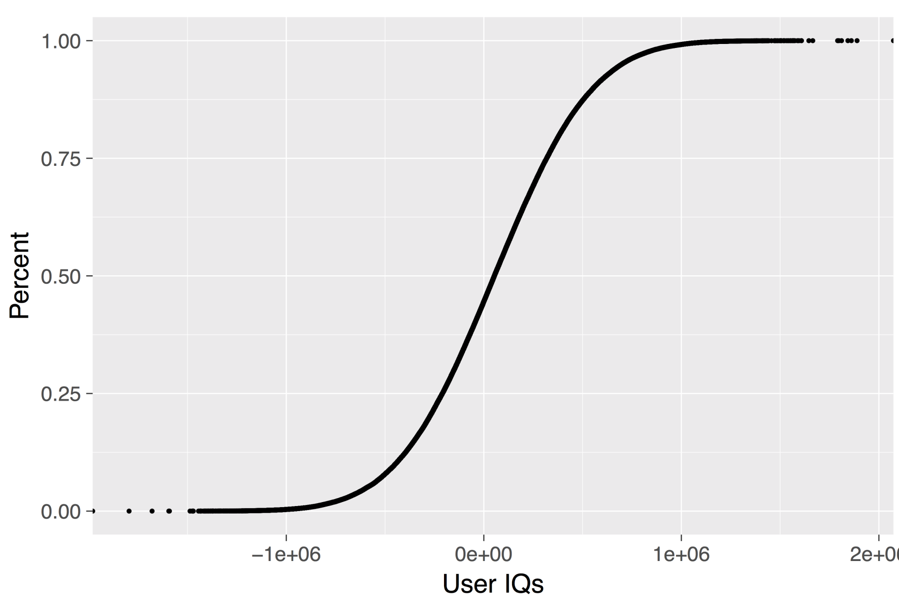


Figure 6.3: Edits IQs Distribution based on ECDF

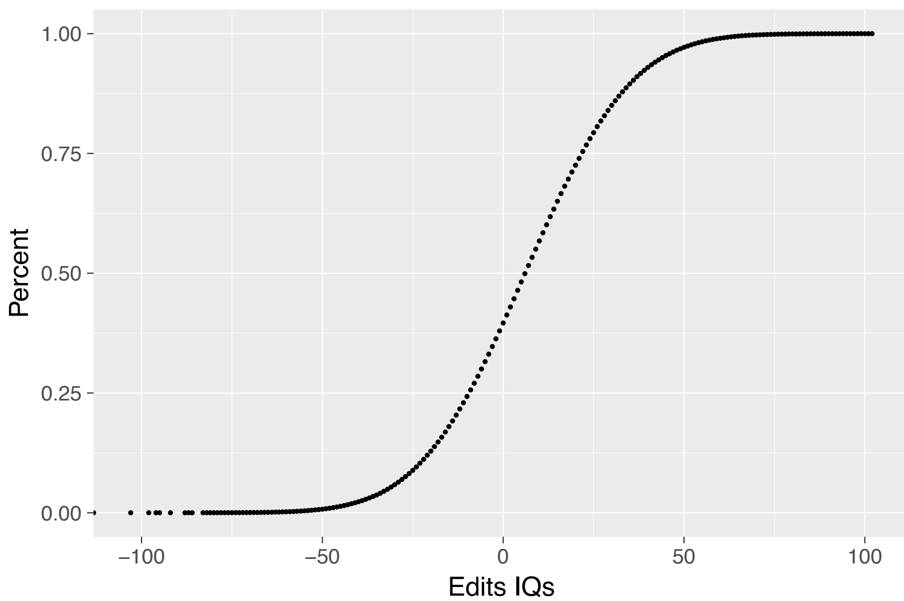
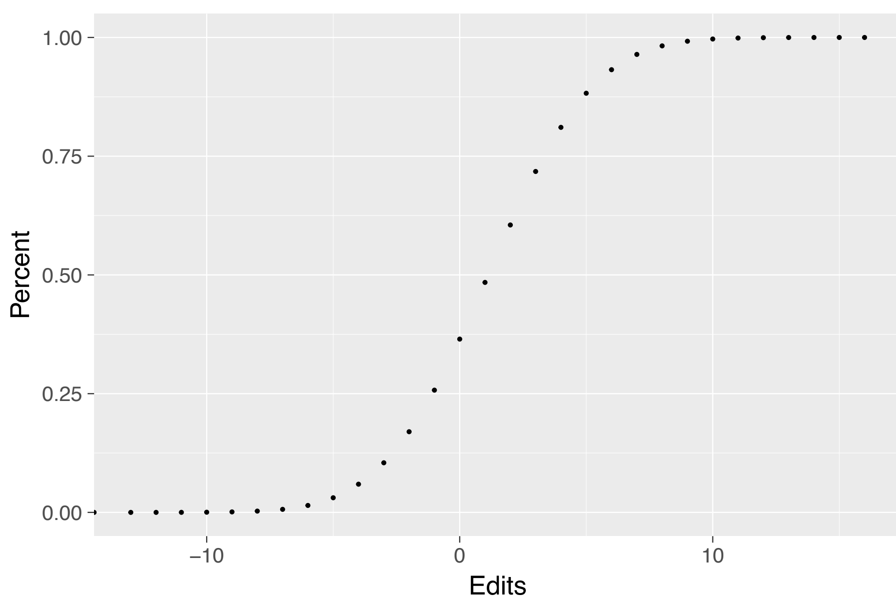


Figure 6.4: Edits Distribution based on ECDF



These Figures 6.2, 6.3 and 6.4 show the distributions of the various dimensions of the trust analyzed using ECDF.

6.5 Trust

This section reviews existing definitions of trust and derives the required criteria for defining a reliable trust model.

Table 6.3: Data distribution based on ECDF and Trust Degree Translator

Trust Degree	Percentage	Edits number	Edits IQ	User IQ
Very Trusted (VT)	25%	>5	>35	>1000
Trusted (T)	31.3%	2 to 5	5 to 35	0 to 1000
Untrusted (U)	6.25%	0 to 2	0 to 5	-100 to 0
Very Untrusted (VU)	37.5%	<0	<0	<-100

This Table shows the groups of the annotations observed on Genius over a certain time period. Each group is illustrated: Its trust degree, its percentage to the whole, how many edits it has, how many IQs it earned and what is the IQs count of its users or rather contributors.

Trust Definition

Trust as a personal merit depends on a context that changes over time Bansal et al. [11]. Trust between two parties is based on the reliability and integrity they share with each other Morgan and Hunt [54]. Unfortunately, trust is a complex concept and a generally accepted definition does not exist [12]. We value the definitions of trust proposed by Mayer et al. [48] "The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" and by Corritore et al. [19] "An attitude of confident expectation in an online situation of risk that one's vulnerabilities will not be exploited". Both definitions contain common aspects: Willingness, attitude, vulnerability, two parties (trustor and trustee) as well as importance and risk. Mayer et al. [48] mention the ability to monitor or control the trustee. This is contained in the term "in an online situation" by Corritore et al. [19]. The willingness and attitude of a trustee not to exploit vulnerability of a trustor can vary. However, this is further enhanced by the review (stability) and evaluation (credibility) of a user generated content by other users [17]. This study considers trustworthiness as a property of an object (i.e. content). While trust is the process performed by an entity (i.e. user) including interaction on that object with respect of vulnerability. The result of that process is the trustworthiness, which is indicated by user's activities and assessed by three concerns: (1) the number of activities conducted over a time period by other users (stability). (2) The types of such activities, users' review and authorship (credibility) and (3) the nature of the content generated by the members of the elite-cycle or n-top-active users (quality). Based on this analysis and from this study view of point, trust is defined as a correlation function of the dimensions stability, credibility and quality.

6.6 Trust Model Construction

The synthesis of our approach is based on dimensions that are derived from various studies in literature with context similarity to Genius. Such dimensions are modified and merged so that users' trust in information provided can be considered and calculated. The mechanism developed is derived mainly from approaches of the Wikipedia Trust Calculator (WTC) of Dondio et al. [22] (calculation of the dimensions) and also from the model of Abdul-Rahman and Hailes [1] (interpretation of the calculated results). We adopt the terms: (1)

Stability as proposed in Dondio et al. [22] meaning the growth of the contributions rating instead of the difference of the article version history. (2) Credibility that is defined in Metzger et al. [51]; Pranata and Susilo [59] as the believability in provided information. This information is characterized by the most popular raters, also known as authority and reputation in Warncke-Wang et al. [68]. (3) Quality, introduced in Warncke-Wang et al. [68] as an entirely objective assessment against a reference standard in domains, where critical decisions rely on it. Quality differs in Wikipedia context across place and time [27], which influences trust and can be estimated by user and content (goodness) [25]. Stability is presented by the annotation edits' distance over a period of time. According to Kittur et al. [37] stability has significant impact on user's trust. We calculate the edits' distance in an annotation, since these edits include significant features such as rating, author's attributes and so on. The case of Wikipedia, from which this part of calculation is derived, is different; the article is the research object (in our case, it is an annotation) and the distance is calculated by the change in text over its versions. Credibility is defined by Flanagin and Metzger [24] and Metzger and Flanagin [50] with criteria accuracy, authorship, objectivity, coverage and currency as shown in Table 6.4. The authors see that it is necessary to determine the correctness (accuracy), who the author is (authorship), what his aim is (objectivity), the degree and depth in meaning (coverage) and the up-to-dateness of the provided information (currency). Accuracy, authority, and coverage are reflected in the Intelligence Quotient (IQ) of users and edits. That is, the user's IQ rating indicates the experience required to authorize and accept content. While the edits' IQ rating represents the agreements of the reader on the edits' accuracy and coverage. Objectivity and currency are not represented in this dimension, as neither the collected data set nor our approach is intended to measure them. However, until now there is no scientific evidence that ratings provided by users are always credible and trustworthy [59]. Nevertheless, we are prone to trust interactions with the most popular users (Tucker [66]; Pranata and Susilo [59]). Quality includes the most active user for a given number of n-top active users and their weighted edit types. The weighting is derived from the number of earned IQs (see Table 6.1) for specific activities (see Table 6.2). This consideration gives a direction for the distinction between annotations based on the quality of certain contributors, which could not be implied in the credibility dimension by the edits IQs calculation, since a specific user could be lost in the crowd of editors. Such an n-top active user will help to detect annotation patterns in further works. Table 6.2 illustrates the predicates of the Genius interaction design, which are classified into LWPP and HWPP and grouped separately by a horizontal line [2]. For example: The predicate followed (LWPP from the second group) may combine with the objects Song Page or user, but not with the object comment (LWPP from the first group). Table 6.4 describes the dimensions: stability, credibility and quality derived from Dondio et al., [22], Flanagin and Metzger [24] and (Warncke-Wang et al.[68] and Gamble and Goble [27]), respectively. This table summarizes the represented dimensions and describes their factors applied at calculation.

Table 6.4: Description of trust Dimension

<table>
<tr><th>Dimension</th><th colspan="3">Description</th></tr>
<tr><td>Stability</td><td colspan="3">Calculated by the annotation edit distance over time period</td></tr>
<tr><td rowspan="3">Credibility</td><td rowspan="3">Calculated by rating, attribution , user role and user IQ</td><td colspan="2">Criteria that are</td></tr>
<tr><td>regarded</td><td>Not regarded</td></tr>
<tr><td>Accuracy, authority and coverage</td><td>Objectivity and currency</td></tr>
<tr><td>Quality</td><td colspan="3">Calculated by the n%-set of the top most active users and the weighted edit type based on LWPP, HWPP and the earned IQ with regard to that edit type that are classified in the Table 6.2.</td></tr>
</table>

Table 6.4 describes the dimensions: Stability, credibility and quality derived from [22],[24] and ([27, 68]), respectively.

Trust Dimension Calculation

Stability (S) is presented by the annotation edits' distance, which includes the function (E(t)) that specifies the number of edits at a given time stamp t:

$$\{E(t) : t \rightarrow \Phi \quad | E : edits\, function, t : time\, stamp, \Phi \in \mathbb{Z}\} \tag{6.1}$$

Here, $\mathbb{Z}$ is a set of all integers.

$$\{S = \sum_{t=t_0}^{t=p} E(t) \quad | t \ \& \ p : time\, stamp\} \tag{6.2}$$

Annotation consists of various edit types (HWPP and LWPP as described in Seciton Doman Analysis) performed by different users. On the basis of this diversity of users, it is necessary to differentiate between their edits on the basis of their edit percentage, roles, and IQ counts (annotation's IQ and user's IQ). Due to this differentiation, we calculate a User Credibility Correction Factor (UCCF) consisting of attribution[7], role power and user IQs. Role power is derived from the permissions of each user role as introduced in Section Domain Analysis. This pre-calculation reflects the user investment in an annotation's edit more thoroughly than the sole user IQ.

In the definition given below, we use the symbol e to refer to an editor. *editors* are the group of users who have edited the annotation in consideration. It applies to each e (editor) from *editors* so that the $UCCF$ of e is extended by *attribution*, *rolepower* and IQ of this e.

$$\forall e \in editors\{UCCF+ = IQ_{user} \times attribution \times rolepower\} \tag{6.3}$$

Similarly, to the extension of the user's IQ by considering the attribution and role power, and as described in Section Domain Analysis, there are complex activities (e.g., annotation

[7] An Attribution is the percentage of user's edit consignment.

creation) called HWPP that require agility from the user during execution. These are also ranked higher in Genius than simple activities that are done by click (e.g. up vote), the so-called LWPP. This differentiation allows weighting of edits according to their types and logically, leads to a higher accuracy of the calculation than the calculation based on annotation's IQ only.

$$editsTypes = IQ_{annotation} \times (|HWPP| \div |LWPP|) \tag{6.4}$$

IQ is the count of IQs of annotation. $|HWPP|$ and $|LWPP|$ are the counts of the edits of the annotation, which are distinguished on the basis of HWPP and LWPP. Thus, the credibility is a function of $UCCF$ and $editsTypes$ and reads as:

$$credibility = f(UCCF, editsTypes) \tag{6.5}$$

$$f \text{ is the arithmetic mean function.} \tag{6.6}$$

Remember: The members of the elite cycle or n-top-active users are experienced contributors who usually create high quality content, which reduces the risk of performing a particular action in an online situation. This is important for the trustor and for updating his trust. This is why we pay special attention to this group.

quality is a function of $UCCF$ and $editsTypes$ of the n-top active users.

$$quality = f(UCCF_{n-topActiveUser}, editsTypes_{n-topActiveUser}) \tag{6.7}$$

$UCCF_{n-topActiveUser}$ this an extension of User Credibility Correction Factor. This extension includes the $n-top$ active users, where n is to be determined by the viewer. $editsTypes_{n-topActiveUser}$ refers to the editsTypes of the $n-top$ active users.

By dividing the trust value equally across these dimensions, our mechanism calculates a trust value from an annotation. The distinction between these dimensions and their impact on trust is reflected in the calculation.

$$trustworthiness = f(stability, credibility, quality) \tag{6.8}$$

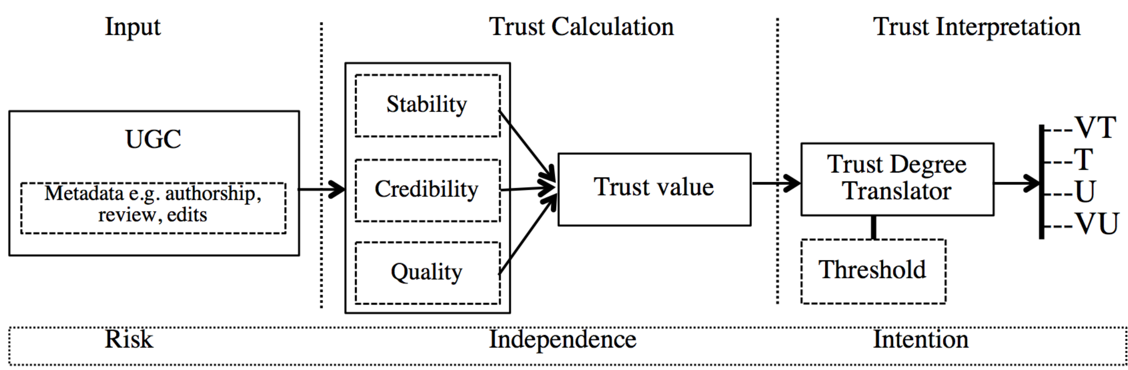
trustworthiness is a function (see Definition 6.6) of three factors which are *stability*, *credibility* and *quality*.

To provide a human-readable result, the trust degree translator in Table 6.3 uses a threshold table to map the numerically calculated value into one of the trust classes.

6.7 Results

The theory behind this study is that in the domain of user-generated content, it should be differentiate between content-change (growth, reduction) and content-evaluation. Content-change indicates the degree of user interest, which could be more than a passing fancy, in case of many activities. Whereas, content-evaluation is performed on two levels: "Normal" user level and elite-member level. Elite-members tend to be more (double as much) trusted and whose reviews are considered as more significant ([66, 67]). The trust model proposed in this paper is presented in Figure 6.5. It consists of four stages: 1) Input, which contains user-generated content to be examined and a set of metrics consisting of metadata (e.g. rating, authorship, etc.). These metrics are considered in the next step; 2) compilation unit that uses the equations (1 to 8) to calculate a numeric trust value based on the metrics. The calculated trust value is passed into the 3) interpretation step that applies a predefined threshold table to classify the value and maps in the 4) outputs into one of the trust classes (very trustworthy, trustworthy, trustworthy, unreliable, and very trustworthy). The ECDF is used to calculate the threshold table by examining the distribution of the data. However, any distribution function or other technique (e.g. k-mean) can also be used.

Figure 6.5: Trust Model



6.8 Discussion

Higher-level contributors to Genius are generally trustworthy, since they have proven their ability to generate high quality annotations that are identified by Genius as well written, error-free in grammar and contain solid knowledge. These aspects are considered and examined by the trust dimensions: stability, credibility and quality.

During our observation, we found out that the positive[8] edits number of most annotations is small, but some of them have a higher stability with more edits. High credibility and high quality are assumed for the same group. It seems logical, but not necessarily true. This is confirmed by comparing with the annotations IQs and credibility. The picture does not differ from the previous one, where annotations with high IQ rates are the same as annotations, on which a high IQ and a high role user interacts, who are involved with high attribution as well. That is, that a high voting count of an annotation does not describe an overall picture of trustworthiness. The same case is expected for the quality.

Based on the analysis of the user groups' permissions, we calculate a factor[9] as role power, since a contribution generated by an expert differs qualitatively from that of a non-expert. Additionally, we take into account the attribution, which gives how much proportionally a user contribution is. These factors are combined with the user IQ and

[8]Edits that include activities like upvote, accept and merge, negative edits include activities like down-vote, reject and delete.

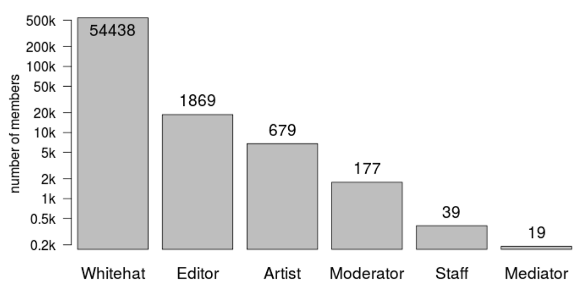
[9]0.025 computed as 1 divided over 40, the number of the observed permissions.

User Credibility Correction Factor (UCCF), which reflects more depth in sense than the sole user IQ. There are users with a higher count of IQs, but a lower user role.

The Mediator's and Regulator's responsibilities imply organizational issues like coaching and promoting users and their focus is less on generating content. Editors and Moderators, whose influence focuses on editing as they have the relevant permissions, compared to Whitehats, who mainly generate content, give equal participation. This in depth view reconfirms our assertion that the sole voting presented in IQs of users and annotations does not provide a complete picture of an annotation's quality. This has been asserted in our previous work [2].

In Figure 6.6 we can see that 95.13% of the users are whitehats, followed by editors with 3.26% (1.18% Artist, 0.30% Moderator, 0.06% Staff and 0.03% Mediators). This suggests that Genius is relatively young, but is growing very fast, or that there is an indication that promotion to a higher role is difficult, or both. But who is most active with Genius? Figure 6.7 confirms Figure 6.6 and shows at first glance that the most active users are Whitehats, who produce 61.14% of the annotations. Second and third are Editors with 18.28% and Moderators with 10.87%, while the rest are Staff (6.56%), Artists (2.64%) and Mediators (0.48%). This is the distribution for each role as a group, but in order to make a statement about the active participation and productivity among users in different roles, the relationships of each person in each group to the group itself must be calculated and compared with the previous results. 466,448 comments were made by the Whitehat role group, which contains 54,438 users, which corresponds to 8.56 comments per Whitehat. We have carried out the same calculation for each group and get 74.64 per Editor, 122.19 per Moderator, 282.82 per Staff, 517.56 per Artist and 193.94 per Mediator, as shown in Table 6.5.

