

# Open Research for Open Data

## Integration of Netnography and Text Mining for Consumer Insight Generation from Online Social Media

**Abstract:** The thesis aims at enhancing the understanding of netnography and text mining methods, as well as creating a mixed methods research framework that integrates the two methods for consumer insight generation from online social media in a commercial consumer research context. It relies on a qualitative and abductive approach. The thesis finds that the suitability for integration of netnography and text mining is high. The thesis also finds that the approach consists of three key participants: researcher, insight user and netizen and follows a within-and-across stage mixed-model design typology. A research model for the mixed methods research approach is further created, which consists of the following sections: (1) New research requirement; (2) Definition of research field; (3) Social media selection and data collection; (4) Netnographical and text mining analysis, interpretation and legitimation; (5) Insight; (6) Value. The information exchange dynamics between the key participants, different stages and the two methods are also analyzed. The thesis finds that each stage in the process can result in moving to the next or previous stage and that during this process the two methods interact through elements, inter-transform and holistically unite as two opposites. The thesis ends by a discussion of the results, including their managerial implications, reflection of the research as well as providing suggestions for future research.

**Keywords:** Big data, Mixed methodology, Market research, Netnography, Text mining

**Author:**

Wei Wei (40481)

**Tutor:**

Robin Teigland

**Opponents:**

Monika Ociecek

Ludvig Segelmark

**Examiner:**

Angelika Löfgren

---

# ACKNOWLEDGEMENTS

The author would like to give special thanks to...

## **Mentor:**

### **Robin Teigland**

Thank you for your great knowledge, constructive guidance, valuable inspiration and contact throughout the thesis. Your incredible dedication and passion for knowledge have truly been an inspiration to me for my thesis, career and spiritual life.

## **Interviewees:**

### **ResearcherX**

Thank you for sharing with me your insightful thoughts and practical knowledge about the mixed methods research, your experience of relevant projects' management as well as testing the model built in the thesis with real-life data together with me.

### **ResearcherY**

Thank you for sharing with me your deep academic and practical knowledge regarding ethnography, netnography and their application in business as well as your experience in relevant research projects.

### **ResearcherZ**

Thank you for sharing with me your rich experience and opinion regarding the management of research projects that are relevant to the thesis and valuable insight regarding the clients' feedback.

### **Lars Iversen**

Thank you for sharing with me your in-depth and “disruptive” understanding of consumer insight and the role it plays in innovation and marketing, your rich experience in innovation as well as your creative and thought-provoking advice for the method's model building.

### **Paola Peretti**

Thank you for sharing with me your experience of conducting academic research that integrate netnography and text mining as well as your unique opinion on this method.

### **Magnus Söderlund**

Thank you for sharing with me your profound understanding of consumer insight, its implication and the comparison of different methods of consumer insight generation.

### **Hans Kjellberg**

Thank you for sharing with me your rich knowledge and insight for consumer insight and different research methods, especially the thick and contextual understanding of consumer.

*“Science is wonderfully equipped to answer the question ‘How?’  
but it gets terribly confused when you ask the question ‘Why?’.”*

*-Erwin Chargaff*

# Table of content

<b>Chapter 1 Introduction</b>	<b>7</b>
1.1 Background	7
1.2 Problem area	9
1.3 Purpose	9
1.4 Research questions	9
1.5 Delimitations	10
1.6 Expected knowledge contribution	10
1.7 Definitions	10
1.8 Disposition and overview	11
<b>Chapter 2 Research background and theoretical overview</b>	<b>12</b>
2.1 A changing research environment – social media & user generated content is redefining the boundary of market research	12
2.1.1 Social Media	12
2.1.2 Social media data	13
2.1.2.1 Social media content	13
2.1.2.2 Social media data vs. big data	14
2.1.3 Social media analytics bring in new opportunities	14
2.1.4 Social media data analysis guideline	15
2.2 Netnography	15
2.2.1 Market research's thick description and netnography	15
2.2.2 Netnography's research process	16
2.2.3 Strengths and weaknesses of netnography	17
2.3 Text mining	17
2.3.1 Overview	17
2.3.2 Text mining's research process	18
2.3.3 Strengths and weaknesses of text mining	19
2.4 Mixed methods research	19
2.4.1 Debate between qualitative and quantitative research	19
2.4.2 Similarities among different paradigms	20
2.4.3 Need for mixed methods research	20
2.4.4 Interdisciplinary commercial market research innovation	22
2.4.5 Factors and guidelines of mixing research methods	23
2.5 Research model: a synthesis of literature	25