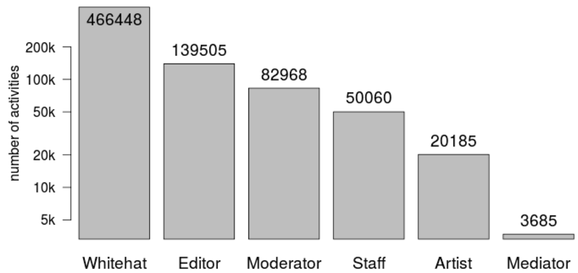
Figure 6.6: Number of Users per Roles



This figure shows users' distribution over the different roles.

The cross-comparison shown in Table 6.5 of generated annotations by users of each role, based on the period of analysis, appears to give a different view of the distribution and represents the active participation of a single user from the different groups in the sequence: Artists, Staff, Mediators, Moderators, Editors and at the end are Whitehats, which is reasonable to expect if we consider the type of allowed activities associated with users permissions. Collaboration between the users provides high-quality User-Generated-Content (UGC). Typically, a small number of users generate a large amount of annotation activities, and vice versa. Table 6.5 shows the association roles to the number of the generated annotations by the number of the users derived from the Figure 6.6 and Figure 6.7 based on the period of analysis [2].

Figure 6.7: Roles and Activities



This figure shows the distribution of the annotation activities over each group of the different roles

Table 6.5: Generated Annotations by Users/Role [2]

Role	Annotations	Users
Whitehat	466,448	54,438
Editor	139,505	1,869
Moderator	82,968	679
Staff	50,060	177
Artist	20,185	39
Mediator	3,685	19

This table shows the association roles to the number of the generated annotations [2] by the number of the users derived from the Figures 6.7 and 6.6 based on the period of analysis.

Managerial Implications

Nowadays, it is easy to find information, especially recommendations, about almost everything on the web. This remote service usually has a credibility problem. It is not always easy to assess the quality of the information provided. At the same time, the lack of known sources of information has made it more difficult to interact (consume) with such information. To meet this challenge, it is necessary to rely on alternative strategy that is available within this information. Our trust model uses the metadata of the information to simplify decision-making, identify and consume trusted information. The input of the trust model is the metrics (e.g. number of comments, reader rating, author rating, etc.) of information in the form of annotation on social media, which are calculated together, and the result is classified. The user receives a human-readable interpretation of the result that the annotation can be trusted or not. Factors that affect trust must be understood in order to develop successful applications that inspire trust and in which users are willing to participate.

The proposed trust model supports the identification of trusted information in collaborative environments. It can be used in various online communities that provide the appropriate metadata for the information provided. The trust model helps to filter the information and thus reduces the information overload shared on the Web. Applications can integrate the trust model into their development to increase the likelihood of their use, as users are able to easily identify trusted information. This work is aimed at promoting valuable knowledge sharing by improving application development using the proposed trust model. Thus, the model serves as a reference for the development of collaborative annotation applications, as we have shown that trust plays an important role as a bridge between information quality and information usage.

The trust model has broad dimensions and is based on assessments and metrics derived from in-depth literature research. By separating the components of its mechanism, it is variable and flexible in its attitude to include additional metrics that are relevant for integration into other but similar areas.

6.9 Conclusion

We analyzed Genius and presented concrete statistics on the distribution of activities created over a time span. These statistics have shown that Genius is still at a nascent stage, but it is growing very rapidly. This conclusion is confirmed by the number of Whitehats, as mostly new comers, in comparison to the other roles. We demonstrated trust in annotations in three dimensions: stability, credibility, and quality derived from the literature and mapped into Genius context based on data analysis. The number of edits of annotation versions calculates the stability dimension during a time span. Credibility dimension builds on accuracy, authority and coverage, which are reflected by the IQ ratings of users and edits. User IQ is corrected by the UCCF that is calculated by the attribution, role power, and user IQ. UCCF adds value to user IQ by taking into account who the user is (role) and how much they contribute (attribution). The quality dimension consists of the most active users and the edits type of an annotation. The edits types were weighted based on the classification LWPP and HWPP proposed by Haythornthwaite [30], as well as the number of IQ points awarded from Genius for various activities. We could propose a theoretical trust model that can measure trust in Genius as a collaborative environment. By considering Table 6.3, we can see the higher the trust degree is, the higher the number of edits. This means that trust increases interaction between users. In the context of Genius, the annotations are mainly divided into trusted or very untrusted. This indicates the existence of a specific personal trait and social response, which should be investigated in further works.

The proposed model can be integrated into other communities that provide the necessary input. In another paper [4], we evaluated the trust model based on user preferences, which confirmed its construction.

Finally, this work will form the basis for further research on clustering the trust objects. In our case, these objects are the annotations, to obtain particular pattern and its features. We answer the question of what makes an annotation having such characteristics to be very trusted, trusted, untrusted and very untrusted? We may propose a template that describes how to or not to provide information, so that a trustor can be motivated to consume the information in making decisions.

Bibliography

[1] A. Abdul-Rahman and S. Hailes, *Supporting trust in virtual communities*, in System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on, IEEE, 2000, pp. 9–pp.

[2] J. Al Qundus, *Generating trust in collaborative annotation environments*, in Proceedings of the 12th International Symposium on Open Collaboration Companion, ACM, 2016, p. 3.

[3] J. Al Qundus, *Technical analysis of the social media platform genius*, tech. rep., Freie Universität Berlin, 03 2018.

[4] J. Al Qundus and A. Paschke, *Investigating the effect of attributes on user trust in social media*, in International Conference on Database and Expert Systems Applications, Springer, 2018, pp. 278–288.

[5] A. M. Aladwani and Y. K. Dwivedi, *Towards a theory of sociocitizenry: Quality anticipation, trust configuration, and approved adaptation of governmental social media*, International Journal of Information Management, 43 (2018), pp. 261–272.

[6] A. A. Alalwan, *Investigating the impact of social media advertising features on customer purchase intention*, International Journal of Information Management, 42 (2018), pp. 65–77.

[7] L. Alzahrani, W. Al-Karaghouli, and V. Weerakkody, *Investigating the impact of citizens' trust toward the successful adoption of e-government: A multigroup analysis of gender, age, and internet experience*, Information Systems Management, 35 (2018), pp. 124–146.

[8] T. M. Amabile, C. Patterson, J. Mueller, T. Wojcik, P. W. Odomirok, M. Marsh, and S. J. Kramer, *Academic-practitioner collaboration in management research: A case of cross-profession collaboration*, Academy of Management Journal, 44 (2001), pp. 418–431.

[9] D. Artz and Y. Gil, *A survey of trust in computer science and the semantic web*, Web Semantics: Science, Services and Agents on the World Wide Web, 5 (2007), pp. 58–71.

[10] A. Baier, *Trust and antitrust*, Ethics, 96 (1986), pp. 231–260.

[11] G. Bansal, F. M. Zahedi, and D. Gefen, *Do context and personality matter? trust and privacy concerns in disclosing private information online*, Information & Management, 53 (2016), pp. 1–21.

[12] A. Beldad, M. De Jong, and M. Steehouder, *How shall i trust the faceless and the intangible? a literature review on the antecedents of online trust*, Computers in Human Behavior, 26 (2010), pp. 857–869.

[13] T. Berners-Lee, J. Hendler, O. Lassila, et al., *The semantic web*, Scientific american, 284 (2001), pp. 28–37.

[14] M. Biehl, W. Cook, and D. A. Johnston, *The efficiency of joint decision making in buyer-supplier relationships*, Annals of Operations Research, 145 (2006), pp. 15–34.

[15] C. Castelfranchi, R. Falcone, and G. Pezzulo, *Trust in information sources as a source for trust: a fuzzy approach*, in Proceedings of the second international joint conference on Autonomous agents and multiagent systems, ACM, 2003, pp. 89–96.

[16] X. Cheng, S. Fu, and G.-J. de Vreede, *Understanding trust influencing factors in social media communication: A qualitative study*, International Journal of Information Management, 37 (2017), pp. 25–35.

[17] B. Choi and I. Lee, *Trust in open versus closed social media: The relative influence of user-and marketer-generated content in social network services on customer trust*, Telematics and Informatics, 34 (2017), pp. 550–559.

[18] E. R. Comer, *Domain analysis: a systems approach to software reuse*, in Digital Avionics Systems Conference, 1990. Proceedings., IEEE/AIAA/NASA 9th, IEEE, 1990, pp. 224–229.

[19] C. L. Corritore, B. Kracher, and S. Wiedenbeck, *On-line trust: concepts, evolving themes, a model*, International journal of human-computer studies, 58 (2003), pp. 737–758.

[20] L. Denoue and L. Vignollet, *An annotation tool for web browsers and its applications to information retrieval*, in Content-Based Multimedia Information Access-Volume 1, LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, 2000, pp. 180–195.

[21] M. Deutsch, *Trust and suspicion*, Journal of conflict resolution, 2 (1958), pp. 265–279.

[22] P. Dondio, S. Barrett, S. Weber, and J. M. Seigneur, *Extracting trust from domain analysis: A case study on the wikipedia project*, in International Conference on Autonomic and Trusted Computing, Springer, 2006, pp. 362–373.

[23] A. Figueiredo, O. Nabuco, T. Al-Chueyr, and M. Rodrigues, *Framework proposal to evaluate trustworthiness in an online community*, in Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 03, IEEE Computer Society, 2009, pp. 579–582.

[24] A. J. Flanagin and M. J. Metzger, *Digital media and youth: Unparalleled opportunity and unprecedented responsibility*, Digital media, youth, and credibility, (2008), pp. 5–27.

[25] B. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, et al., *What makes web sites credible?: a report on a large quantitative study*, in Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, 2001, pp. 61–68.

[26] J. W. Fritch and R. L. Cromwell, *Delving deeper into evaluation: Exploring cognitive authority on the internet*, Reference Services Review, 30 (2002), pp. 242–254.

[27] M. Gamble and C. Goble, *Quality, trust, and utility of scientific data on the web: Towards a joint model*, in Proceedings of the 3rd international web science conference, ACM, 2011, p. 15.

[28] T. Grandison and M. Sloman, *A survey of trust in internet applications*, IEEE Communications Surveys & Tutorials, 3 (2000), pp. 2–16.

[29] K. F. Hashim and F. B. Tan, *The mediating role of trust and commitment on members' continuous knowledge sharing intention: A commitment-trust theory perspective*, International Journal of Information Management, 35 (2015), pp. 145–151.

[30] C. Haythornthwaite, *Crowds and communities: Light and heavyweight models of peer production*, in System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on, IEEE, 2009, pp. 1–10.

[31] J. B. Horrigan and L. Rainie, *The Internet's growing role in life's major moments*, vol. 181, Pew Internet & American Life Project Washington, DC, 2006.

[32] S. Javanmardi and C. V. Lopes, *Modeling trust in collaborative information systems*, in Collaborative Computing: Networking, Applications and Worksharing, 2007. CollaborateCom 2007. International Conference on, IEEE, 2007, pp. 299–302.

[33] C. Jensen, S. Poslad, and T. Dimitrakos, *Trust Management: Second International Conference, iTrust 2004, Oxford, UK, March 29-April 1, 2004, Proceedings*, vol. 2995, Springer, 2004.

[34] S. Kamboj, B. Sarmah, S. Gupta, and Y. Dwivedi, *Examining branding co-creation in brand communities on social media: Applying the paradigm of stimulus-organism-response*, International Journal of Information Management, 39 (2018), pp. 169–185.

[35] K. Kelton, K. R. Fleischmann, and W. A. Wallace, *Trust in digital information*, Journal of the American Society for Information Science and Technology, 59 (2008), pp. 363–374.

[36] P. Kirs and K. Bagchi, *The impact of trust and changes in trust: A national comparison of individual adoptions of information and communication technologies and related phenomenon*, International Journal of Information Management, 32 (2012), pp. 431–441.

[37] A. Kittur, B. Suh, and E. H. Chi, *Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia*, in Proceedings of the 2008 ACM conference on Computer supported cooperative work, ACM, 2008, pp. 477–480.

[38] M. Laroche, M. R. Habibi, and M.-O. Richard, *To be or not to be in social media: How brand loyalty is affected by social media?*, International Journal of Information Management, 33 (2013), pp. 76–82.

[39] R. S. Lazarus, *Emotion and adaptation*, Oxford University Press on Demand, 1991.

[40] J. Lee and I. B. Hong, *Predicting positive user responses to social media advertising: The roles of emotional appeal, informativeness, and creativity*, International Journal of Information Management, 36 (2016), pp. 360–373.

[41] H. Li, J. Jiang, and M. Wu, *The effects of trust assurances on consumers' initial online trust: A two-stage decision-making process perspective*, International Journal of Information Management, 34 (2014), pp. 395–405.

[42] S. C. Lu, D. T. Kong, D. L. Ferrin, and K. T. Dirks, *What are the determinants of interpersonal trust in dyadic negotiations? meta-analytic evidence and implications for future research*, Journal of Trust Research, 7 (2017), pp. 22–50.

[43] T. Lucassen and J. M. Schraagen, *Trust in wikipedia: how users trust information from an unknown source*, in Proceedings of the 4th workshop on Information credibility, ACM, 2010, pp. 19–26.

[44] T. J. Ma and D. Atkin, *User generated content and credibility evaluation of online health information: a meta analytic study*, Telematics and Informatics, 34 (2017), pp. 472–486.

[45] P. Manuel, *A trust model of cloud computing based on quality of service*, Annals of Operations Research, 233 (2015), pp. 281–292.

[46] S. P. Marsh, *Formalising trust as a computational concept*, (1994).

[47] C. C. Marshall, *Toward an ecology of hypertext annotation*, in Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space—structure in hypermedia systems: links, objects, time and space—structure in hypermedia systems, ACM, 1998, pp. 40–49.

[48] R. C. Mayer, J. H. Davis, and F. D. Schoorman, *An integrative model of organizational trust*, Academy of management review, 20 (1995), pp. 709–734.

[49] M. J. Metzger, *Making sense of credibility on the web: Models for evaluating online information and recommendations for future research*, Journal of the American Society for Information Science and Technology, 58 (2007), pp. 2078–2091.

[50] M. J. Metzger and A. J. Flanagin, *Credibility and trust of information in online environments: The use of cognitive heuristics*, Journal of Pragmatics, 59 (2013), pp. 210–220.

[51] M. J. Metzger, A. J. Flanagin, K. Eyal, D. R. Lemus, and R. M. McCann, *Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment*, Annals of the International Communication Association, 27 (2003), pp. 293–335.

[52] M. Miceli and C. Castelfranchi, *The role of evaluation in cognition and social interaction*, (2000).

[53] C. L. Miltgen and H. J. Smith, *Exploring information privacy regulation, risks, trust, and behavior*, Information & Management, 52 (2015), pp. 741–759.

[54] R. M. Morgan and S. D. Hunt, *The commitment-trust theory of relationship marketing*, The journal of marketing, (1994), pp. 20–38.

[55] E. W. Ngai, S. S. Tao, and K. K. Moon, *Social media research: Theories, constructs, and conceptual frameworks*, International Journal of Information Management, 35 (2015), pp. 33–44.

[56] T. M. Nisar, G. Prabhakar, and P. P. Patil, *Sports clubsâĂŹ use of social media to increase spectator interest*, International Journal of Information Management, 43 (2018), pp. 188–195.

[57] J. R. Nurse, I. Agrafiotis, M. Goldsmith, S. Creese, and K. Lamberts, *Two sides of the coin: measuring and communicating the trustworthiness of online information*, Journal of Trust Management, 1 (2014), p. 5.

[58] M. Pournader, A. Kach, S. H. R. Hajiagha, and A. Emrouznejad, *Investigating the impact of behavioral factors on supply network efficiency: insights from banking's corporate bond networks*, Annals of Operations Research, (2017), pp. 1–26.

[59] I. Pranata and W. Susilo, *Are the most popular users always trustworthy? the case of yelp*, Electronic Commerce Research and Applications, 20 (2016), pp. 30–41.

[60] T. U. Rahman, S. Khusro, I. Ullah, and Z. Ali, *Exploiting user expertise and willingness of participation in building reputation model for scholarly community-based question and answering (cqa) platforms*, in Computer Science On-line Conference, Springer, 2018, pp. 436–444.

[61] S. Y. Rieh and D. R. Danielson, *Credibility: A multidisciplinary framework*, Annual review of information science and technology, 41 (2007), pp. 307–364.

[62] P. K. Roy, J. P. Singh, A. M. Baabdullah, H. Kizgin, and N. P. Rana, *Identifying reputation collectors in community question answering (cqa) sites: Exploring the dark side of social media*, International Journal of Information Management, 42 (2018), pp. 25–35.

[63] M. Söllner, A. Hoffmann, and J. M. Leimeister, *Why different trust relationships matter for information systems users*, European Journal of Information Systems, 25 (2016), pp. 274–287.

[64] D. Taraborelli, *How the web is changing the way we trust*, Current issues in computing and philosophy, (2008), pp. 194–204.

[65] D. Thomas and R. Bostrom, *Building trust and cooperation through technology adaptation in virtual teams: Empirical field evidence*, EDPACS, 42 (2010), pp. 1–20.

[66] T. Tucker, *Online word of mouth: characteristics of yelp. com reviews*, Elon Journal of Undergraduate Research in Communications, 2 (2011), pp. 37–42.

[67] W. Y. Wang and K. R. McKeown, *Got you!: automatic vandalism detection in wikipedia with web-based shallow syntactic-semantic modeling*, in Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, 2010, pp. 1146–1154.

[68] M. Warncke-Wang, D. Cosley, and J. Riedl, *Tell me more: an actionable quality model for wikipedia*, in Proceedings of the 9th International Symposium on Open Collaboration, ACM, 2013, p. 8.

[69] B. Wu and H. Shen, *Analyzing and predicting news popularity on twitter*, International Journal of Information Management, 35 (2015), pp. 702–711.

Overview

In this phase of the thesis, we have developed a trust model based on the dimensions of stability, credibility and quality. The proposed model must be verified with an appropriate evaluation method. This means that web users shall evaluate the construction against these measures.

In addition, the correlation of the dimensions to each other with regard to weighting could not be determined precisely. All dimensions have equal factors in the equation to calculate a numerical trust degree. To counteract this limitation, the estimated preferences of the respondents can be mapped in importance values and used as factors for the dimensions to refine the confidence calculation. The following is an introduction to the valuation method used before we present the implementation and results in the next chapter.

Conjoint Analysis

This section introduces our approach to conjoint analysis (CA), its design variants, analysis techniques as well as the criteria under consideration when using it.

Conjoint approaches vary according to (1) models and (2) analysis methods. The models differ based on the *types and functions of attributes*, *trade-off models* and *design*. Next, we explain these differences as follows: *types and functions of attributes*: types can be divided into *categorical* that uses words for describing the levels values and *quantitative*, that uses numbers. There are also two classes of attribute functions: *part-worth function*, where the attribute levels tend to be desirable (piecewise linear) and *vector function*, where attribute levels are piecewise nonlinear.

Trade-off models represent the main forms of conjoint models and cover three classes: The first class is the *Choice Based Conjoint*, which follows the strategy of selecting none[10], one[11] or two[12] choice(s) or respondents compile their own preferred concept. This class includes the sub forms: *Discrete Choice Conjoint*, *Trade-off Matrix method*, *Paired Comparison Method*, *Menu Based Conjoint* and *MaxDiff*. The second class is the *Rating/Scaling Based Conjoint*, in which respondents rate one concept or each attribute level separately. The last class is the *Hybrid Model* that combines different techniques together. For example, the *Adaptive Conjoint Analysis*, in which in the first phase, a respondent has to rank each level separately and, in the second phase, rates the attributes.

Design: There are four main categories that can be applied to design a Conjoint Analysis: first, the *Full Profile Trade-off* that uses all possible attributes and level combin-

[10]non-option
[11]Discrete Choice
[12]Trade-off matrix, paired comparison and MaxDiff (best and worst choice)

ations[13]. Secondly, the *Partial Profile Trade-off* that is more suited if the levels or the attributes number is high[14]. It uses a subset of the full profile configuration. For example, it uses the *Fractional Factorial design* [2, 11, 15] that reduces the configuration within a factor. Third, the *Adaptive Method* which is similar to the prior design and reduces the configuration by combining or eliminating undesirable attributes and using random sampling. Fourth, the *Hybrid Methods* that combines multiple designs, such as, the *Incomplete Block design* that has design conditions for the concept appearance (at most once, or exactly number of times as alone or as it pairs with another concept).

The analysis methods will differ based on the trade-off (choice decision) utilities (part-worths) estimation. This estimation can be divided broadly into two classes according to the researchers' point of view: (1) *Logit model* if the estimation is based on taking the entire trade-offs, or (2) *Hierarchical bayes* if the estimation is based on the trade-offs of each respondent, separately.

The scope of this paper is to evaluate the proposed trust model. The evaluation conducted applies the Discrete Choice Conjoint approach (DCC). This approach was chosen because of its widely uses and its solid outputs. The design applied for DCC is the fractional factorial design (mentioned above) within the factor $\frac{1}{2}$ (32 concepts). This design was chosen due to the length of the configuration (64 concepts) that could be produce in case of applying a full profile trade-off. For the utilities estimating we used the Logit Model.

Design Criteria

A DCC is a repeated-design task, in which respondents repeatedly select one concept out. DCC consists of a profile that represents a possible combination of choices. Tasks build a profile and each task includes one portion of the alternatives that illustrate the products properties (attributes giving levels). Respondent's choices are called trade-offs, which can be analyzed for drawing a conclusion about the relative importance (decision utility) of each attribute.

Certain criteria to be comply with when using Conjoint Analysis to achieve reasonable results. According to [12, 25, 26] some of these are as follows: First, *number of levels effect*, that is, the importance of an attribute has a relative value i.e. changing the utility of one level influences the importance of other attributes. Also, changing the number of levels increases its attribute importance. Secondly, *independence from irrelevant alternatives problem* influences the attributes importance. The distribution of the total importance usually depends on the number of attributes. For instance, if there are multiple attributes obtained from one attribute, then the preferences of respondents will be manipulated and consequently the resulting importance. Finally, *attributes interaction*, which means that the attributes should not be in relationship to one another. For instance, two attributes are connected if selecting one of them would influence (increasingly or decreasingly) the selection probability of other one.

The DCC used in this work applies symmetric attributes levels (i.e. each attribute has a

[13] Configuration length = $levels^{attributes}$

[14] There is no fixed threshold for this, however, the configuration length should not make a respondent tire.

similar number of levels) to avoid the number of levels effect. The attributes are independent and cannot be obtained from each other. They will differ in nature and characteristics due to their different types (see Chapter 6). For the same reason the attributes have no connection with each other and they have no common nor cover exclusion criteria[15].

Provision

This section introduces the preparation for DCC and established the instructions respondents received.

Using e-mail, we announced a link to the online survey in Arabic, English and German. In the DCC we described the attributes: (1)*Comments* as "a number that indicates improvement edits created by other readers", (2) *Reader Rating* as "a number of other readers' approval" and (3) *Author Rating* as "a number of voting that the author earned for his/her activities in the social network". We also stated that "The greater the number, the greater the satisfaction". Each number represents the sum of negative and positive assertions. There are comments that were rated negatively and other comments that were rated positively by readers. Negatives were marked with minus and positives with plus numbers. Subsequently, both numbers were summed up. This applies to all properties".

Due to the amount of information in a full-profile design ($4_{levels}^{3_{attributes}}$ makes 64 alternatives), a complied questionnaire would become to extensive. Therefore, we decided to use a randomly fractional factorial design within the factor $\frac{1}{2}$. The conducted DCC consists of 32 concepts and takes an average of 9 minutes to completion for each respondent. They also received four alternative selections for each task. The attributes are *Comments*, *Reader Rating* and *Author Rating* as illustrated in Table 6.6.

[15]Unrealistic choice offer. As a verified or signed statement on social media.

Table 6.6: Attributes and Levels Design

Attribute	Level
Comment	0
	2
	5
	10
Reader Rating	0
	5
	30
	70
Author Rating	-100
	0
	1000
	2000

Tab.6.6 illustrates the attributes and the levels used in the Discrete Choice Conjoint design. The design is symmetric and the level values are derived from the data analysis collected from Genius which is presented in the previous work.

Chapter 7

Investigating the Effect of Attributes on User Trust in Social Media

[1]

7.1 Abstract

One main challenge in social media is to identify trustworthy information. If we cannot recognize information as trustworthy, that information may become useless or be lost. Opposite, we could consume wrong or fake information - with major consequences. How does a user handle the information provided before consuming it? Are the comments on a post, the author or votes essential for taking such a decision? Are these attributes considered together and which attribute is more important? To answer these questions, we developed a trust model to support knowledge sharing of user content in social media. This trust model is based on the dimensions of stability, quality, and credibility. Each dimension contains metrics (user role, user IQ, votes, etc.) that are important to the user based on data analysis. We present in this paper, an evaluation of the proposed trust model using conjoint analysis (CA) as an evaluation method. The results obtained from 348 responses, validate the trust model. A trust degree translator interprets the content as very trusted, trusted, untrusted, and very untrusted based on the calculated value of trust. Furthermore, the results show a different importance for each dimension: stability 24%, credibility 35% and quality 41%.
Keywords Social Media, Trust, Conjoint Analysis.

We refer the reader to the publication:
Title=Investigating the Effect of Attributes on User Trust in Social Media,
author=Al Qundus, Jamal and Paschke, Adrian,
booktitle=International Conference on Database and Expert Systems Applications,
pages=278–288,
year=2018,
organization=Springer
https://link.springer.com/chapter/10.1007/978-3-319-99133-7_23.

[1]The content of this chapter has been published in [6], which is coauthored with A. Paschke

Overview

The first evaluation in the previous chapter deals with the information examined as a black box. That is, we looked at the metadata of the information without paying attention to the text, because the classification was based only on the metadata we intended to verify. After this evaluation phase, which was carried out with conjoint analysis, it is now time to examine the content of the trust classes resulting from the trust model and confirmed by user preferences.

We conduct two investigations: 1) A Natural Language Processing (NLP) approach to extract relevant features (i.e. Part-of-Speech and various readability indexes) (see Chapter 8). We report relative good performance of the NLP study. 2) A machine learning technique in more precise, a Random Forest classifier (RF) using Bag-of-Words model (BoW), which Chapter 9 presents.

Chapter 8

Application Natural Language Processing

In order to identify the rules of the quality relation in each class, we investigate the correlation of various readability indexes and syntactic constructions. Using 20% of the data sets, we tested the average number of sentences, words, characters, complex words, syllables, unique words, words per sentence, syllables per word, part-of-speech and the readability indexes KINCAID, FOG, SMOG, FLESCH, COLEMAN_LIAU and ARI.

8.1 Part-of-Speech and Readability Indexes of Short Text

Part-of-Speech (PoS) tegger[1] and Readability Indexes[2] are widely used in data mining. First, we cleaned the data base using a set of stop-words, which contain standard words identified as noisy and other instances (e.g. smileys, "LOL", "!!!!" etc.) that we observed in the data base. Second, all words were stemmed and lemmatized, with the aim "to reduce inflecting forms and derived related forms of a word to a common basic form"[3]. Finally, using the PoS-tegger the numbers related to the attributes illustrated in the Listing 8.1 are collected.

[1] https://nlp.stanford.edu/software/tagger.shtml

[2] http://www.readabilityformulas.com/

[3] Citation: https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html [Acccessed:3 October 2018]

Listing 8.1: First verbatim

```
@relation quality.symbolic
@attribute numSentences numberic        @attribute numComplexWords numberic
@attribute numWords numberic            @attribute numSyllables numberic
@attribute uniqueWords numberic         @attribute fog numberic
@attribute flesch numberic              @attribute kincaid numberic
@attribute wordsPerSentence numberic    @attribute perComplexWord numberic
@attribute syllabelPerWord numberic     @attribute ari numberic
@attribute smog numberic                @attribute colemanliau numberic
@attribute gunningfog numberic
@attribute quality {VT, T, U, VU}
@data   numeric vector and a label
```

Listing 8.2: Run information

```
Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     quality.symbolic
Test mode:    500-fold cross-validation
```

Listing 8.3: Stratified cross-validation

```
=== Summary ===
Correctly Classified Instances         301               34.0498 %
Incorrectly Classified Instances       583               65.9502 %
Kappa statistic                          0.1272
Mean absolute error                      0.3371
Root mean squared error                  0.4819
Relative absolute error                 89.9373 %
Root relative squared error            111.2642 %
Total Number of Instances              884
```

8.2 Experiment

For this experiment we build multiple-class classifiers that are Random Forest, Decision Tree and Naive Bayes. Decision Tree did not perform well compared with Random Forest and Naive Bays, which have achieved similar accuracy. This accuracy resulted based on Naive Bayes with the test mode of 500-fold cross-validation using tool weka-3-8-2. Listing 8.2 summarizes the run information applied. The multiple-class classifier of Naive Bayes achieves 34% accuracy, which is considered as a good performance towards dealing with short-text against four classes. Listing 8.3 shows the statics of the experiment and Listing 8.4 represents the confusion matrix of the classifier, which performs well in recognition the instances of very-trusted short-text (VT:105) and very-untrusted short (VU:114) compared to the instances of untrusted short-text (U:71) and trusted short-text (T:11). These are good classification results considering the challenge of short-text multiple classifications. Nevertheless, this irregularity only allows determining a pattern with increasing or decreasing properties to a limited extent. Thus, hardly a single generalization can be enforced across all classes.

The Figures [8.1 to 8.10] visualize and Listing 8.5 represents the features investigated in order to train a classifier to category short text. From these figures and listing, we can see that some of the features have a linear verity over the four classes. For example,

Listing 8.4: Confusion Matrix

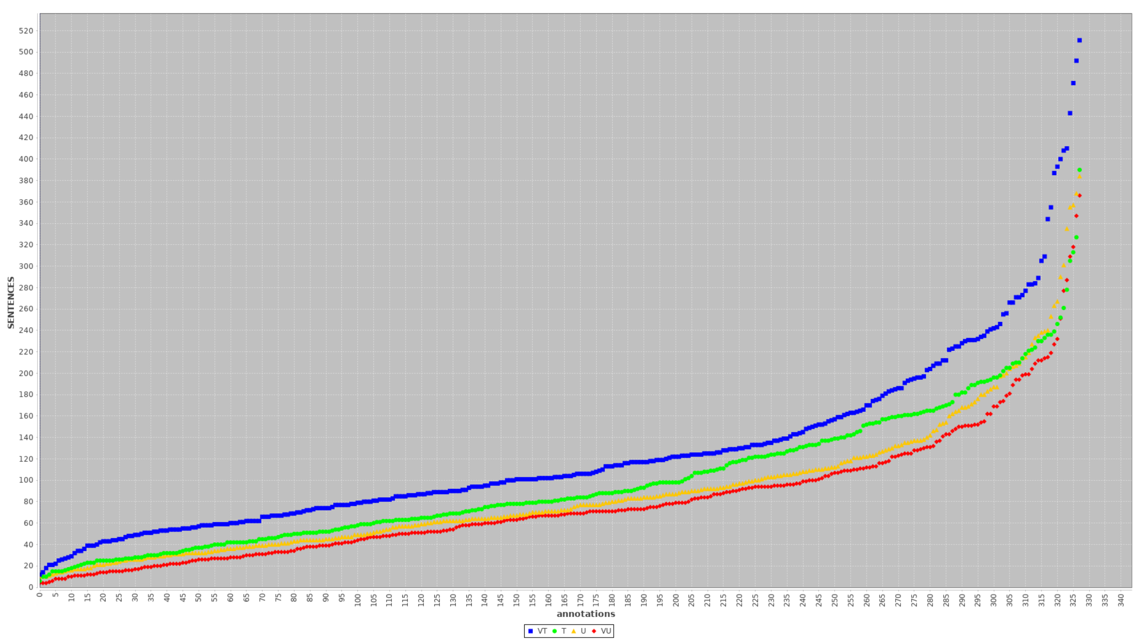
```
   a   b   c   d   <-- classified as
 105  17  36  65 |   a = VT
  71  11  74  85 |   b = T
  43  13  71  78 |   c = U
  37   7  57 114 |   d = VU
```

considering the feature numWords (means and 78.0, 61.1, 55.6 and 48.8 of the classes VT, T, U and VU respectively) that could be apply as a good metric in order to classify short-text regarding trust. While the feature numSentences (means of 4.7, 3.5, 3.3 and 3.1) is less of analytic quality and the feature wordsPerSentence (17.1, 18.19, 17.7 and 16.3) is even inexpressive.

8.3 Discussion

The experiment carried out represents a series of attributes that contain remarkable references to the differentiation of the short texts contained in the trust classes. The oblique line indicates the linear relationship represented by the attributes having the averages of numSentences, numWords, numSyllables and uniqueWords. However, these attributes include relatively high standard deviations, in addition to the numComplexWords attribute, which enable them to hardly predict a trust class of short texts. While in most cases all readability indexes and the attributes wordsPerSentence, perComplexWords, SyllablesPerWord have an irregular distribution, which makes them unsuitable for trust prediction. The following Figures [8.1 to 8.10] illustrate an overview of some attributes examined.

NLP Analysis



(VT T U VU)

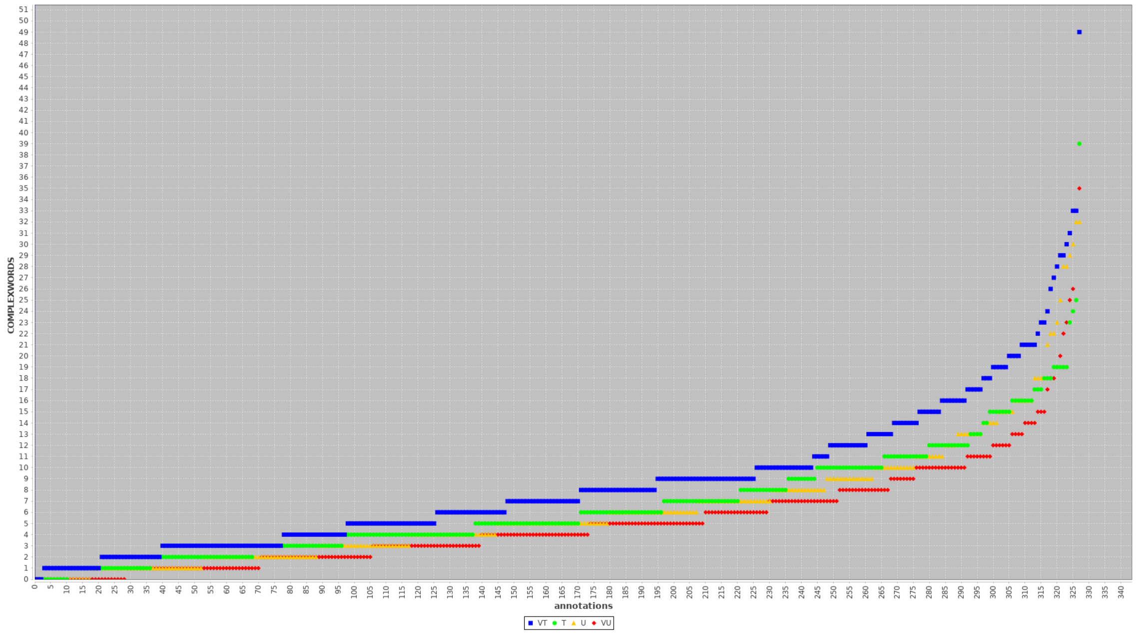
Figure 8.1: The average number of sentences

8.4 Conclusion

This pre-analysis includes exploring several PoS and readability indexes as features of short-text classified in the trust classes. Despite indications that could be able to predict

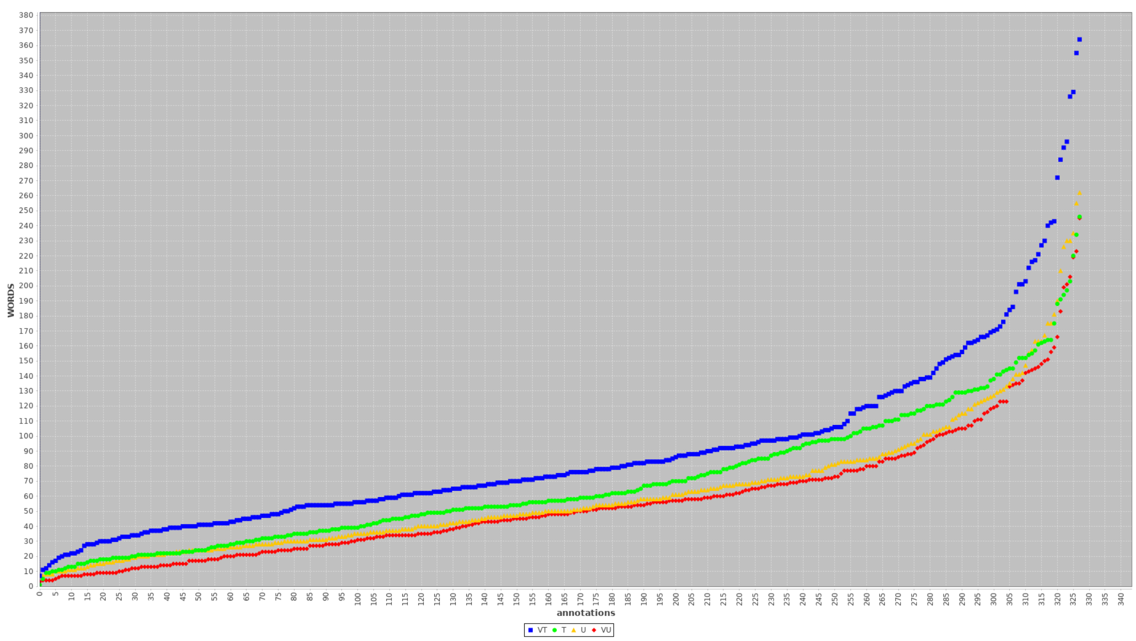
Listing 8.5: Classifier model (full training set)

```
Naive Bayes Classifier
                      Class
Attribute                VT         T         U        VU
                      (0.25)    (0.27)    (0.23)    (0.24)
=======================================================
numSentences
  mean                4.7623    3.5228    3.3805    3.1628
  std. dev.           2.1349    1.8936    1.6144    1.6215
numComplexWords
  mean                8.0463    6.5494    6.3581    5.1268
  std. dev.           4.3942    4.2212    3.7692    3.4062
numWords
  mean               78.0733   61.1876   55.6978   48.8358
  std. dev.          36.7594   34.0534   27.9333   28.4876
numSyllables
  mean               113.065    89.611   82.5455   71.1264
  std. dev.          53.4831   50.2359   41.0283   41.0894
uniqueWords
  mean               57.8249   46.6886   43.6727   38.6368
  std. dev.          22.4026   22.1374   18.4226   19.4957
fog
  mean               10.9715   11.6489   11.7779   10.9657
  std. dev.           2.5858    2.9443    2.9291    2.9769
flesch
  mean               66.8386    63.899   62.6884   65.2206
  std. dev.           9.5233   11.7864   11.2318   12.4922
kincaid
  mean                8.1878     8.866    8.9309    8.2298
  std. dev.           2.3339    2.7044    2.6997    2.7788
wordsPerSentence
  mean               17.1114   18.1977   17.7764   16.3732
  std. dev.           5.5741    6.3048    6.8872    7.0193
perComplextWords
  mean               10.3181   10.9313   11.6678   11.0371
  std. dev.           3.4703    4.4117    4.5355    5.2663
syllabesPerWords
  mean                1.4496    1.4714    1.4906    1.4775
  std. dev.           0.0995    0.1277    0.1291    0.1489
ari
  mean                8.0494    8.8649    8.9173    8.0922
  std. dev.           3.3895    3.5744    3.6805    3.5513
smog
  mean               10.5185   10.8819   11.0257   10.2869
  std. dev.           1.7713    2.0189    1.9472    2.0772
colemanliau
  mean                8.2661    8.6623    8.8431    8.4183
  std. dev.           2.2549    2.9286    2.7123    2.6454
gunningfog
  mean               40.7057   40.7841   40.8075   40.6786
  std. dev.            0.334     0.401    0.3892    0.3842
```



(VT T U VU)