<b>Chapter 3 Methodology</b>	<b>26</b>
3.1 <i>Research strategy</i>	26
3.1.1 Epistemological considerations: hermeneutic	26
3.1.2 Ontological considerations: constructionism	27
3.1.3 Abduction	27
3.1.4 Qualitative approach	28
3.2 <i>Research design</i>	28
3.2.1 Design of theoretical study	28
3.2.2 Design of empirical study	29
3.2.3 Selection of institutions and representatives	30
3.2.3.1 Market research practitioner representative – Company A	30
3.2.3.2 Academic scholar representatives – Stockholm School of Economics and IULM University	30
3.2.3.3 Insight user representative – Springfellow International	31
3.2.4 Empirical data collection and analysis	32
3.3 <i>Limitations</i>	33
3.4 <i>Research quality</i>	33
3.4.1 Credibility	34
3.4.2 Transferability	34
3.4.3 Dependability	34
3.4.4 Confirmability	34
<b>Chapter 4 Empirical findings</b>	<b>35</b>
4.1 <i>Interview with Researcher X</i>	35
4.2 <i>Interview with Researcher Y</i>	36
4.2.1 First interview	37
4.2.2 Second interview	38
4.2.3 Third interview	39
4.3 <i>Interview with Researcher Z</i>	40
4.4 <i>Interview with Lars Iversen</i>	42
4.5 <i>Interview with Paola Peretti</i>	45
4.6 <i>Interview with Magnus Söderlund</i>	47
4.7 <i>Interview with Hans Kjellberg</i>	51
<b>Chapter 5 Analysis</b>	<b>53</b>
5.1 <i>Insight – the objective of market research</i>	53
5.2 <i>Suitability for integrating netnography and text mining methods</i>	54
5.2.1 Strengths of netnography	54
5.2.2 Weaknesses of netnography	56

5.2.3 Strengths of text mining	57
5.2.4 Weaknesses of text mining	58
5.2.5 Synergy created by integrating netnography and text mining	59
5.2.5.1 Complementary strengths	59
5.2.5.2 Overlapping weaknesses	61
5.2.5.3 Other benefits	61
5.2.5.4 Conclusion: a clear suitability for integration	62
5.3 <i>Development of a mixed methods research framework</i>	62
5.3.1 Key participants	62
5.3.2 Typology	63
5.3.2.1 Mixed-model or mixed-method design	63
5.3.2.2 Within or across the stages of the research	63
5.3.3 Strategic process model	64
5.3.3.1 From the cross section angle	64
5.3.3.2 From the vertical section	66
5.3.3.3 Information exchange dynamics	68
<b>Chapter 6 Conclusion</b>	<b>72</b>
<b>Chapter 7 Discussion</b>	<b>73</b>
7.1 <i>Managerial implications</i>	73
7.1.1 Opportunities	73
7.1.2 Challenges	73
7.1.2.1 Open research for open data acceptance	74
7.1.2.2 Evolution of competencies and collaboration	74
7.1.2.3 Concern of overlapping weaknesses	74
7.2 <i>Reflections on the study</i>	75
7.2.1 Generalizability	75
7.2.2 Open data as a window	75
7.2.3 Critical reflections	76
7.3 <i>Suggestions for further research</i>	76
<b>Bibliography</b>	<b>77</b>
<b>Appendix</b>	<b>80</b>

# Chapter 1 Introduction

*This chapter introduces the reader to the field of the thesis leading to the problem area and purpose. It then presents the questions, delimitations and key definitions of the thesis.*

## 1.1 Background

The business competition landscape is changing drastically everyday. With companies' strategic transformation from product-driven to market-driven on to design-driven (Verganti, R., 2009), consumer insight is becoming the cornerstone of companies' navigation and decision-making. Companies spend more than 32 million U.S. dollars globally every year on market research, 95% of which is spent on traditional research such as focus groups and surveys (Esomar, 2009). However, the drawback of traditional market research methodologies is that it relies heavily on consumers' recall of their behavior (Hayward, 2009) rather than ongoing, that researchers' selection of questions may ignore or twist issues that are important for customers and that consumers themselves are sometimes not able to be fully aware of and express their own real experiences and emotions (Kearon and Earls, 2009). The nature of the traditional market research methodologies may risk missing the consumer insights that are truly valuable to enterprises. Therefore, there is a growing realization that the traditional market research models from a top-down authoritarian approach are increasingly under challenges (Comley, 2008), with a declining response rate (Johnston and Harris, 2006). Also, the globalization of business is a road of no return and it requires enterprises to understand various consumer groups globally, which creates challenges for traditional consumer insight generation methods.

Meanwhile, consumers all around the world are going online. To understand consumers' behaviors, one must understand their activities in the online world. In the recent two decades, Internet has transformed from an occasional tool to one of the preliminary ways people connect, shop, learn, entertain and work. In 2000, there were around 400 million<sup>1</sup> Internet users world wide, and in 2013 the number became 2,749 million<sup>2</sup>. Analyzing large data sets from Internet is becoming the basis for companies' competing edge, new waves of profitability growth and innovation. "Leaders in every sector will have to grapple with the implications of big data, not just a few data-oriented managers", according to research by MGI and McKinsey's Business Technology Office<sup>3</sup>.

---

<sup>1</sup> <http://www.websiteoptimization.com/bw/1103/>

<sup>2</sup> <http://nethawk.net/2013/05/2013-q1-internet-world-statistics/>

<sup>3</sup> [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation?cid=other-eml-ttn-mip-mck-oth-1404](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation?cid=other-eml-ttn-mip-mck-oth-1404)

Among today's Internet users, 72% are now active on social media<sup>4</sup>. In the US, 16 minutes<sup>5</sup> of every hour is spent social networking. Social media is becoming the second nature of human beings. Consumers participate, share, collaborate and generate content through social medias. More and more enterprises turn their focus on how to fully utilize information from social medias for purposes such as consumer understanding, marketing communication, consumer involvement and open innovation. Social media and user-generated content have fundamentally changed market research's horizon. Market research's environment is changing in the way that consumers not only provide answers but also play a more and more proactive and creative role in shaping the areas of research (Comley, 2008).

Lyman and Wakeford (1999, p.359) claimed 15 years ago that "...the study of digital and networked technologies is one of the fastest growing fields of research in the social sciences", and this statement has only been strengthened through time. With the drastic development of the Internet, data becomes increasingly cheap and more available. However, more data also creates information overload for analytics. To utilize effective tools to extract actionable insight out of the noise is becoming more and more critical for business. Hence, big data approach is rapidly emerging in recent years due to its unparalleled advantage in volume, velocity and variety (Andrew McAfee and Erik Brynjolfsson, 2012). The digital footprints left by consumers through online medias, with its abundant supply of opinions and real-time occurrences, create a gigantic digital mine for market researchers to dig in. Among all big data analytics techniques, text mining is a powerful tool to mine textual data and transfer into consumer insight. However, there are several criticisms regarding utilizing text mining to understand consumers. One common concern is that, relying on machine learning, text mining can't reveal the social context consumers live in and can't provide an in-depth understanding of consumers' experience and desire, which are the catalyst of opportunity identification, trend spotting and innovation (Leavy, 2010). Therefore how to bridge this gap to use text mining to transfer big data into business competitiveness becomes a critical issue.