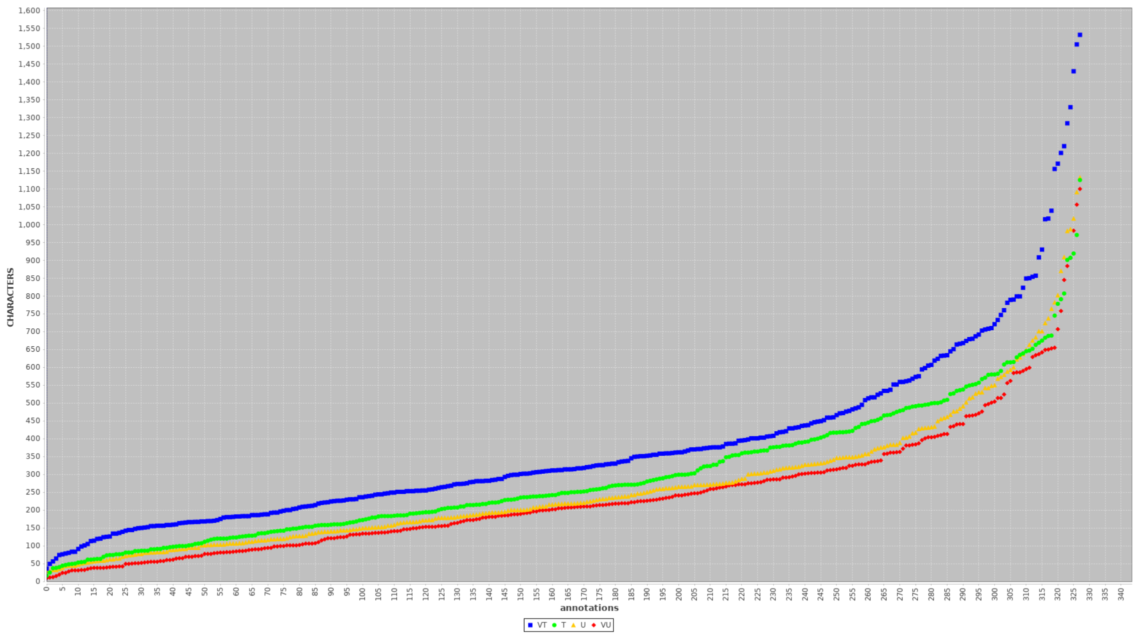
Figure 8.2: The average number of complex words



(VT T U VU)

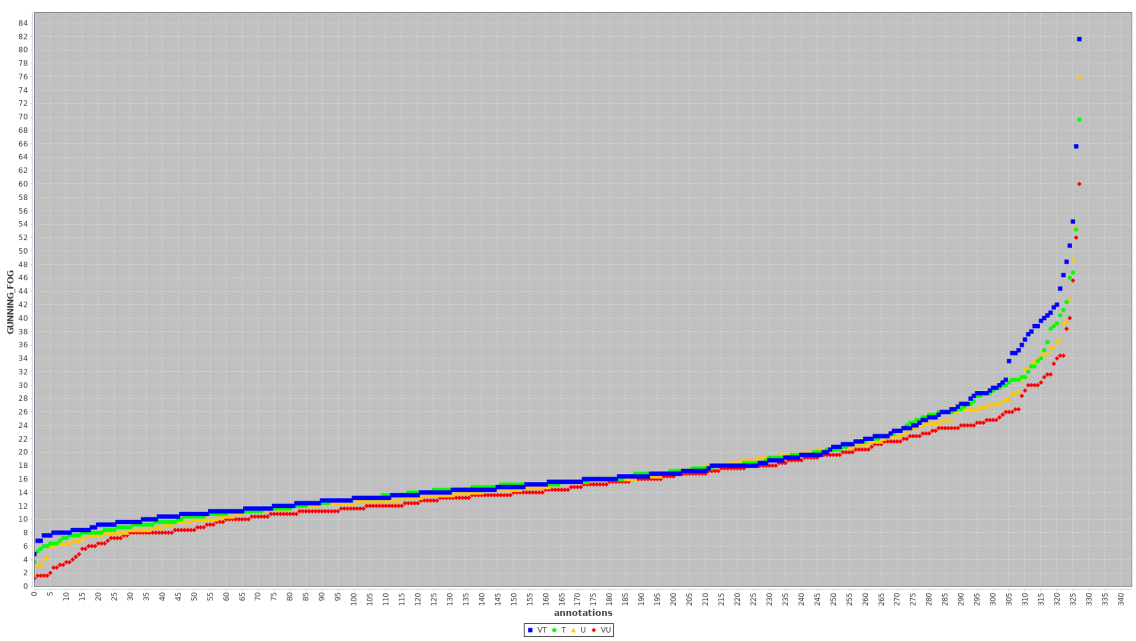
Figure 8.3: The average number of words

trust class of a short-text given, we could not establish enough evidences on the influence of such features on trust. Thus, it is hard to consider these features in the trust model



(VT T U VU)

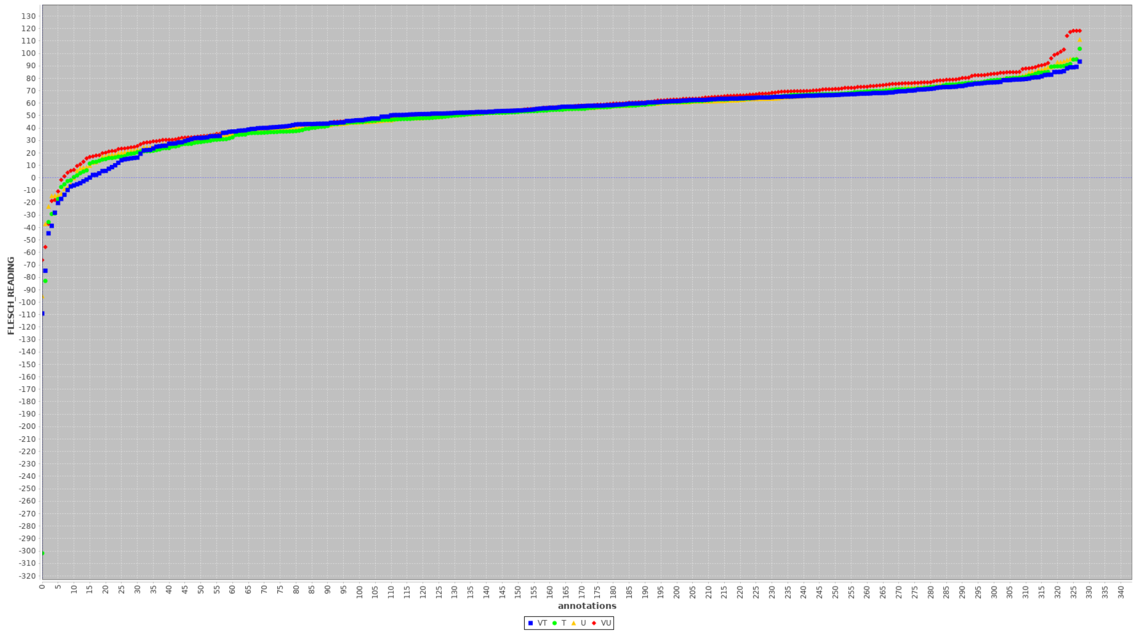
Figure 8.4: The average number of characters



(VT T U VU)

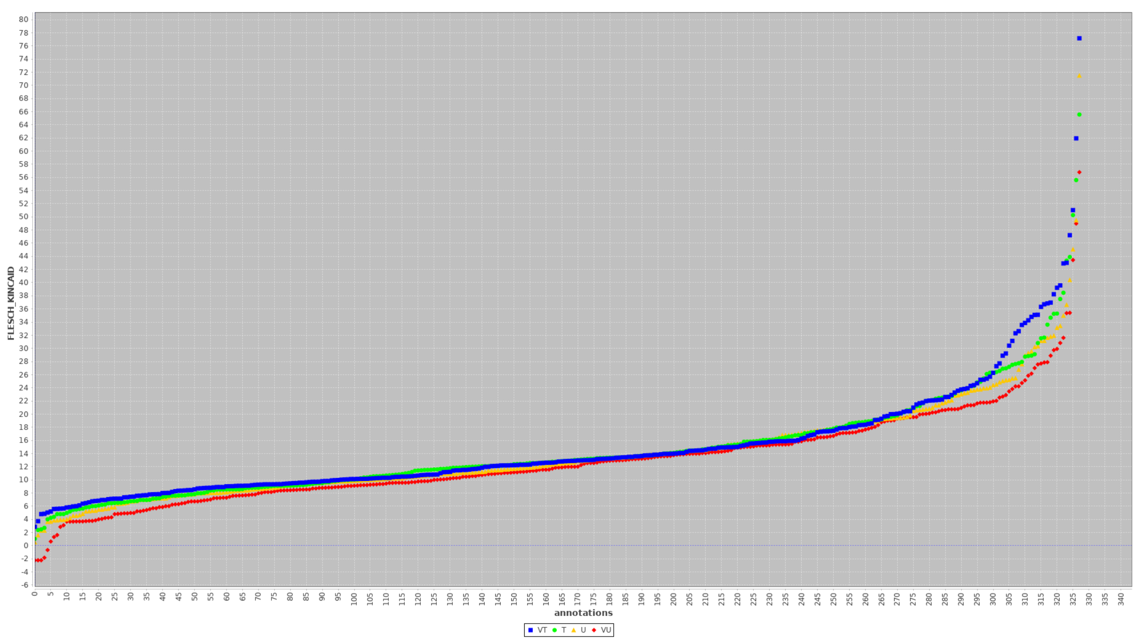
Figure 8.5: FOG index

proposed. That is why; we looked for another way to investigate the correlation between a short-text and the trust class related to.



(VT T U VU)

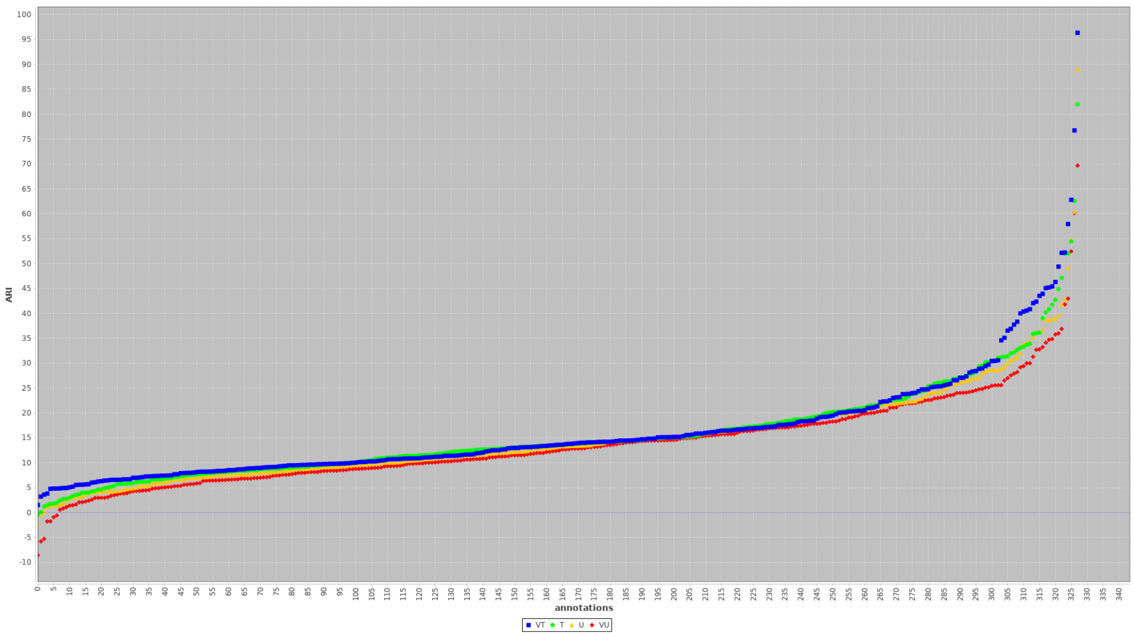
Figure 8.6: FLESCH index



(VT T U VU)

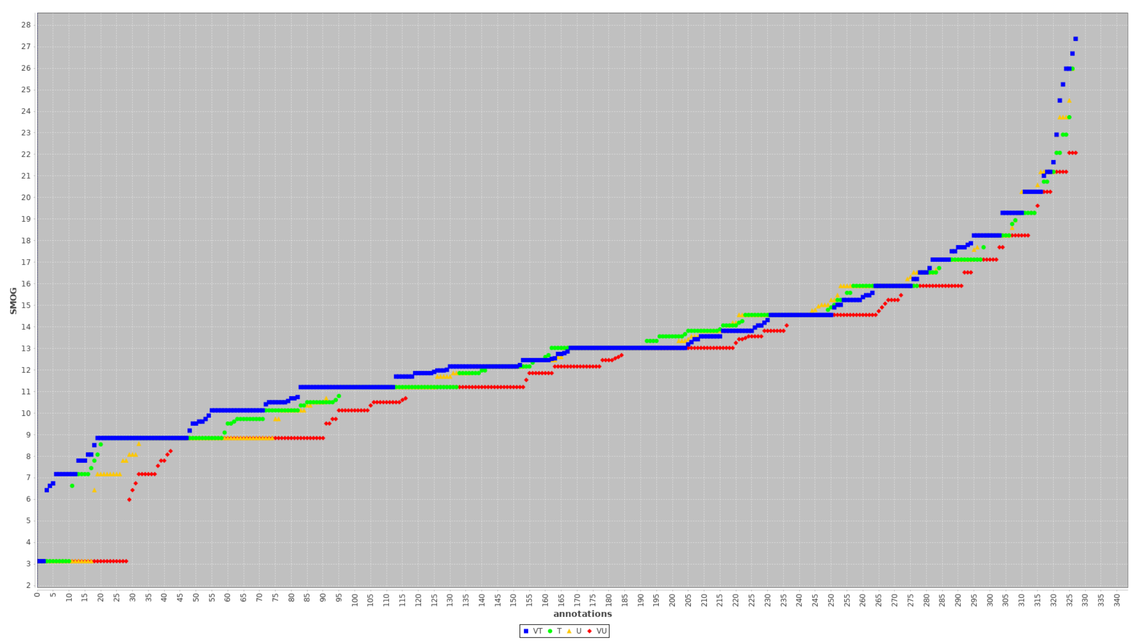
Figure 8.7: KINCAID index

Due to the limited page number of pages that can be published, thus, the results of the NLP and readability indexes are shortly presented in the paper that the next chapter



(VT T U VU)

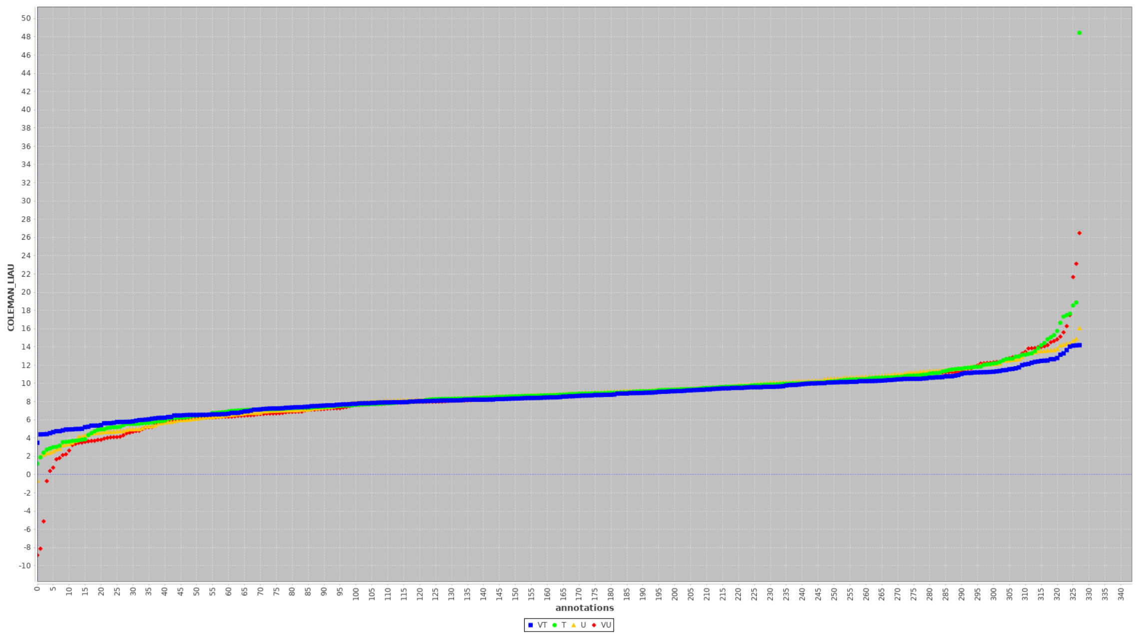
Figure 8.8: Automated Readability Index



(VT T U VU)

Figure 8.9: SMOG index

relies on, and the focus in it was on the results of machine learning technique (Random Forest).



(VT T U VU)

Figure 8.10: COLEMAN_LIAU index

Chapter 9

Design Science Research: Exploring the effects of text complexity on its quality in social media

[1]

9.1 Abstract

Short text classifications with regard to a criterion (e.g. quality, readability, etc.) are usually extended by an external source or its metadata. This enhancement either changes the original text if it is additional text from an external source, or it requires text metadata that is not always available. This study meets this challenge by working with the original short text without extension. The aim of this work in progress is to predict short text quality that leads to trust. The research questions are: Where do short texts differ in its content? To what extent can such differences be used to predict the classification of short texts in terms of quality? Can a relationship be established between metadata and the content of a short text? To address these questions, we apply our trust model to classify data collections based on metadata into four classes: Very trusted, trusted, untrusted and very untrusted. These data collections are gained from the online communities Genius and Stackoverflow. To evaluate short-texts in terms of its trust levels, we conduct Random Forest classifier (RF) using Bag-of-Words model (BoW) and report promising intermediate results (on average 62% accuracy of both online communities) in short-text quality identification that leads to trust.

Keywords Trust, Short-Text Quality, Feature Extraction, Random Forest, Readability, Part-of-Speech

[1]The content of this chapter is under review in French Journal of Management Information Systems (SIM), which is coauthored with A. Paschke, S. Gupta and M. Yousef

9.2 Introduction

It is generally known that social media are growing and become the source of information for many users. The whole world is becoming a small town with an unlimited space, where the information is spreading very quickly, and, in many cases, we cannot identify its trigger. In addition, the structure of providing information take the form of many-to-many instead of the traditional form one-to-many. The many-to-many structure makes identifying of high-quality information more challenging.

The features of texts were examined as early as 1948's [6] in various domains. The goal is to formulate text to fit the age and reading skills of a target group's category. For example, in educational field, it is important to generate text with a level fitting to the students' class level [18]. Feature extraction rests on various approaches, e.g. syntactic, vocabulary, coherence, cohesion, discourse etc. and applied various arguments or features, e.g. text parse tree height, bag-of-words based on word frequency, cosine similarity, information ordering etc. as introduced in [3, 18]. Depending on the type of text, some features perform well and are therefore correlated with e.g. readability and some are not.

It is obvious that depth in meaning, easy to read, well formated structure, error free in grammar and more, are characteristics of qualitative text. These can be measured using readability indexes and natural language processing (NLP) approach to analyze Part-of-Speech (PoS). The major aim is the transmission of high quality information, whereby the topic and target group are limited or already given. However, in the context of social networks as Genius[2], Stackoverflow[3] or Twitter[4] these are very broad or even not known. We do not need to prove the importance and influence of these communities here. On the other hand, it is appropriate to clarify the challenges involved in investigating their contents when it comes to using traditional machine learning techniques. Especially social media, which are in focus for our work.

The social media Genius allows text interpretation in form of annotations on various topics. Annotations are interpretation placeholder and provide metadata such as authorship, reader rating and suggestions. Genius users are distributed into six roles (Whitehat, Artist, Editor, Mediator, Contributor and Staff or Regulator) that are assigned different permissions (more details in our technical report [1]). Interpretations are free (unstructured) short texts that are unlimited and can contain pictures, links (URLs and cross-references) and symbols.

Numerous examinations on short text exist. [19] investigate short data mining for information extraction and use base principles of NLP (stemming, tokenizing, N-grams and bag-of-word) for applying clustering tasks (term frequency inverse document frequency). [5] examine the linguistic and discoursal characteristics (short cut, co-language, slangs, symbols etc) in massages to predict the authorship (male or female). Short text clustering in social media is an increasingly application investigations, [10] determine the community semantic words to create semantic concept vectors, to which short texts are mapped to identify their cluster. [20] apply a short text clustering and introduce an information filtering approach for tweets. [23] propose a tweet segmentation system that splits short text into semantic *expression in English i.e., worldwide setting* or setting data *expression*

[2] https://genius.com/
[3] https://stackoverflow.com/
[4] https://twitter.com/

inside the clump of tweets i.e., nearby setting. [17] extract user preferences to improve marketing strategy development by classification of emotion in short text.