At the same time, online ethnography, termed 'netnography' by Kozinets (1997) as an online qualitative marketing research technique has been invented and developed for providing consumer insight. "Netnography" means ethnography adapted to the online community study. As a method, "netnography" is faster, simpler, and cheaper than traditional ethnography, and more naturalistic and unobtrusive than other qualitative research methods such as focus groups or interviews. It reveals

---

<sup>4</sup> <http://www.jeffbullas.com/2014/01/17/20-social-media-facts-and-statistics-you-should-know-in-2014/>

<sup>5</sup> <http://ivn.us/2013/10/02/7-social-media-stats-2013-will-surprise/>



information on the symbolism, meanings, and consumption patterns of online consumer groups (Kozinets, 2002), which is what text mining approach lacks.

## 1.2 Problem area

Text mining approach could in theory be integrated with netnography to create a synergy effect by generating more holistic consumer insight which could not only demonstrate the general landscape, trends, clusters and connections from a macro perspective but also zoom in to provide more in-depth and context-rich insight in a micro level from online social media. Some commercial market research institute and academic researchers are exploring the possibilities of combining both, however there is a lack of systematic research approach for the process and therefore leaves space for this research.

## 1.3 Purpose

The purpose of this research is to enhance the understanding of the nature of netnography and text mining, explore the potential for the two methods' integration and as an extension develop a research method framework that integrate the two methods and serve as a guideline to generate consumer insights from online social media in commercial market research context.

## 1.4 Research questions

In order to attain the purpose, the research questions below will be explored and answered:

Firstly, in order to understand the compatibility between the methods, a high level analysis of netnography and text mining will be performed focusing on the question: *Based on the strengths and weaknesses of netnography and text mining, what is the suitability for a mixed methods research approach integrating them?*

Secondly, the key participants and typology of the approach will be determined. Due to this the thesis aims to answer: *Who are the key participants and what are the typology of a mixed methods research approach between netnography and text mining?*

Thirdly, to further increase the understanding of the method, the process of the mixed methods research and the information exchange between its components needs to be analyzed. The result of this analysis will be a method model with the description of key elements, attributes and information flow dynamics. The thesis thereby aims to answer: *What is the method model for a mixed methods research approach between netnography and text mining for consumer insight generation from online social media?*

## 1.5 Delimitations

Firstly, this research narrows the analysis to commercial market research usage where the source is online social media. Therefore, this research's result doesn't necessarily directly apply for consumer research for academic purpose. However, please note that the findings can still be applicable as a foundation for future academic research regarding methods for commercial market research usage.

Secondly, this research will demonstrate one way of how text mining can be integrated with netnography. However, it won't cover all the ways of which these two methods combine. Both text mining and netnography are immature techniques that evolve drastically all the time and therefore the research in combining them two should be open for potential changes.

## 1.6 Expected knowledge contribution

The intent is that through this research, practitioners in functionalities such as marketing, communication, customer service and new product development can gain a profound knowledge and a systematic model for how to utilize text mining and netnography to generate consumer insight from online social media. Insights are at the heart of modern corporations and the landscape from which they are generated is becoming continuously more information dense. Through a structured mixed methods research approach between netnography and text mining, the author hopes that insights can be attained with greater precision at a lower cost. The intent has inspired the researcher to combine knowledge from both theory and practice as well as integrate insight from both academia and practitioners to make the finding academically sound and practically applicable. Finally, the research can also serve as a foundation for further mixed methods research with regard to qualitative and quantitative methods integration in a social media context.

## 1.7 Definitions

This section will provide the definitions and explanations for the core concepts of this research.