Since short texts are usually created in an intimate and natural way, they are all the more valuable for deriving user needs. In addition, since NLP is not suitable for short texts and therefore the results obtained are inaccurate, the necessity for adapted methods has become urgent. The information gained from short texts can be used to support and improve search/filtering engines, preferences prediction (marketing), knowledge sharing, vandalism and fake news recognition etc. That is why, identification of high quality short texts and their transmitting in such a way, that human being can understand, are essential.

This research addresses questions: where do short texts differ in their content? To what extent can such differences be used to predict the classification of short texts in terms of quality? Can a relationship be established between Metadata and the content of a short text? Our approach is novel in addressing the problem of trust classification based on the content solely and without text manipulation.

The aim of this work is to present our approach in progress and to discuss the promising initial results of examination quality of short texts. These can either be transferred to social collaboration such as Wikipedia or Yelp etc., or form the basis for it.

This paper is structured as following: Section 9.3 introduces the related works that have motivated our work. Section 9.4 represents the background of the current work by introducing our trust model used for the basis-classification as a reference for the approach in Section 9.5 including the machine learning technique applied, and the computational theory behind such technique. Section 9.6 represents the results that are discussed in Section 9.7. Conclusion, Limitation and Future Scope of Research take place in Section 9.8.

9.3 Related Work

This section introduces the works that inspire our approach. Barzilay and Lapata [3] propose an entity-based statistical model that examines entity coherence and lexical cohesion elements e.g. the number of pronouns or definite articles per sentences to qualify topic continuity from sentence to sentence, average cosine similarity and word overlaps. The authors reported good results, e.g. more entities and more verb phrases reduce readability. The operation case of this work is a corpus as a whole. This is indeed an essential first step towards gaining a vision about the distribution of text properties. However, our purpose is to put sub sets or even individual short texts in competition. The work addresses also that so called *ranking problem -text ordering and summary evaluation-*, which is difficult to apply on short texts. Todirascu et al. in [22] examine various cohesion aspects to investigate readability of text. The authors extend the texts with the corpus as external knowledge to manually annotate on the texts. We adopt some of the features used, e.g. the *Entity density* and plan to deploy the *co-reference chain properties*. This approach uses a corpus that originates from a specific topic (French as a foreign language), which simplifies the annotation process but cannot be applied in the context we address. Unlike most approaches to exploring the similarity of documents, Hatzivassiloglou et al. in [7] approach acts similarity at the level of text parts (paragraph- and sentences-level). Each text part is a representative of the action contained in the original text. Accordingly, dealing with such text parts is more difficult, because the smaller the text, the less likely it is to find matches between words on which the measurement of similarity is based. The work provides *composite features* that are pairs correlations of *primitive features* such as noun phrases and semantically similar verbs. A machine learning method for mapping the feature values into a similarity measure. The work suggests a text summarizing approach, since we are interested more in the semantically similar grammar, we can modify the technique used and adopt it for our purpose. For example, they calculate the semantic distance between phrases, we can use the same calculation to evaluate the distance of the number of relation triples (subject-verb-object) that occur in the sub sets in consideration. Heilman et al. combine (1) grammer- (using a classifier - identifying features, algorithm applies the features and a component for training data) and (2) vocabulary (patterns of use) model-based approaches to predict readability of first and second language texts. The authors stated that these approaches are suitable for web documents and short texts [8]. We want to pursue this as part of our approach. In addition to Heilman et al. approach, Pitler and Nenkova in [15] combine discourse feature to predict, but in this case, the quality of the text. Even if discourse relations based on vocabulary are not so important, we were still interested in this work. Due to the wide array of the characteristics examined. In this work, reader rated (readability measure) journal articles selected from the Wall Street Journal corpus. The authors tried to restore the results of the rating by calculating the likelihood of an array of features and using a linear regression to measure the features correlation. The last two works are the closest to our work with the differences of context (social network), audience (open), text type (short text, interpretation), focus (quality leads to trust), features (fitting to all previous ones) and the way we divide the corpus into four segments to be investigated.

9.4 Background

In this section, we introduce our trust model used for the basis-classification, the machine learning technique applied in this paper, and the computational theory behind such technique, briefly.

9.4.1 Trust model

In our prior work [2], we developed a trust model that classifies Genius annotation into four classes (very trusted, trusted, untrusted and very untrusted) as shown in Table 9.1. These classes were created on the basis of the Empirical Cumulative Distribution Function (ECDF) and the manual observation of the database analysis. The model consists of three dimensions categorized based on the metadata metrics of the annotations. These dimensions are 1) Stability, which is calculated based on the number of edits of annotations over a period of time. 2) Credibility is based on user ratings and activity type, and 3) Quality depends on the profiles of authors and editors. Each dimension has a weight that is calculated on the basis of the measure of user preferences. These measures and their weights give a trust degree of an annotation, which is then assigned to one of the trust classes.

In this paper, however, we hold the same classes, in its place; we propose a new approach that classifies these instances based on the content (text) using the Random Forest technique.

9.4.2 Random Forest (RF)

According to Klassen and Patrui [11], random forest is a meta-learner, which consists of a collection of individual trees (logical conjunction of disjunctions). Each tree votes on an overall classification for the given set of data. The random forest algorithm decides to select the individual classification with the most votes. Each decision tree is created from a random subset of the training data set, using that so called replacement, in performing this sampling. That is, the entities contained in a subset for building a decision tree are possible candidates of the next subset for creating the next decision tree, which leads to that some entities are included more than once in the sample, and others won't appear at all. When building each decision tree, a model based on a different random subset of the training data set and a random subset of the available variables is used to select how best to partition the data set at each node. Each decision tree is designed for its maximum size, with no pruning performed. Together, the resulting decision tree models of the random forest represent the final ensemble model where each decision tree votes for the result, and the majority wins.

In order to evaluate the performance of the RF classifier, a set of the following measures are considered: (1) sensitivity (SE) which represents the true positive rate, (2) specificity (SP) which represents the true negative rate (complement of sensitivity), (3) precision (PR) which represents the ability of correctly predicted positive target condition to the total, (4) accuracy (ACC) represents the classifier ability to predict the target condition correctly, (5) F-measure (F-measure) represents the classifier ability to predict the target condition correctly (comparing to ACC, it tells a lot more in case of imbalanced date set, since it considers both PR and SE), (6) the Matthews correlation coefficient (MCC) which indicates the correlation degree of the tree decisions; according to the following formulations:

$$(1)\ SE = TP/\ (TP + FN)$$
$$(2)\ SP = TN/\ (TN + FP)$$
$$(3)\ PR = TP/\ (TP + FP)$$
$$(4)\ ACC = (TP + TN)/\ (TP + TN + FP + FN)$$
$$(5)\ F\text{-}Measure = 2 \times (PR \times SE)/\ (PR + SE)$$
$$(6)\ MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{((TP+FP)(TP+FN)(TN+FP)(TN+FN))}}$$

With TP is the number of true predicted positives, TN the number of true predicted negatives, FP the number of false predicted positives and FN the number of false predicted negatives. This work uses the RF from the platform KNIME [4] with the following parameters and options: numTrees which consists of defining the number of trees to generate (equals 100). Seed, which means the random number seed used (equals 1) . numExecutionSlots (1) means the number of execution slots (threads) to use for constructing the ensemble. maxDepth (0 for unlimited) means the maximum depth of the trees. numFeatures (0) means the number of attributes used in random selection. The next section describes the workflow applied.

Table 9.1: Genius and Stackoverflow Corpus Overview

Community	Corpus Size	Very Trusted	Trusted	Untrusted	Very Untrusted
Genius	212,397	162,910	40,370	7,477	1,640
Stackoverflow	2,212	844	904	347	117

9.5 Approach

The evaluation of design artifacts and design theories is an important part of design science research Hevner et al [9]. March and Smith [13] stated that design science research (DSR) consider artifacts that could be constructs, models, methods etc. Accordingly, the trust model proposed is the artifact of this DSR. Using the hierarchy of criteria for information system artifact evaluation introduced by Prat et al. [16], the main aim of this work is the artifact evaluation, which addresses the artifact views 1) goal including the evaluation criteria validity and 2) activity consisting of the evaluation criteria consistency and accuracy. In the following we describe in more details our approach.

Pre-processing Machine learning algorithms perform poorly in working on texts in their original form. The corresponding form represents a vector of numerical features. Therefore, a pre-processing step is necessary to convert the texts into a clearer representation. The text segmentations of the classes are cleaned using the following tasks (see Table 9.2): Punctuation Erasure, N-chars Filter, Number Filter, Case Converter (lower case), Stop-words Filter, Snowball Stemmer and Term Filtering. We used a language detector provided by Tika-collection to process on English-text only.

The output of this stage is a new file represented in a vector space based on the bag-of-words model (BoW) combined with Term-Frequency (TF). The representation is actually 1 if the word is present and 0 otherwise. The TF is the number of times a word appears in the instance (text).

We first experiment with the effect of changes in random forest's parameters on its performance. The number of trees to be generated numTrees was set to 10, 100 and 150. The number of tree depth was set to 10 and then to 0 to create trees of any depth. The number of runs (loop) was set to 10, 100, 150 and 1000. Within each run the data is split again into training and testing data and at the end of the loop the average on measures is calculated. For all data sets, the RF performs stable on the numTrees equals 100, unlimited tree depth (0) and the loop size of 100.

Model Building The classifiers were trained and tested, with the division into 80% training data and 20% test data from the data generated by the pre-processing phase. The data sets used by the classifier are imbalanced (see Table 9.1), which can influence the classifier to the advantage of the set with more samples and is so called the problem of the imbalanced class distribution. We have applied an under-sampling approach that reduces the number of samples of the majority class to the minority class, thus reducing the bias in the size distribution of the data subsets. Negative and positive examples were forced to equal amounts when performing a 100-fold Monte Carlo Cross Validation (MCCV) [24] for a model setup. This data was resulted by merging each two-trust class into one corpus in

Table 9.2: Pre-processing Phase of the Knime-workflow including the Nodes and their description.

Node	**Description**: Source `https://nodepit.com`
Tika Language Detector	This node uses the Apache Tika library to detect the language of a given String/Document value. The newly detected languages will be appended to the input table. The list of all supported languages can be seen here. If the text contains mixed languages, the detector will, by default, return the language with the most confidence value
Punctuation Erasure	Removes all punctuation characters of terms contained in the input documents.
N Chars Filter	Filters all terms contained in the input documents with less than the specified number N characters.
Number Filter	Filters all terms contained in the input documents that consist of digits, including decimal separators "," or "." and possible leading "+" or "-". There is also an option to filter all terms that contain at least one digit.
Case Converter	Converts all terms contained in the input documents to lower or upper case.
Stop Word Filter	Filters all terms of the input documents, which are contained in the specified stop word list.
Snowball Stemmer	Stems terms contained in the input documents with the Snowball stemming library.
Term Filtering	based on document frequency

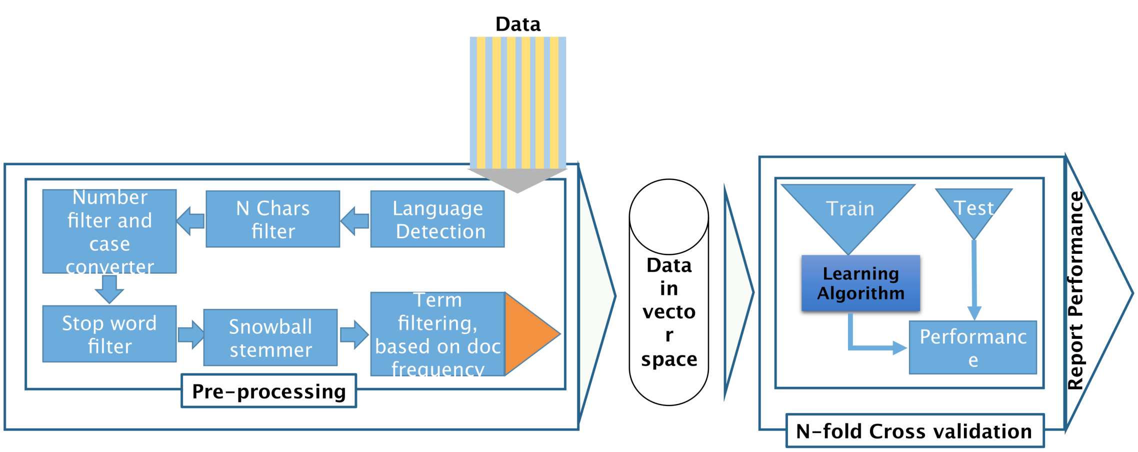


Figure 9.1: Knime Workflow

This figure illustrates the workflow applied for the short-text analysis. It consists of two phases: Pre-processing phase, on the left side, and the learning stage On the right side.

Table 9.3: Binary Combination of the Trust Classes

Community	Combination	Size	#key-words
Genius	very trusted vs. trusted	98,497	300
	very trusted vs. untrusted	72,156	299
	very trusted vs. very untrusted	67,181	309
	trusted vs. untrusted	39,289	280
	trusted vs. very untrusted	34,314	288
	untrusted vs. very untrusted	7,973	228
	very trusted + trusted vs. untrusted + very untrusted	106,470	299
Stackoverflow	very trusted vs. trusted	1,748	730
	very trusted vs. untrusted	1,191	784
	very trusted vs. very untrusted	961	755
	trusted vs. untrusted	1,251	674
	trusted vs. very untrusted	1,021	673
	untrusted vs. very untrusted	464	755
	very trusted + trusted vs. untrusted + very untrusted	2,212	700

order to conduct binary-classification. For the feature extraction, we use the bag-of-words model, which scans each class to build a collection of words presented and sum up their frequency. This bag-of-words is then used to calculate Term Frequency (TF) that indicates the similarity degree between a text and a document, in our case; these are an instance and a trust class. The resulted combination are recognized in the Table 9.3.

In the following, an illustrative example describes the two phases of the approach carried out.

Illustrative Example

We shall prepare a file containing the trust classes very trusted (vt) and very untrusted (vu) for processing. In the pre-processing phase, two file readers load the classes vt and vu in their raw form independently of each other. Each class is converted into a document marked with its class label. After other languages have been removed and only English texts have been stored by a Tike language detector, the two classes are merged into one file (vt_vs_vu) via a concatenation node. This file passes through several nodes, including cleaning steps as represented in Figure 9.1.

Now we get a clean document and term frequencies as a matrix table, based on which a binary vector can be created with a document vector node. The final step is to write the output file containing in each row the text of an instance, keywords and their frequencies, and the trust label.

In the processing phase, the generated file is forwarded to the row filter node to remove rows within a text length of less than 10 words. A column filter isolates the keywords, their frequencies, and labels, and then passes the file to a count loop node. In each loop, the samplings included in the file is equalized (under-sampling) and split randomly into training subset to create the decision trees and testing subset for the prediction by the learning algorithm (RF). In the prediction phase, the RF applies the testing subset over the generated decision tree models and represents its performance measures. At the end of the loops (100 runs), only the overage on the several performance measures, which should be stable now, is considered.

9.6 Results

This binary-classification is our reference for the short-text mining aimed at in this work. The classifier developed must be able to restore this classification with the best possible accuracy. According to our logic, very trusted class must contain more texts of higher quality than trusted class, which contains more texts of higher quality than untrusted class, and this class contains more texts of higher quality than very untrusted class. We can see in Table 9.1 the number of instances per trust class is imbalanced. Very trusted class of Genius and trusted class of Stackoverflow are more biased towards other classes. It is desirable to have a classifier that offers high prediction accuracy across all classes. This is a challenge and can be bypassed by binary classification.

Our novel work on interpretations could have a different perspective, for example, than Pitler's and Nenkova's in [15], who found that "longer articles are less well written and harder to read than shorter ones". This can be the opposite in the case of interpretations. As a rule, longer descriptions provide the necessary explanation. Under this assumption, this work carries out natural language processing (NLP) i.e. part-of-speech and several readability indexes as well as a machine leaning technique (ML) based on bag-of-words model applied by the random forest classifiers. The aim is to evaluate the interdependency of the features and their influence on quality. The NLP analysis results that there is a linear relation between quality of text and the metrics *present 3rd person singular* (VBZ), *present tense* (VBP), *base verb* (VB), *adverb* (RB), *possessive marker* (POS), *common noun* (NN), *adjective* (JJ), *gerund, present participle* (VBG), *past tense* (VBD), *plural common nouns* (NNS) and *plural proper noun* (NNP) (see Figure 9.2). This linear relation exists also within the metrics *number of characters*, *number of words*, *number of syllables* (see Figure 9.3). While such relation could not be found in terms of readability indexes and it is limit regarding the *number of complex words*.

However, the results of the random forest technique based on the bag-of-words model look much more promising. Table 9.4 gives an overview on the percentages of the factors considered.

The best classifiers performances are applied on the combinations of (very) trusted versus (very) untrusted, while the worst performances can be found on the combinations, which are close to each other (very trusted vs. trusted and untrusted vs. very untrusted). This is represented by *vt vs. t* (57%, 57% and 61%, 62% as F-measure and ACC of Genius and Stackoverflow respectively) and *vu vs. u* (54%, 55% and 57%, 53% as F-measure and ACC of Genius and Stackoverflow respectively), in contrast to higher performances in case of all other combination, mostly. For example, the highest distance can be found in the combination *vt vs. vu*, which is an evidence of their dissimilarity (69%, 68% and 70%, 66% as F-measure and ACC of Genius and Stackoverflow respectively). This is confirmed by

Figure 9.2: Part-of-Speech Analysis

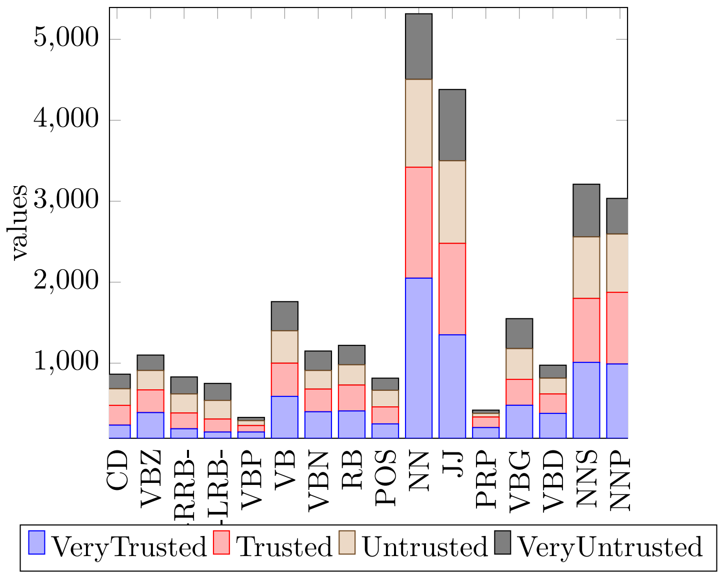
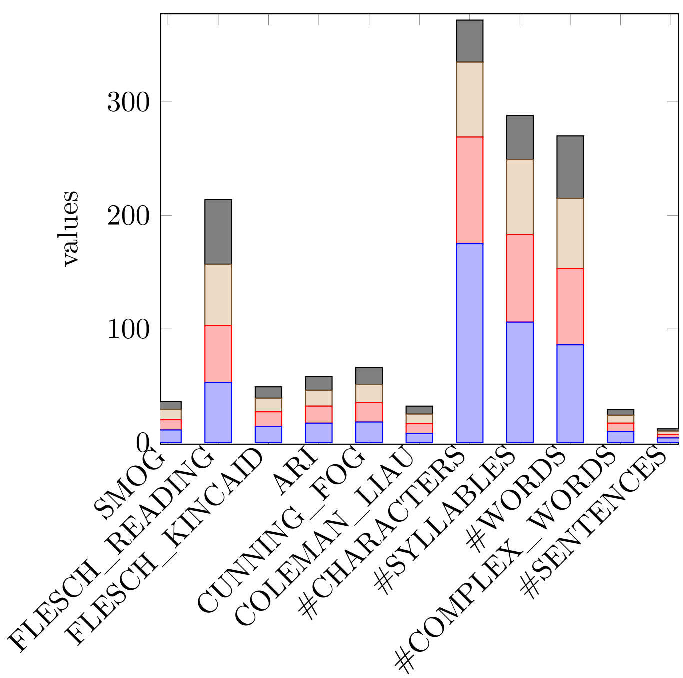


Figure 9.3: Readability Indexes Analysis



the highest MCC[5] measure of each combination and applies in both online communities and indicates the correctness of trust classification.