**Big data:** Regarding the definition of big data, most definitions focus on the gigantic quantity of data in storage. However, Russom (2011) argues that there are two other critical dimensions of big data Size matters, namely data variety and data velocity. "The three Vs of big data (volume, variety, and velocity) constitute a comprehensive definition, and they bust the myth that big data is only about data volume. In addition, each of the three Vs has its own ramifications for analytics"(Philip Russom, 2011). IBM added another dimension "Veracity" to the 3V model, meaning that big data contains uncertainty due to "data inconsistency and incompleteness, ambiguities, latency, deception, model approximations" (IBM, 2012).

**Market research:** Market research, which is often interchanged with marketing research, is defined as “... the function that links the consumer, customer, and public to the marketer through information-information used to identify and define marketing opportunities and problems; generate, refine, and evaluate marketing actions; monitor marketing performance; and improve understanding of marketing as a process<sup>6</sup>” (AMA, 2004).

**Consumer insight:** According to Stone (2004, p.2), consumer insight includes two forms. “Firstly, there are Insights (plural)-flashes of inspiration, or penetrating discoveries that can lead to specific opportunities. Market research or customer databases can deliver these, and often do. However, much bigger than this, and central to what companies need today, is Insight (singular), defined as ‘the ability to perceive clearly or deeply’, a deep, embedded knowledge about our consumers and our markets that helps structure our thinking and decision-making. Everyone involved in marketing needs this form of Insight”.

**Social media:** Ahlqvist, Toni, et al (2008, p. 13) defines social media as “the interaction of people and also to creating, sharing, exchanging and commenting contents in virtual communities and networks”.

**User-generated content:** It refers to “media content created or produced by the general public rather than by paid professionals and primarily distributed on the Internet”(Daugherty, 2008, p. 16).

**Netnography:** “Netnography, or ethnography on the Internet, is a new qualitative research methodology that adapts ethnographic research techniques to the study of cultures and communities emerging through computer-mediated communications” (Kozinets, 2002).

**Text mining:** It belongs to data mining and “refers to the process of extracting interesting and non-trivial patterns or knowledge from text documents” (Tan, 1999).

**Mixed methods research:** It is defined by Burke and Anthony (2004, p. 17) as “... the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study”.

## 1.8 Disposition and overview

**Chapter 2** combines research background introduction, literature review with theoretical overview of the research field to build a theoretical platform for the thesis and provide a holistic comprehension of the research field. The chapter ends with a theoretical research model to serve as

---

<sup>6</sup> <https://www.ama.org/AboutAMA/Pages/Definition-of-Marketing.aspx>

an analytical framework and theoretical foundation for the analysis in chapter 5.

**Chapter 3** presents how the research was designed and conducted. Research strategy wise, a qualitative and abductive approach were utilized. Research design wise, multiple semi-structured interviews were performed.

**Chapter 4** presents the empirical findings from the semi-structured interviews.

**Chapter 5** presents the analysis of the theoretical and empirical findings in the previous chapters. A mixed methods research model is also developed in this chapter.

**Chapter 6** presents the study's main conclusions.

**Chapter 7** provides the final discussions of the thesis, including the managerial implications, reflections of the study and the suggested directions of future research.

## Chapter 2 Research background and theoretical overview

*This chapter contains an overview of literatures and theories relevant to the research question, followed by a theoretical framework that will be the base for the analysis. It is divided by 5 blocks. The last block provides a theoretical research model based on the integration of theories which will be employed to structure and analyze the empirical findings.*

### 2.1 A changing research environment – social media & user generated content is redefining the boundary of market research

#### 2.1.1 Social Media

A study by Nielsen shows that users of Internet spend more time in online social media than any other kind of website in 2012<sup>7</sup>. Experian Marketing Services also did a study which shows that in the United States for every hour online, 16 minutes was spent on social networking and forums, which is roughly 27% of all time spent on PCs<sup>8</sup>. Social media enables its users to connect and interact with a gigantic amount of others easily, quickly and simultaneously and is gaining exponential growth everyday. Social media is different from traditional medias from multiple perspectives, such as quality, reach, frequency, usability, immediacy and permanence (Morgan et al., 2012). Gundecha