Considering the measures sensitivity and specificity, we can see that the classifiers are more able to correctly classify instances as (very) trusted in contrast with (very) untrusted,

[5]MCC has a range of [-1,1], where -1 specifies a totally incorrect binary classifier, while 1 specifies a totally correct binary classifier.

Table 9.4: Classifier Performance of the Random Forest based on Bag-of-Words Model

Community	Combination	Precision	Sensitivity	Specificity	F-Measure	Accuracy	MCC
Genius	vt vs t	0,57	0,58	0,56	0,57	0,57	0,13
	vt vs u	0,66	0,67	0,65	0,66	0,66	0,32
	vt vs vu	0,67	0,70	0,65	0,69	0,68	0,36
	t vs u	0,61	0,61	0,60	0,61	0,61	0,22
	t vs vu	0,64	0,65	0,63	0,65	0,64	0,28
	u vs vu	0,56	0,53	0,58	0,54	0,55	0,11
	vtt vs vuu	0,65	0,66	0,64	0,65	0,65	0,30
Stackoverflow	vt vs t	0,63	0,60	0,65	0,61	0,62	0,25
	vt vs u	0,65	0,73	0,60	0,69	0,67	0,33
	vt vs vu	0,64	0,80	0,52	0,70	0,66	0,34
	t vs u	0,58	0,57	0,58	0,57	0,58	0,15
	t vs vu	0,58	0,70	0,59	0,63	0,59	0,19
	u vs vu	0,52	0,64	0,42	0,57	0,53	0,06
	vtt vs vuu	0,62	0,64	0,60	0,63	0,62	0,24

This table summarizes the classifier performance of the Random Forest based on Bag-of-Words Model. vt=very trusted,t=trusted,u=untrusted, vu=very untrusted, vtt=very trusted+trusted, vuu=very-untrusted+untrusted, vtt=very trusted merged with trusted, vuu=very untrusted merged with untrusted.

in case of Genius (at highest 70%), while instances as (very) trusted in contrast with (very) untrusted, in case of Stackoverflow (at highest 80%). This indicates the ability to recognize trusted instances, which differ based on their content from other. On the other hand, in case of very trusted merged with trusted versus very untrusted merged with untrusted (vtt vs vuu), the true recognition of negative instances (very untrusted merged with untrusted vtt) is higher (66% and 64% of Genius and Stackoverflow respectively). This indicates that despite the proved dissimilarity of both classes, the classes trusted and untrusted are relative near located to each other, which is reflected by the performance decreased between these two classes specially in Stakoverflow (from 69% to 65% and 70% to 63% as F-measures of Genius and Stackoverflow respectively). This can be observed more clearly in the relative limited performance measures of the combination untrusted and very untrusted in both online communities.

In case of Stackoverflow, we investigate the distribution of answers that are marked as accepted by users. Accepted answers are located by 37%, 27%, 12% and 0% in the classes very trusted, trusted, untrusted and very untrusted, respectively as illustrated in Table 9.5. This is a one more evidence that high trust classes provide high quality content and that the classification based on the trust model is correct.

Table 9.5: Distribution of Accepted Answers over Trust Classes

Trust Class	Accepted Answers	Percentage
VT	313	37%
T	249	27%
U	44	12%
VU	1	0%

9.7 Discussion

This study investigates short-texts in the forms of interpretations and posts gained from the online communities Genius and Stackoverflow. These short-texts were classified in terms of trust by our trust model based on its metadata.

To answer our research questions, we applied syntactic analysis using an NLP approach combined with several readability indexes (see previous Chapter 8). Despite achieving some indications of the content difference of each class such as verbs and nouns frequencies, relation triples, number of complex words etc. due to the nature of short-text these indications cannot apply as strong evidences to distinguish the content. However, the RF classifier based on BoW model as features establishes such evidences.

9.7.1 RF classifier using BoW Model

The performance measures, resulted of both online communities Genius and Stackoverflow, show an average of accuracies 62% and 61% respectively and can be applied to recognize the texts of different qualities. This means that, the classification of the trust classes can be reconstructed to a certain degree based on its content. This can be used as a pre-process for the trust model to increase performances or in the case that metadata is not available. By considering the different performance measures of the binary-classification of the trust classes, it clearly shows that the classifier performs linear with the logical distance of the trust classes. That is, it shows relatively low accuracy and F-measure by making decisions on combinations of trust classes that are logical close to each other e.g. the combination of very-trusted vs. trusted, untrusted vs. very-untrusted or even trusted vs. untrusted . While it performs better on the combinations consisting of trust classes that have logical long distance between each other e.g. very-trusted vs. very-untrusted, very-trusted vs. untrusted. In addition, the results show that the accepted-answers in Stackoverflow corpus are distributed according to the trust degree of each class. That is, the higher the trust degree, the higher the percentage of the accepted-answers can be found; the percentages of accepted-answers are 37%, 27%, 12% and 0% in the trust classes Very Trusted, Trusted, Untrusted and Very Untrusted respectively.

9.7.2 Theoretical Contributions

This work follows the General Design Theory (GDT) provided by Takeda et al. [21]. GDT consists of the following processes: 1) Define goal: Short text regarding its quality check through the reconstruction of confidence classes. 2) Proposal: Approaches of the techniques to be applied (NLP and ML). 3) Development: Implementation using appropriate tools (Weka and KNIME). 4) Evaluation: Accepting and refining the results. 5) Conclusion: Decide which candidates form the solution. Our investigation on text complexity and lexical analysis is consistent with the Information Manipulation Theory (IMT), presented in the next subsection 9.7.3. The IMT uses key words (i.e. clearly, accurate, relate to and representation), which are reflected in this study by examining the multiple readability indexes and the tokens (BoW) used in creating and providing information. Exploring the readability indexes such as FOG, KINCAID, ARI etc., which calculate the text complexity degree, addresses the quantity aspect of the IMT. While, the metrics deployed in the trust model (e.g., authority and reader rating) imply consideration of relevance and presentation of information. Accordingly, this study follows the principle and supports IMT, which is widely considered as one of the most significant explanations for data manipulation in communication.

9.7.3 Information Manipulation Theory (IMT)

McCornack [14] developed the Information Manipulation Theory (IMT) that explores the behavior during information providing. According to Levine [12], IMT offers a multidimensional approach to the design of misleading messages and uses maxims as a framework for describing a variety of misleading message forms. The maxims that a truthful conversation includes are: Quantity represents a set of information a receiver is given in order to communicate clearly. That is, the degree of how much detail is delivered to the receiver to get idea about the information transferred. Quality refers to which extend is an information factual and accurate. Relevance refers to whether the provided information is related to the situation or topic of the conversation, and manner that considers information representation rather that the actual information itself.

9.7.4 Managerial Implications

Online communities can get benefit from our approach for reviewing the information provided on their platform. The first experiment based on Natural Language Processing (NLP) provides metrics such as verbs and nouns frequencies, relational triples and number of complex words that can be used as a guideline for users to improve the style of their content. On the one hand, an online community can evaluate the quality of contributions based on such metrics and then perform the appropriate actions (improve, return or remove). The second experiment with the Random Forest Classifier (RF), based on the Bag-of-Words (BoW) model, performs well and is able to identify content in terms of trustworthiness. This supports content filtering and reduces the overhead of low-quality content. Identifying high-quality content will improve the offering of an online community and increase the likelihood of users viewing it as a source of high-quality content.

9.8 Conclusion, Limitation and Future Scope of Research

This study examines short texts in the form of interpretations and contributions from the online communities Genius and Stackoverflow. These short texts were classified into four trust classes by our trust model based on its metadata. In order to investigate the relationship between a short text and the associated trust class, we examined several PoS and readability indices as characteristics of the short text classified in trust classes. Despite evidence that might be able to predict the trust level of a given short text, we have not been able to provide sufficient evidence of the influence of such characteristics on trust. Therefore, it is difficult to take these characteristics into account in the proposed trust model. However, the RF classifier, which is based on the BoW model as characteristics, specifies such evidence. The performance measurements resulting from the two online communities Genius and Stackoverflow show an average accuracy of 62% and 61%, respectively, and can be used to identify texts of different quality. In addition, the results show that the assumed responses are distributed in the Stackoverflow corpus according to the trust level of each class. That is, the higher the trust level, the higher the percentage of accepted answers can be found; the percentage of accepted answers is 37%, 27%, 12%, and 0% in the Very Trusted, Trusted, Untrusted, and Very Untrusted trust classes. The nature of the short texts means that, machine learning is limited as compared to such studies of "long" texts. In the case of online communities and especially social media, the short texts are usually informal and noisy (words in other languages, shortcuts, tokens, etc.). This makes the problem that needs to be addressed much more complex. Further work shall explore the possibility of developing a method that maps informal text into formal text by determining and replacing the meaning of the noises used by the user. In terms of the structure and writing style of the text, it can be a similar approach to improve the short text by modification, without additions that could change its meaning.

Bibliography

[1] J. Al Qundus, *Technical analysis of the social media platform genius*, tech. rep., Freie Universität Berlin, 03 2018.

[2] J. Al Qundus and A. Paschke, *Investigating the effect of attributes on user trust in social media*, in International Conference on Database and Expert Systems Applications, Springer, 2018, pp. 278–288.

[3] R. Barzilay and M. Lapata, *Modeling local coherence: An entity-based approach*, Computational Linguistics, 34 (2008), pp. 1–34.

[4] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, and B. Wiswedel, *Knime-the konstanz information miner: version 2.0 and beyond*, AcM SIGKDD explorations Newsletter, 11 (2009), pp. 26–31.

[5] A. Chalak, *Linguistic features of english textese and digitalk of iranian efl students*, Research in Applied Linguistics, 8 (2017), pp. 67–74.

[6] R. Flesch, *A new readability yardstick.*, Journal of applied psychology, 32 (1948), p. 221.

[7] V. Hatzlvassiloglou, J. L. Klavans, and E. Eskin, *Detecting text similarity over short passages: Exploring linguistic feature combinations via machine learning*, in 1999 Joint SIGDAT conference on empirical methods in natural language processing and very large corpora, 1999.

[8] M. Heilman, K. Collins-Thompson, J. Callan, and M. Eskenazi, *Combining lexical and grammatical features to improve readability measures for first and second language texts*, in Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, 2007, pp. 460–467.

[9] A. R. Hevner, S. T. March, J. Park, and S. Ram, *Design science in information systems research*, Management Information Systems Quarterly, 28 (2008), p. 6.

[10] C. Jia, M. B. Carson, X. Wang, and J. Yu, *Concept decompositions for short text clustering by identifying word communities*, Pattern Recognition, 76 (2018), pp. 691–703.

[11] M. Klassen and N. Paturi, *Web document classification by keywords using random forests*, in International Conference on Networked Digital Technologies, Springer, 2010, pp. 256–261.

[12] T. R. Levine, *Dichotomous and continuous views of deception: A reexamination of deception ratings in information manipulation theory*, Communication Research Reports, 18 (2001), pp. 230–240.

[13] S. T. March and G. F. Smith, *Design and natural science research on information technology*, Decision support systems, 15 (1995), pp. 251–266.

[14] S. A. McCornack, *Information manipulation theory*, Communications Monographs, 59 (1992), pp. 1–16.

[15] E. Pitler and A. Nenkova, *Revisiting readability: A unified framework for predicting text quality*, in Proceedings of the conference on empirical methods in natural language processing, Association for Computational Linguistics, 2008, pp. 186–195.

[16] N. Prat, I. Comyn-Wattiau, and J. Akoka, *Artifact evaluation in information systems design-science research-a holistic view.*, in PACIS, 2014, p. 23.

[17] Y. Rao, H. Xie, J. Li, F. Jin, F. L. Wang, and Q. Li, *Social emotion classification of short text via topic-level maximum entropy model*, Information & Management, 53 (2016), pp. 978–986.

[18] S. E. Schwarm and M. Ostendorf, *Reading level assessment using support vector machines and statistical language models*, in Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2005, pp. 523–530.

[19] J. L. Solka et al., *Text data mining: theory and methods*, Statistics Surveys, 2 (2008), pp. 94–112.

[20] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, *Short text classification in twitter to improve information filtering*, in Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, ACM, 2010, pp. 841–842.

[21] H. Takeda, P. Veerkamp, and H. Yoshikawa, *Modeling design process*, AI magazine, 11 (1990), p. 37.

[22] A. Todirascu, T. François, D. Bernhard, N. Gala, and A.-L. Ligozat, *Are cohesive features relevant for text readability evaluation?*, in 26th International Conference on Computational Linguistics (COLING 2016), 2016, pp. 987–997.

[23] B. Vijay and D. D. Rao, *Improving accuracy of named entity recognition on social media*, IJSEAT, 5 (2017), pp. 809–814.

[24] Q.-S. Xu and Y.-Z. Liang, *Monte carlo cross validation*, Chemometrics and Intelligent Laboratory Systems, 56 (2001), pp. 1–11.

Chapter 10

Conclusion

This chapter summarizes the thesis by discussing the objects investigated and reminding the reader about the objectives resulted briefly. In addition, we present the limitation, the benefits of this work and further work.

10.1 Discussion of the Objects Explored

The literature review shows that trust is mixed up and there is no clear distinction between several aspects. These aspects, e.g. quality, credibility, etc. could be related to trust, while many focus on manners such as reputation and vandalism systems other than on trust. Trust is investigated as a relationship between users with transitivity characteristic or between a user and a system and takes the values {0%, 100%}. The considered (Information) trust in this thesis builds on user-generated content using templates as annotations or posts providing metadata mainly related to other readers and authorship.

Trust is investigated from different perspectives, i.e. individual, interpersonal, relational, societal and in different relationship forms, i.e. one-to-one, one-to-many, many-to-one or many-to-many structures. The authors provide similar definitions of trust, which contain keywords such as belief, vulnerability, truthfulness, etc. and conditions such as risk, interdependence and a given situation. In addition, due to trust being a personal merit, this variety of dealing with such concern shows its complexity or so-called unpredictable simplicity. That is, almost everyone has an idea of what trust is, but it is difficult to accurately describe and generalize because individuals differ in personality and judgment. These are usually based either on personal experience or on an experience of trusted entities and here we have come back to the question of what is meant by "trusted entity".

The collaborative, free online encyclopedia platform Wikipedia is one of the most researched communities in terms of trust, quality, credibility, readability, etc. The focus is on identifying relevant features that are derived from the content-revision history and author information. Assessment criteria such as trust should play a role in changing the way information from social media is perceived and used. From the point of view of this work, the nature of the content explored in Wikipedia differs regarding length, format, and structure. There are also a couple of works that consider Twitter as a case study for the same concern. In order to meet the challenge of Twitter's short text, many of the offered approaches decide to extend the text with an external source based on topic similarity or text predictors for completing/adding sentences. Other approaches addressing the research area of machine learning consider author metadata as additional features to existing text mining techniques to improve their performance.

However, models, algorithms and different approaches have been developed on the topic of trust and are based on a broad of characteristics, independently - we bound such characteristics together regarding our case study into dimensions that are explored more in details in the next Section 10.2-. Related works research the aspect stability as the measure of content length in a period of time and develop functions that return the number of edits done at a time. The article stability, including the number of changes is considered as a factor of risks and applied mainly for examining vandalism. Other works deal with credibility as a concept related to trustworthiness. Credibility consists of five criteria, including accuracy, authority, objectivity, currency, and coverage (see Chapter 3) and builds on experiences of either known or credible people, that so-called trust-circle. Quality has been widely investigated as a fundamental criterion of text information and it is defined as "fitness for use" or as "user satisfaction". Quality is assessed based on classes that are content-based (e.g. text-style, structure), context-based (e.g. topic-focused, authority) and rating-based (e.g. search-ranking, popularity).

Algorithms for calculating trust, rely on several metrics. For example, the metrics (average time on a website, number of website visits, average daily visits and bounce rate) build the basis for computing trust. In some other works, the metrics (competence, proximity, popularity, recency and corroboration) are considered for such assessment (see Section 3.2.1). The aim is to map trust into a numeric value that indicates the trust degree of the text information. Natural language processing (NLP) based techniques are linguistically motivated and are applied in exploring texts regarding its lexical and syntactic features as well as semantic analysis. Mostly, such texts are noisy and need to be cleaned from some lexical items contained.

In terms of information trust modeling, relevant approaches can be divided into (1) logical approaches, including models that are based on mathematical logic, (2) computational approaches aiming to integrate trust models implementation into automatic systems and (3) socio-cognitive approaches, including models considering trust on the basis of direct experience (source statistical) or on the basis of a set of trustor- and environmental-features (cognitive). The majority of these models takes their outcomes as a feedback influencing the input factors and focus more on entities (trustor or agents) rather than content (see Section 3.3), in contrast to our work, which retains a certain balance, as will be explained in more detail in the following section.

10.2 Conclusion

In the last decade, the world has moved rapidly and purposefully towards digitization as a result of increasing globalization. For example, the number of printed news items is gradually decreasing. The flow of information is no longer one-to-many, but many-to-many. As a result, a lot of information has become available on the Internet and users can create and share information efficiently and easily. Nevertheless, the evaluation of user-generated content in terms of trust is becoming an important issue.

The interaction between users in online communities requires trust in the sub-objects of that community. Trust is indicated and evaluated by the activities of users on the information provided. Trust encourages users to consume information and make critical decisions. For example, trust between users has a strong influence on interactions, which can lead to the exchange of high-quality content. Therefore, it is necessary to develop models and strategies for user-generated content that offers an improved policy of user participation. That is, factors influencing trust must be understood in order to develop successful applications so that users are willing to participate in. This requires extending the human-to-human dimension by a human-to-machine dimension to build trust models.

There are a lot of terms related to trust e.g. un-trust, dis-trust, blind-trust, etc. (see Chapter 2). Trust is considered from the aspects of user-generated content and user. The user-aspect is classified in trust between users (human-human), trust between user and object (human-system) and trust between objects (system-system or agent-agent). This work focuses on the trust based on user-generated content that so-called information trust.

As trust is a complex subject and despite providing many definitions of it in the literature, it is hard to establish a universal definition of trust. Several definitions are introduced in Chapter 3. We value the definitions of trust proposed by the authors Mayer et al., which described by the willingness of one party to be vulnerable to the actions of another party, based on the expectation that the other party will perform a particular action important to the trustor, regardless of the ability to monitor or control that other party. And the definition of Corritore et al., which proposed as an attitude of confident expectation in an online situation with the risk that one's own vulnerability will not be exploited (see Chapter 6).

This study defines trustworthiness as a property of an object (i.e. content). While trust is the process performed by an entity (i.e. user) including interaction of that object with respect to vulnerability. The result of that process is the trustworthiness, which is indicated by user's activities and assessed by three concerns: (1) the number of activities conducted over a time period by other users (stability); (2) The types of such activities, user review and authorship (credibility); and (3) the nature of the content generated by the members of the elite-cycle or n-top-active users (quality). Based on this analysis, trust is formally defined as a correlation function of the dimensions: stability, credibility, and quality under the conditions 1) A risk, a user (trustor) should be vulnerable in the use of information provided, that is, the information is important. 2) Independence, an information provider (trustee) cannot be controlled. If I can control the provider, so the trust question is senseless, and 3) intention, the information provided could be incorrect, but it is not intentionally manipulated.

The approach of this dissertation relies on real data gathered from two online communities that are Genius and Stackoverflow. The corpus of Genius contains 1,306,560 activities carried out by 162,747 users on 77,806 unique pages. An overview of Genius's social- and technical aspects is provided in Chapter 5. The Stackoverflow's corpus contains 3,623 posts and 4,092 comments carried out by 3179 users. These corpora were analyzed regarding the metadata, e.g. voting, author-profile, etc. In the next step in this analysis phase, related works were explored in terms of metrics (edit history, ranking, accuracy etc.) related to trust. These metrics were compared and merged with metadata metrics according to their semantic and influence into three frameworks we call dimensions (stability, credibility, and quality). In the generation phase, a trust model was proposed that classify user-generated content based on the calculation of the dimensions into one of four possible classes (Very Trusted, Trusted, Untrusted and Very Untrusted), which were derived from the empirical analysis of the database distributions using the empirical cumulative distribution function (ECDF). Two evaluations of the trust model were performed. The first evaluation was based on a conjoint analysis approach to experience user preferences regarding the dimensions of the trust model. The respondents confirmed our trust model by accepting its dimensions. In addition, we could estimate the respondents' relative importance towards each dimension, which allowed improving the weights of the calculation used to value the trust degree. This is introduced and described in Chapter 7. In order to address the limitation of the trust model being built on only metadata and to conduct a second evaluation, we applied a natural language technique and a machine learning method to investigate the text content of the trust classes. The aim was to reconstruct the classification based on the text-embedded features. We could train a random-forest classifier based on the bag-of-words model to classify short-text towards trust with an accuracy of 62% on average. A multiple-class classifier based on a naive Bayes algorithm achieved an accuracy of 34% based on the natural language technique using lexical analysis, part-of-speech and readability indexes. The approaches and the results of this evaluation stage are introduced in Chapter 9.