---

<sup>7</sup> <http://www.nielsen.com/us/en/newswire/2012/social-media-report-2012-social-media-comes-of-age.html>

<sup>8</sup> <http://marketingland.com/study-27-of-time-online-in-the-us-is-spent-on-social-networking-40269>

and Liu (2012) created a classification scheme with 9 different types of social media and the description of them as shown below.

Type	Characteristics
Online social networking	Online social networks are Web-based services that allow individuals and communities to connect with real-world friends and acquaintances online. Users interact with each other through status updates, comments, media sharing, messages, etc. (e.g., Facebook, Myspace, LinkedIn).
Blogging	A blog is a journal-like website for users, aka bloggers, to contribute textual and multimedia content, arranged in reverse chronological order. Blogs are generally maintained by an individual or by a community (e.g., Huffington Post, Business Insider, Engadget).
Microblogging	Microblogs can be considered same a blogs but with limited content (e.g., Twitter, Tumblr, Plurk).
Wikis	A wiki is a collaborative editing environment that allow multiple users to develop Web pages (e.g., Wikipedia, Wikitravel, Wikihow).
Social news	Social news refers to the sharing and selection of news stories and articles by community of users (e.g., Digg, Slashdot, Reddit).
Social bookmarking	Social bookmarking sites allow users to bookmark Web content for storage, organization, and sharing (e.g., Delicious, StumbleUpon).
Media sharing	Media sharing is an umbrella term that refers to the sharing of variety of media on the Web including video, audio, and photo (e.g., YouTube, Flickr, UstreamTV).
Opinion, reviews, and ratings	The primary function of such sites is to collect and publish user-submitted content in the form of subjective commentary on existing products, services, entertainment, businesses, places, etc. Some of these sites also provide products reviews (e.g., Epinions, Yelp, Cnet).
Answers	These sites provide a platform for users seeking advice, guidance, or knowledge to ask questions. Other users from the community can answer these questions based on previous experiences, personal opinions, or relevent research. Answers are generally judged using ratings and comments (e.g., Yahoo! answers, WikiAnswers).