The trust model proposed in this thesis is illustrated in Figure 1.1 Chapter 1. It consists of four stages: 1) Input that contains an annotation to be investigated and a set of metrics, which consists of metadata (e.g. rating, authorship etc.) and text-embedded features (i.e. bag-of-words). These metrics are taken into account at the next stage; 2) Calculation that assesses a concrete value of trust using equations operating on the metrics; the calculated trust value is passed at the 3) Interpretation stage that applies a predefined threshold to classify the value and maps it to 4) the Output in one of the trust classes (Very Trusted, Trusted, Untrusted and Very Untrusted).

The solutions presented in the work offer a reasonable answer to the research questions (see Chapter 1). The atomic components of the proposed trust model (see Chapter 6) represent the socio-technical parameters for building trust in collaborative environments. The metrics, number of edits, contribution evaluation (e.g. user rating), author evaluation (e.g. author rating) and author role (e.g. staff), represent the social parameters and must be accessible and recognizable for all users. While the metrics at the technical level of the application are time stamps of each edit, the distinction of LWPP and HWPP edits, percentage contribution assignment of each author (attribution), author(s) role(s) power and assigned permissions. These technical parameters must be provided by the application, for example, at a public interface.

Trust is recognized based on user activities. It does not matter whether a user makes an activity towards another user-generated content positively (e.g. generate an improvement) or negatively (e.g. disagree by down-voting). Activities being taken show that the content has raised users' interest and that the content is important. That is why users also make a decision regarding trust. Usage also includes consumption, which is an important aspect of this work. Consumption, even if we regard it as an activity, can still be performed as reading-and-applying-elsewhere. This kind of activity leaves invisible traces and is very difficult for us to track. For example, you will find an answer to your issue on Stackoverflow. We assume that you accept and trust this answer. But you do not give any rating or comment. This case can be very difficult to pursue in order to be taken into account in our trust model. The number of views of user-generated content may give an indication of such activity, but it is difficult to distinguish the number of views on bounce rate (negative) and consumption (positive).

In our work, quality is an important component of trust, but does not equal trust (see Chapter 6). Quality is assigned to authors of high role levels, who are experienced and follow a certain style when creating content. They ensure that the content is accurate, understandable, meaningful and supported by arguments. This is the only way they can increase their profile capital. That is why our trust model estimates quality based on metadata related to authors. These characteristics define the quality and have been proven in related work. However, quality is related to trust and a pre-stage towards trust.

As mentioned, trust is recognized based on user activities, the more users interact with content, the more or less trust there is in it. In other words, trust leads to a consequent increase in the number of interactions. The success of an application is measured by its degree of usage. Usage implies that users interact with the content. Therefore, trust in the contents of an application leads to the success of this application (see Chapter 1 and 3).

From the point of view of the trust model, an application should follow the proposed instructions regarding its socio-technical parameters in order to increase its chances of success. These parameters are illustrated by the trust dimensions and described in detail by the metrics contained in each dimension.

10.3 Limitation

The addressed topic covers many disciplines and attracted trust community. We received relevant feedback and ideas through reviewers of conferences and journals. Time and budget were in many places the main obstacles against developing the work further.

For instance, during our investigation for metrics related to trust, we planned to interview experts of Genius (staff or moderator of Genius) to obtain more insights and background of these metrics. Unfortunately, we did not receive any response that could bring us beyond this work. Providing such information would put us on the right track much earlier and saved us a lot of time to elaborate other directions of the current work.

Another example is the evaluation conducted through a survey (see Chapter 7). The evaluation was planned for local respondents (known people near in their location, who could we interview) in order to interview them and finding more about their impressions. However, due to lack of respondents, we decided on the second approach, where we asked the international respondents (unknown people, who get our survey request over a third connection channel); even that international respondents have the advantage that the results can be generalized and thus support the generalization of our model.

10.4 Recommendation

Nowadays, it is easy to find information, especially recommendations, about almost everything on the web. This remote service usually has a credibility problem. It is not always easy to assess the quality of the information provided. At the same time, the lack of known sources of information has made it more difficult to interact (consume) with such information. To meet this challenge, it is necessary to rely on an alternative strategy that is available within this information. Our trust model receives as input the metrics (e.g., number of comments, reader rating, author rating, etc.) of information in the form of annotation in social media. Next, it calculates the dimension stability, credibility, and quality and classifies the annotation into the output. As input, the trust model takes the metrics (e.g., number of comments, reader rating, author rating, etc.) of information in the form of annotation in social media, which are calculated together, and the result is classified. The user receives a human-readable interpretation of the result that the annotation can be trusted or not. The proposed trust model supports the identification of trusted information in collaborative environments. It can be used in various online communities that provide the appropriate metadata for the information provided. The trust model helps to filter the information, and thus reduces the information overload shared on the Web. Online communities can integrate the trust model into their development to increase the likelihood of their use, as users are able to easily identify trusted information. This work is aimed at promoting valuable knowledge sharing by improving application development using the proposed trust model. Thus, the model serves as a reference for the development of collaborative annotation applications, as we have shown that trust plays an important role as a bridge between information quality and information usage. The trust model has broad dimensions and is based on assessments and metrics derived from in-depth literature research. By separating the components of its mechanism, it is variable and flexible in its attitude to include additional metrics that are relevant for integration into another but similar areas.

This work provides a set of benefits that the following two lists summarize:

- Trust this thesis investigates is intuitive, limited and its formalism using a tangible model.
- This thesis analyzed the user real-data and implements a strategy to support user making-decision to or to not trust information provided.
- This thesis aims to encourage valuable knowledge sharing by improving application development using the trust proposed model.
- The trust model involves comprehensive dimensions and relies on evaluations and metrics derived from solid literature investigations, but due to the separation of the components of its mechanism, it is variable and flexible in its attitude to include further metrics that are relevant for integration in other but similar domains.

The following benefits are in terms of managerial implications:

- Providing a model of trust for UGC-oriented software, including a machine learning-based approach to classify short-text.
- Increasing the probability of software to be used and, as a consequence, to be successful.
- Assist users in making decisions to trust the information provided by making it easier for them to distinguish information to save time and effort.
- Better distinction between information of different quality. This reduces the explosion of the information provided and the complexity of the information validation techniques used.

10.5 Further Work

Additional directions for this work include implementing an application based on the trust model proposed, considering sense-making and fake-news and investigating the correlation among trust varieties e.g. un-trust, distrust, mistrust etc.

Implementing an application would have a high probability to gain high user-acceptance and become successful. This requires human sources, time and budget to be able to compete with other applications.

During exploring the topic of this dissertation, the near topics sense-making and fake-news gained some of our attention. Sense-making is that process, on which we create relations and connections between given inputs of our senses (retrieval information) and the objects (mental models) in our minds. We already experienced such process to deal, understand or react to an environment. This would support the tracing of human behaviours, such as decision-making on a complex issue such as trust. Nevertheless, it would be hard to generalize this concept for the same reason of dealing with trust and more research is needed.

In contrast, the topic of Fake-news has its challenge on the source side rather than on the content side. The content here is intentionally generated, that is, it carefully requires more research by the contributor, who are usually very influential organizations or entities. The more influence this information should have, the better the information will be created and presented. It will be very difficult to examine this information in terms of trust, as the focus should be more on the creators and their backgrounds and intentions. Therefore, new and improved mechanisms are required.

There are many concepts associated with trust. Some even form their own research area (e.g. untrust), others are a kind of illustration or complement to each other (mistrust vs. dismistrust). Although, the definitions of these concepts are mostly vague, they open up new dimensions for investigation and research. It would have been very interesting for this work to investigate the relationship of these concepts, but that would be far from the scope of this work.

Regarding trust in its variety, widely-believed and reputation or word-of-mouth are terms we faced during our research. The question is, which term has more influence on trust, i.e. the judgment that is given by a wide circle of experts from different disciplines or by a few experts in the field of the information provided?

Appendix

Appendix A

Table 10.1: Activity Types

Annotation activities (52 from 78)	**Other activities**
accepted annotation accepted their annotation	**accepted a suggestion** **accepted comment**
added suggestion to annotation added suggestion to description added suggestion to their annotation added suggestion to their description	**added a photo** **added comment to text**
archived comment	-
cosigned annotation cosigned description cosigned their annotation cosigned their description	-
created annotation created description	**created text**
deleted annotation deleted their annotation	**deleted text**
downvoted annotation downvoted comment downvoted description downvoted suggestion	**downvoted post**
edited annotation edited description edited their annotation edited their description	**edited metadata** **edited text**
-	**followed**
-	**gave access to forum**
incorporated annotation integrated comment integrated suggestion	
-	**locked**

Continued on next page

Table 10.1 – continued from previous page

Annotation activities (52 from 78)	Author activities
- - -	**made editor** **made educator** **made mediator**
marked annotation marked description marked their annotation	**marked as spam**
mentioned	-
merged annotation edit merged their annotation edit	**merged discography**
-	**moved post**
-	**pinned**
-	**posted**
proposed edit to annotation proposed edit to description proposed edit to their annotation proposed edit to their description	-
pyonged annotation pyonged description pyonged their annotation pyonged their description	**pyonged**
-	**registration**
rejected annotation rejected annotation edit rejected suggestion rejected comment rejected their annotation rejected their annotation edit	-
replied annotation replied their annotation	-
-	**unlocked**
-	**unpinned**
upvoted annotation upvoted suggestion upvoted description upvoted post upvoted comment	**upvoted post**
-	**verified lyrics**

This table shows the annotation activity types and all activity types, in which dead end activities are in bold.

Table 10.2: Activity Descriptions

Id	**Predicate**	**Object**	**Description**	**Regex**
1	accepted	- (their) annotation - (a) suggestion	Member with certain permissions accepted the annotation and (a) suggestion. a suggestion refers to annotation, while suggestion to page text http://genius.com/3289744, which we call comment.	$\{ab(c)d \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$
2	added	- photo - (their) annotation - (their) description - text	- member added a profile photo - member added a suggestion to (their) annotation - member added a suggestion to (their) description - member added a comment to page text.	$\{ab(c)d \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$
3	archived	- suggestion	Archive just hides the suggestion (an option between accepted and more likely rejected).	$\{abd \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$
4	cosigned	- (their) annotation - (their) description	You agree with the person. Similar to ^, but it can also be used when someone posts something that is not directly above you. Person 1: Illmatic's best song is *Halftime*. Person 2: I don't really like that song. Person 3: Cosign Person 1. http://genius.com/2541962.	$\{ab(c)d \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$

Continued on next page

Table 10.2 – continued from previous page

Id	Predicate	Object	Description	Regex
5	created	- annotation - description - text	-member created annotation on a pice of text - member created description to a page text - member created a page text	$\{abd\| a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$
6	deleted	- (their) annotation - text	-member deleted an annotation - member deleted a page text	$\{ab(c)d\| a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$
7	downvoted	- annotation - suggestion - comment - description - post	Decrement annotation's / suggestion's / post's IQ score. a suggestion is of a description or an annotation, while comment refers to the page text.	$\{ab(c)d\| a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$
8	edited	- (their) annotation - (their) description - metadata - text	Change annotation / description/ text or it's metadata.	$\{ab(c)d\| a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$

Continued on next page

Table 10.2 – continued from previous page

Id	Predicate	Object	Description	Regex
9	followed	- member - page	A member can follow a another member or a page.	$\{abd \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$
10	gave	access to forum	A member with certain permission like an Editor gave another member an access to forum.	$\{abd \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$
11	incorporated	annotation	Annotation participates in *transcription contest* into Genius annotation. http://genius.com/Scribe-a-thon-september-2015-annotated/.	$\{abcd \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, c \in \sum_{subject}, d \in \sum_{object}\}$
12	integrated	- suggestion - comment	Integrate lets you integrate the suggestion into the annotation (or the comment into the page text. Like archived but integrated is intensify accepted.	$\{ab(c)d \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$

Continued on next page

Table 10.2 – continued from previous page

Id	Predicate	Object	Description	Regex
13	locked	- page	Page is locked means 600 IQ's is required to edit it. When a Regulator or Moderator locks a page, only Editors and above can edit the text! Check out this annotation http://genius.com/3288589 for more details on locked pages. Verified Lyrics/Texts: when an artist *verifies* their text, the page is locked to everyone except staff. If an editor comes across text that appear incorrect but have been *verified* by an artist, they should hit up a member of staff to assist them! http://genius.com/3289756.	$\{abd \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$
14	made	- editor - educator - mediator - moderator - staff	A member has been promoted. Moderator and staff promotion happens rarely.	$\{abd \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$

Continued on next page

Table 10.2 – continued from previous page

Id	Predicate	Object	Description	Regex
15	marked	- (their) annotation - description - as spam	color coded (if as spam then it will no longer exist!) - marked this as Restates the line: that means your audience feels like you're simply saying what the artist said in different words. http://genius.com/7507130 - marked this as it's a stretch: means your audience finds your interpretation unlikely or hard to believe. http://genius.com/7507130 - marked this as Missing something: that means you should check the suggestions and proposed edits to try and improve it. http://genius.com/7507158.	$\{ab(c)d\| a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$
16	mentioned	- member - page	Member or content is referred in another content.	$\{ab\| a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}\}$

Continued on next page

Table 10.2 – continued from previous page

Id	Predicate	Object	Description	Regex
17	merged	- (their) annotation edit - discography	Often people will explain one part of a whole line because they don't understand the other. This can be easily fixed by choosing the best annotation (Annotation 1) then rejecting/deleting the other (Annotation 2) while incorporating the important comment from it (Annotation 2) into the better annotation (Annotation 1). http://genius.com/1435708 discography: artist ID will be removed and the discography will be added into the list of lyrics of the artists. Example: http://genius.com/activity_stream/show_details?[]=32188270.	$\{abd \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$
18	moved	- threads	Threads/contribution will be moved into *right* or suitable forums/ sections (Lit, sport...).	$\{ab \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}\}$
19	pinned	- forum threads	Up to five threads can be pinned to the top of any forum at a given Time. Think of the possibilities! Pinned forum expectations! Pinned threads for album clean up! All the pinned threads! http://genius.com/discussions/172847-New-feature-pinned-forum-threads	$\{ab \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}\}$

Continued on next page

Table 10.2 – continued from previous page

Id	Predicate	Object	Description	Regex
20	posted	- threads	Threads has been posted to a forum.	$\{ab \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}\}$
21	proposed edit to	- (their) annotation - (their) description	Making a propose to improve.	$\{abd \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$
22	pyonged	- member - page	Share a given page with all of their followers. http://genius.com/2544094	$\{abd \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$
23	registration		New member.	*Your * friend X is on Genius!* * Facebook, Google and Twitter

Continued on next page

Table 10.2 – continued from previous page

Id	Predicate	Object	Description	Regex
24	rejected	- (their) annotation (edit) - (a) suggestion	Content will not appear any more because it is rejected.	$\{ab(c)d \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$
25	replied	- (their) annotation	Only replies can be written to annotations created by the page artist (no suggestions).	$\{ab(c)d \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$
26	unpinned		see 19 pinned.	
27	upvoted	- annotation - suggestion - comment - description - post	Increment annotation's / suggestion's / post's IQ score. Suggestion is of a description or an annotation, while comment refers to the page text.	$\{ab(c)d \mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, (c \in \sum_{subject}), d \in \sum_{object}\}$
28	unlocked		See 13 locked.	

Continued on next page

Table 10.2 – continued from previous page

Id	Predicate	Object	Description	Regex
29	verified	- lyrics	Page is tagged (green marked) as checked.	$\{abd\mid a \in \sum_{subject} \setminus \{\epsilon, their\}, b \in \sum_{predicate}, d \in \sum_{object}\}$

This table shows description of the activities

Table 10.3: Role Permissions

Permissions	Whitehat	Artist	Mediator	Editor	Moderator	Staff
appearance	White	Green	Magenta	Yellow	Purple	Steel-blue
access						
chat	No	No	Yes	Yes	Yes	Yes
editorial board	No	No	Yes	Yes	Yes	Yes
facebook/twitter	No	No	Yes[a]	No	Yes[a]	Yes
general forum	Yes	Yes	Yes	Yes	Yes	Yes
moderation forum	No	No	Yes	Yes	Yes	Yes
user aliases	No	No	Yes	Yes	Yes	Yes
user report	No	No[a]	Yes	Yes	Yes	Yes
action						
annotate locked song	600+	600+	600+	Yes	Yes	Yes
chart beat	No	No	Yes	Yes	Yes	Yes
clear votes less 0	No	No	Yes	Yes	Yes	Yes
lock/unlock pages	No	own	600+	Yes	Yes	Yes
massage user	300+	300+	300+	Yes	Yes	Yes
mark as spam	No	No	Yes	No	Yes	Yes
move threads	No	No	Yes	No	Yes	Yes
penalty box	No	No	Yes	No	Yes	Yes
pin threads	No	No	Yes	No	Yes	Yes
warn/ban in chat	No	No	No	No	Yes	Yes
create						
content (ann., text, vote)	Yes	Yes	Yes	Yes	Yes	Yes
forum	No	No	No	No	No	Yes
forum-post	150+	150+	Yes	Yes	Yes	Yes
postlets	No	No	No	Yes	Yes	Yes
update album-tracklist	1000+	1000+	1000+	Yes	Yes	Yes

Continued on next page

Table 10.3 – continued from previous page

Permissions	Whitehat	Artist	Mediator	Editor	Moderator	Staff
upload a profile pic	Yes	Yes	Yes	Yes	Yes	Yes
delete						
forum-thread/post	No	No	No	Yes	Yes	Yes
song page	No	own	No	No	Yes	Yes
text page	No	own	No	Yes	Yes	Yes
edit						
album-tracklist	No	own	No	Yes	Yes	Yes
artist page	No	own	No	Yes	Yes	Yes
forum posts	No	No	Yes	No	Yes	Yes
locked page	No	own	600+	Yes	Yes	Yes
postlets	No	No	No	Yes	Yes	Yes
text page	No	own	No	Yes	Yes	Yes
promotion						
Editor	No	No	No	Yes	Yes	Yes
De-editor	No	No	No	No	Yes	Yes
Mediator	No	No	No	No	Yes	Yes
Moderator / De-mod	No	No	No	No	Mod. commune	Yes
verify Artist	No	No	No	No	Yes	Yes

This table shows the permissions of each user role. They are not complete.
[a]It could not be determined.

Appendix B

R Code :

```
library(ggplot2)
#-------------------------------------
#-- stability calculation based on the number of edits ---|
#-------------------------------------

#-- read file into a table: This file contains annotation IDs, ordering sequence
#and the stability
stability_tab =
read.table(" file:/../ stability_ann_id_rownr_orderd_stability.txt", sep=";",
col.names=c("annid","seq", "stability"), fill=FALSE, strip.white=TRUE)

#-- read values as numeric
z<-as.numeric(unlist(stability_tab[1]))
x<-as.numeric(unlist(stability_tab[2]))
y<-as.numeric(unlist(stability_tab[3]))
```

```
#-- Simulating the values from the normal distribution by the mean and standard
deviation.
stability <- round(rnorm(y, mean=mean(y), sd=sd(y)))

#-- converting values into data frame
stability_dataframe <- data.frame(stability)

#-- ploting data frame using Empirical Cumulative Density Function
ggplot(stability_dataframe, aes(stability)) + stat_ecdf(geom = "point")+
  labs(title="Empirical Cumulative Edits Function", y = "Percent", x="Edits")+
  theme(panel.grid.minor = element_line(color = "gray", size=0.5))+
  theme_set(theme_gray(base_size = 24))

#----------------------------------------
#-- credibility calculation based on the count of edits IQs ---|
#----------------------------------------
credibility_tab = read.table(" /../ credibility_annotationsIQs.txt", sep=";",
         col.names=c("seq", "annoIQ"),
         fill=FALSE,
         strip.white=TRUE)
x_c<-as.numeric(unlist(credibility_tab[1]))
y_c<-as.numeric(unlist(credibility_tab[2]))

#-- Simulating the values from the normal distribution by the mean and standard
deviation.
credibility_ann <- round(rnorm(y_c, mean=mean(y_c), sd=sd(y_c)))

#-- converting values into data frame
credibility_dataframe <- data.frame(credibility_ann)

#-- ploting data frame using Empirical Cumulative Density Function
ggplot(credibility_dataframe, aes(credibility_ann)) + stat_ecdf(geom = "point")+
 labs(title="Empirical Cumulative Edits Function",
    y = "Percent", x="Edits IQs")+
 theme(panel.grid.minor = element_line(color = "gray", size=0.5))+
 theme_set(theme_gray(base_size = 24))


#----------------------------------------
#-- credibility calculation based on the author IQ and author attribution ---|
#----------------------------------------
credibility_tab =
read.table("file: /../ credibilty_ann_id_authoriqMulAuthorattribution.txt",
sep=";", col.names=c("seq", "userIQ"), fill=FALSE, strip.white=TRUE)
x_u<-as.numeric(unlist(credibility_tab[1]))
y_u<-as.numeric(unlist(credibility_tab[2]))
```

```
#-- Simulating the values from the normal distribution by the mean and standard
deviation.
credibility_u <- round(rnorm(log(y_u), mean=mean(y_u), sd=sd(y_u)))

#-- converting values into data frame
credibility_dataframe <- data.frame(credibility_u)

#-- ploting data frame using Empirical Cumulative Density Function
ggplot(df_u, aes(credibility_dataframe)) + stat_ecdf(geom = "point")+
 labs(title="Empirical Cumulative User Function",
    y = "Percent", x="User IQs")+
 theme(panel.grid.minor = element_line(color = "gray", size=0.5))+
 theme_set(theme_gray(base_size = 24))


#------------------------------------------
#-- quality based edits types n-set of the top most active user. This can be
|
#-- calculated for specific n. These are two examples of whitehat and artist.|
#------------------------------------------
quality_u = read.table("file: /quality_most_active_user_role.txt", sep=";",
          col.names=c("seq","ann_id","whitehat_co", "artist_co","editor_co","mediator_co"
"moderator_co","regulator_co"),
          fill=FALSE, strip.white=TRUE)
x_u<-as.numeric(unlist(quality_u[1]))
ann_id<-as.numeric(unlist(quality_u[2]))
y_w<-as.numeric(unlist(quality_u[3]))
y_a<-as.numeric(unlist(quality_u[4]))
y_e<-as.numeric(unlist(quality_u[5]))
y_me<-as.numeric(unlist(quality_u[6]))
y_mo<-as.numeric(unlist(quality_u[7]))
y_r<-as.numeric(unlist(quality_u[8]))

#-- Simulating the values from the normal distribution by the mean and standard
deviation.
quality_u_w <- round(rnorm(y_w, mean=mean(y_w), sd=sd(y_w)))

#-- converting values into data frame
quality_dataframe_q_u <- data.frame(quality_u_w)

#-- ploting data frame using Empirical Cumulative Density Function
ggplot(quality_dataframe_q_u, aes(quality_u_w)) + stat_ecdf(geom = "point")+
 labs(title="Empirical Cumulative User Role Function",
    y = "Percent", x="Whitehat")+
 theme(panel.grid.minor = element_line(color = "gray", size=0.5))
```

```
#------ artist
#-- Simulating the values from the normal distribution by the mean and standard
deviation.
quality_u_a <- round(rnorm(y_a, mean=mean(y_a), sd=sd(y_a)))

#-- converting values into data frame
quality_dataframe_q_u <- data.frame(quality_u_a)

#-- ploting data frame using Empirical Cumulative Density Function
ggplot(quality_dataframe_q_u, aes(quality_u_a)) + stat_ecdf(geom = "point")+
 labs(title="Empirical Cumulative User Role Function",
    y = "Percent", x="Artist")+
 theme(panel.grid.minor = element_line(color = "gray", size=0.5))
```

Appendix C

Trustworthiness Java Code: **1. Genius**:

```
package main;

import pattern.Annotation;
import pattern.Author;
import pattern.RolePower;
import pattern.TrustDegree;
import static java.lang.Math.toIntExact;
import java.util.ArrayList;
import java.util.List;

public class TrustworthinessCalculator {

        public double stabilitylog;
        public double credibilitylog;
        public double qualitylog;
        public double stability;
        public double credibility;
        public double quality;
        public double trustworthiness;


        /**
         * {E(t) : t → φ |E : edits function,t : timestamp,φ ∈ Z}
         * Here, Z is a set of all integers.
         * {S = SUM{t=p,t=t0} E(t) |t &p : time stamp, E : edits function
         * @param edits
         * @return
         */
        private double calculateStability(Annotation annotation) {
```

```
        // take the number of edits since creation time to the last edit detected
        this.stability=LWPP + HWPP;
        this.stabilitylog = LogBaseX(this.stability, 10);
        return this.stability;
}
/**
 * ∀ e ∈ editors{UCCF+ = attribution × rolepower \times IQ}
 * editsTypes = IQ × (|HWPP| ÷ |LWPP|)
 * credibility = f(UCCF,editsTypes)
 * @param value
 * @return
 */
private double calculateCrediblity(Annotation annotation) {
        // calculate UCCF
        double UCCF = calculateUCCF(annotation);

        //calculate editsTypes
        double EDITSTYPES = annotation.votes_total × calculateHL_PP(annotation);

        this.credibility = (UCCF + EDITSTYPES)÷ 2;

        this.credibilitylog = LogBaseX(this.credibility, 10);

        return this.credibility;
}

double HWPP =0, LWPP=0;
/**
 * @param annotation
 * @return
 */
private double calculateHL_PP(Annotation annotation) {
        if(annotation.pinned)
                LWPP+=1;
        if(annotation.twitter_share_message != null)
                if(!annotation.twitter_share_message.equals("\"?\" -@Genius"))
                LWPP+=1;
        LWPP += annotation.votes_total + annotation.pyongs_count;

        if(annotation.accepted_by != null)
                HWPP+=1;
        if(annotation.custom_preview != null)
                HWPP+=1;

        //for created_at
        HWPP+=1;

        HWPP += annotation.comment_count + annotation.proposed_edit_count;

        return (HWPP/LWPP);
}
/**
```

```
 *
 * @param annotation
 * @return
 */
private double calculateUCCF(Annotation annotation) {
        double tmp =0;

        // calculate UCCF = attribution × role power user IQ
        for(int i =0; i< annotation.authors.length; i++)
                tmp += annotation.authors[i].attribution ×
                calculateROLEPOWER(annotation.authors[i]) ×
                annotation.authors[i].user.iq;
        return tmp;
}
/**
 *
 * @param author
 * @return
 */
private double calculateROLEPOWER(Author author) {

        // the roles NULL OR CONTRIBUTOR are Whitehats
        if(author.user.role_for_display == null)
                return author.attribution * (RolePower.FACTOR ×
                RolePower.WHITEHAT_PERMISSION);
        else if(author.user.role_for_display.equals(RolePower.WHITEHAT))
                return author.attribution * (RolePower.FACTOR ×
                RolePower.WHITEHAT_PERMISSION);
        else if(author.user.role_for_display.equals(RolePower.ARTIST))
                return author.attribution × (RolePower.FACTOR ×
                RolePower.ARTIST_PERMISSION);
        else if(author.user.role_for_display.equals(RolePower.EDITOR))
                return author.attribution × (RolePower.FACTOR ×
                RolePower.EDITOR_PERMISSION);
        else if(author.user.role_for_display.equals(RolePower.MEDIATOR))
                return author.attribution × (RolePower.FACTOR ×
                RolePower.MEDIATOR_PERMISSION);
        else if(author.user.role_for_display.equals(RolePower.MODERATOR))
                return author.attribution × (RolePower.FACTOR ×
                RolePower.MODERATOR_PERMISSION);
        else if(author.user.role_for_display.equals(RolePower.STAFF))
                return author.attribution × (RolePower.FACTOR ×
                RolePower.STAFF_PERMISSION);
        return 0;
}
/**
 * quality = f(UCCFâĂš,editsTypesâĂš)
 * where:
 * UCCFâĂŹ symbolises User Credibility Correction Factorn-topActiveUser.
 * editsTypesâĂŹ symbolises editsTypesn-topActiveUser .
 * @param value
```

```
 * @return
 */
private double calculateQuality(Annotation annotation, int n_top_active_user) {

        if ( n_top_active_user >= annotation.authors.length) {
                System.out.println("n_top_active_user is too high");
                return this.quality = this.credibility;
        }
        // calculate UCCF
        double UCCFq = calculateUCCFq(annotation, n_top_active_user);

        //calculate editsTypes
        double EDITSTYPESq = annotation.votes_total ×
        calculateHL_PPq(annotation, n_top_active_user);
        this.quality = (UCCFq + EDITSTYPESq)÷2;
        this.qualitylog = LogBaseX(this.quality, 10);
        return 0;
}
/**
 *
 * @param annotation
 * @param n_top_active_user
 * @return
 */
private double calculateHL_PPq(Annotation annotation, int n_top_active_user) {
        double tmp =0;

        // calculate UCCF = attribution ∗ role power user IQ
        for(int i =0; i< n_top_active_user ; i++)
                tmp += annotation.authors[i].attribution ×
                calculateROLEPOWER(annotation.authors[i]) × annotation.authors[i].user.iq
        return tmp;
}
/**
 *
 * @param annotation
 * @param n_top_active_user
 * @return
 */
private double calculateUCCFq(Annotation annotation, int n_top_active_user) {
        if(annotation.pinned)
                LWPP+=1;
        if(annotation.twitter_share_message != null)
                if(!annotation.twitter_share_message.equals("\"?\" -@Genius"))
                LWPP+=1;
        if((annotation.votes_total + annotation.pyongs_count) <=n_top_active_user)
                n_top_active_user = 0;
        LWPP += annotation.votes_total + annotation.pyongs_count - n_top_active_user;

        if(annotation.accepted_by != null)
                HWPP+=1;
        if(annotation.custom_preview != null)
```

```
                        HWPP+=1;

                //for created_at
                HWPP+=1;

                if((annotation.comment_count + annotation.proposed_edit_count)
                <=n_top_active_user)
                        n_top_active_user = 0;

                HWPP += annotation.comment_count + annotation.proposed_edit_count
                - n_top_active_user;

                return (HWPP/LWPP);
        }
        /**
         *
         * @param edits
         * @return
         */
        private double calculateTrustwotheniss() {
                return this.trustworthiness = (0.24 × this.stability + 0.35 ×
                this.credibility + 0.41 × this.quality)/3;
        }
        public TrustDegree trustDegreeTranslator(Annotation annotation, int n_top_active_user) {
                calculateCrediblity(annotation);// it must be frist calculated
                calculateStability(annotation);
                calculateQuality(annotation, n_top_active_user);
                calculateTrustwotheniss();

                // do logarithm modification
//              System.out.print("vor "+this.trustworthiness);
                this.trustworthiness = LogBaseX(this.trustworthiness, 10);
//              System.out.println(" nach "+this.trustworthiness);
                // set label
                double VT,T, U, VU;
                if(this.trustworthiness >= VT)
                        annotation.TRUST_LABEL=TrustDegree.VERY_TRUSTED;
                else if(this.trustworthiness >= T)
                        annotation.TRUST_LABEL=TrustDegree.TRUSTED;

                else if(this.trustworthiness >= U)
                        annotation.TRUST_LABEL=TrustDegree.UNTRUSTED;

                else //VU
                        annotation.TRUST_LABEL=TrustDegree.VERY_UNTRUSTED;

                return annotation.TRUST_LABEL;

        }

  private double LogBaseX(double x, double base){
    double value= Math.log(x) ÷ Math.log(base);
```

```
    return Math.rint(value*1000)÷1000;
   }

}
```

2. Stackoverflow:

```
package main;

import pattern.Answer;
import pattern.Post;
import pattern.TrustDegree;
import static java.lang.Math.toIntExact;
import java.io.BufferedWriter;
import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileOutputStream;
import java.io.IOException;
import java.io.OutputStreamWriter;
import java.io.PrintWriter;
import java.io.UnsupportedEncodingException;
import java.nio.file.Files;
import java.nio.file.Paths;
import java.util.ArrayList;
import java.util.List;
import org.jsoup.Jsoup;
import org.unbescape.html.HtmlEscape;

public class TrustworthinessCalculator {

        public double stabilitylog;
        public double credibilitylog;
        public double qualitylog;
        public double stability;
        public double credibility;
        public double quality;
        public double trustworthiness;


        /**
         * {E(t) : t → φ|E : edits function,t : timestamp,φ ∈ Z}
         * Here, Z is a set of all integers.
         * {S = SUM{t=p, _t=t0} E(t) |t &p : time stamp, E : edits function
         * @param edits
         * @return
         */
        private double calculateStability(Answer answer) {

                // take the number of edits since creation time to the last edit detected
                this.stability=answer.post.viewCount + answer.post.commentCount +
                answer.post.favoriteCount;
                this.stabilitylog = LogBaseX(this.stability, 10);
```

```
        return this.stability;
}
/**
 * ∀ e ∈ editors{UCCF+ = attribution × rolepower × IQ}
 * editsTypes = IQ âĹŮ (|HWPP| Ãů |LWPP|)
 * credibility = f(UCCF,editsTypes)
 * @param value
 * @return
 */
private double calculateCrediblity(Answer answer) {
        // calculate UCCF
        double UCCF = calculateUCCF(answer);
        //calculate editsTypes
        double EDITSTYPES = answer.post.score × calculateHL_PP(answer);

        this.credibility = (UCCF + EDITSTYPES)÷2;

        this.credibilitylog = LogBaseX(this.credibility, 10);


        return this.credibility;
}

double HWPP =0, LWPP=0;
/**
 *
 * @param annotation
 * @return
 */
private double calculateHL_PP(Answer answer) {
        HWPP=answer.comments.size()+answer.post.answerCount;
        //+ answer.post.tags.split(";").length;
        LWPP = answer.post.viewCount+ answer.post.favoriteCount+ answer.post.score
        /*activity of voting*/;

        return (LWPP > 0?HWPP/LWPP:HWPP);
}
/**
 *
 * @param annotation
 * @return
 */
private double calculateUCCF(Answer answer) {
        double tmp =0;

        // calculate UCCF = attribution × role power user IQ
        for(int i =0; i< answer.authors.size(); i++)
                tmp += (answer.authors.get(i).contributionTextLength
                /sumOfTextLength(answer))
                                /* * calculateROLEPOWER(annotation.authors[i])*/
                                * answer.authors.get(i).reputation;
```

```
        return tmp;
}
/*
 * attriubtion based on the text length of a user to the rest.
 */
private int sumOfTextLength(Answer answer) {
        int length_comments_texts =answer.post.body.length();
        for(int i=0; i<answer.comments.size();i++)
                length_comments_texts+=answer.comments.get(i).text.length();
        return length_comments_texts;
}
/**
 * quality = f(UCCFâĂš,editsTypesâĂš)
 * where:
 * UCCFâĂŹ symbolises User Credibility Correction Factorn-topActiveUser.
 * editsTypesâĂŹ symbolises editsTypesn-topActiveUser .
 * @param value
 * @return
 */
private double calculateQuality(Answer answer, int n_top_active_user) {

        if ( n_top_active_user >= answer.authors.size()) {
                System.out.println("n_top_active_user is too high");
                return this.quality = this.credibility;
        }
        // calculate UCCF
        double UCCFq = calculateUCCFq(answer, n_top_active_user);
        //calculate editsTypes
        double EDITSTYPESq = answer.post.score ×
        calculateHL_PPq(answer, n_top_active_user);
        this.quality = (UCCFq + EDITSTYPESq)÷2;
        this.qualitylog = LogBaseX(this.quality, 10);
        return 0;
}
/**
 *
 * @param annotation
 * @param n_top_active_user
 * @return
 */
private double calculateHL_PPq(Answer answer, int n_top_active_user) {
                HWPP=answer.comments.size()+answer.post.answerCount
                /*+ answer.post.tags.split(";").length*/ - n_top_active_user ;
                LWPP = answer.post.viewCount+ answer.post.favoriteCount+ answer.post.score
                /*activity of voting*/;
        return (LWPP > 0?HWPP/LWPP:HWPP);
}
/**
 *
 * @param annotation
 * @param n_top_active_user
```

```
         * @return
         */
        private double calculateUCCFq(Answer answer, int n_top_active_user) {
                double tmp =0;

                // calculate UCCF = attribution * role power user IQ
                for(int i =0; i< n_top_active_user; i++)
                        tmp+=(answer.authors.get(i).contributionTextLength/sumOfTextLength(answer))
                                        /* * calculateROLEPOWER(annotation.authors[i])*/
                                        * answer.authors.get(i).reputation;
                return tmp;
        }
        /**
         *
         * @param edits
         * @return
         */
        private double calculateTrustwotheniss() {
                return this.trustworthiness = (0.24 × this.stability + 0.35 ×
                this.credibility + 0.41 × this.quality);
        }
        public List<Integer> vt_ids = new ArrayList<>();
        public List<Integer> t_ids = new ArrayList<>();
        public List<Integer> u_ids = new ArrayList<>();
        public List<Integer> vu_ids = new ArrayList<>();
        public TrustDegree trustDegreeTranslator(Answer answer, int n_top_active_user) {
                calculateCrediblity(answer);// it must be frist calculated
                calculateStability(answer);
                calculateQuality(answer, n_top_active_user);
                System.out.print(answer.post.Id);
                calculateTrustwotheniss();

                // do logarithm modification
//              System.out.print("vor "+this.trustworthiness);
                this.trustworthiness = LogBaseX(this.trustworthiness, 10);
//              System.out.println(" nach "+this.trustworthiness);
                // set label
                double VT, T, U, VU;
                if(this.trustworthiness >= VT) {
                        answer.TRUST_LABEL=TrustDegree.VERY_TRUSTED; vt_ids.add(answer.post.Id);
                        printIntoFile(HtmlEscape.unescapeHtml(answer.post.body)+
                        "\n -----POST------","data/veryTrusted_stackoverflow.txt", 0);
                }else if(this.trustworthiness >= T) {
                        answer.TRUST_LABEL=TrustDegree.TRUSTED; t_ids.add(answer.post.Id);
                        printIntoFile(HtmlEscape.unescapeHtml(answer.post.body)+"
                        \n -----POST------","data/trusted_stackoverflow.txt", 1);
                }else if(this.trustworthiness >= U) {
                        answer.TRUST_LABEL=TrustDegree.UNTRUSTED; u_ids.add(answer.post.Id);
                        printIntoFile(HtmlEscape.unescapeHtml(answer.post.body)+
                        "\n -----POST------","data/untrusted_stackoverflow.txt", 2);
                }else { //VU
                        answer.TRUST_LABEL=TrustDegree.VERY_UNTRUSTED; vu_ids.add(answer.post.Id);
```

```
                        printIntoFile(HtmlEscape.unescapeHtml(answer.post.body)+
                        "\n -----POST------","data/veryUntrusted_stackoverflow.txt", 3);
                }
  return answer.TRUST_LABEL;
        }
  private double LogBaseX(double x, double base){
    double value= Math.log(x) / Math.log(base);
    return Math.rint(value*1000)/1000;
   }

  public PrintWriter writerVT = null, writerT = null, writerU = null, writerVU = null;
        private void printIntoFile(String str, String file, int filenr) {
                try {
                        if(filenr == 0) {// vt
                                if(writerVT == null)
                                        writerVT = new PrintWriter(file, "UTF-8");

                                writerVT.println(str);
                        }
                        else if(filenr == 1) {// t
                                if(writerT == null)
                                        writerT = new PrintWriter(file, "UTF-8");

                                writerT.println(str);
                        }
                        else if(filenr == 2) {// u
                                if(writerU == null)
                                        writerU = new PrintWriter(file, "UTF-8");

                                writerU.println(str);
                        }
                        else if(filenr == 3) {// vu
                                if(writerVU == null)
                                        writerVU = new PrintWriter(file, "UTF-8");

                                writerVU.println(str);
                        }
                } catch (FileNotFoundException | UnsupportedEncodingException e) {
                        // TODO Auto-generated catch block
                        e.printStackTrace();
                }
        }
}
```