**Figure 1.** Characteristics of different types of social media. Source: Gundecha and Liu, 2012

## 2.1.2 Social media data

### 2.1.2.1 Social media content

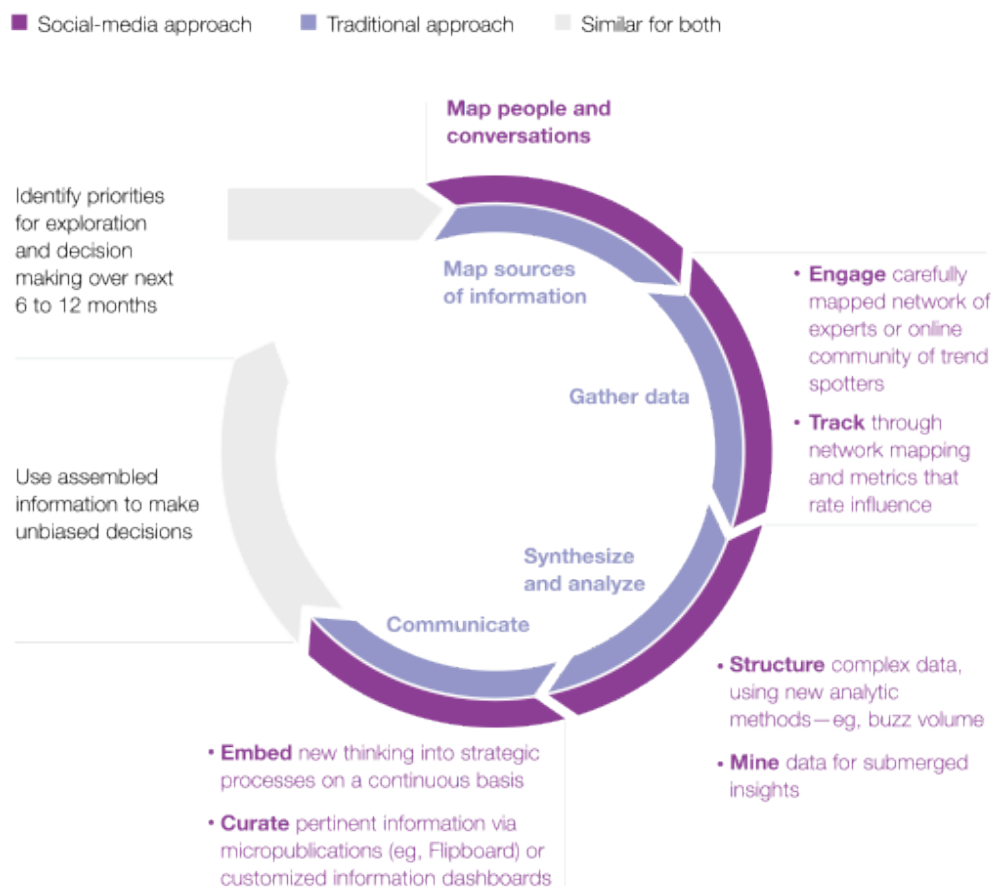
Social media content is the content generated through interactions of users in social media. Social media content contains “different kinds of media contents created and published by amateurs who have just been at the consuming end in the past” (Clever et al., 2008). Social media content is often termed as User Generated Content (UGC). Kaplan and Haenlein (2010) identified social media content as UGC. They also argued that UGC is “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content” (Kaplan and Haenlein, 2010, p. 61).

### 2.1.2.2 Social media data vs. big data

In practice, social media data is often mixed with big data. However, social media data is not equal to big data but rather one type of it. Besides social media data, big data can also come from sources such as “supply chain applications, text data from call center applications, semi-structured data from various business-to-business processes, and geospatial data in logistics” and also “data coming off of any kind of device or sensor, say robotic manufacturing machines, thermometers sensing temperature, microphones listening for movement in a secure area, or video cameras scanning for a specific face in a crowd” (Russom, 2011, p. 7).

### 2.1.3 Social media analytics bring in new opportunities

The rapid growth and ready availability of social media data is changing the way people collect data, analyze data and generate insight and intelligence from data (Cristianini, 2010). In particular, social media data is becoming a crucial channel for companies to create business intelligence. Harrysson et al. (2012) identified the way social media enhances the intelligence cycle.



**Figure 2.** The way social media enhances the intelligence cycle. Source: Harrysson, et al. (2012)

### 2.1.4 Social media data analysis guideline

An overview of the approach for analyzing social media data, one type of text data, is provided by Veeck (2013):

1. **Develop a problem definition and research objectives:** This may be the most important step and includes understanding what the direction should be which is especially important due to the amount of data.
2. **Identify key search terms:** Often an iterative process starting with broader searcher followed by for example combinations.
3. **Identify social media data sources:** This depends on the research objectives and can include for example social network sites, video-sharing sites or blogs.
4. **Organize data:** Important data can in other forms than text and preferably the organization of data should thereby allow for this. It is also important to know when to stop searching for sources due to the large amount of available data on the web.
5. **Analyze data:** The practice here is the same as for other qualitative data.
6. **Present findings:** Here social media can stand out due to e.g. the ability to get quotes.
7. **Outline limitations:** It is as important, if not even more important, to highlight limitations when utilizing social media data compared to other research.
8. **Strategize:** Utilize the findings for actionable recommendations.

## 2.2 Netnography

### 2.2.1 Market research's thick description and netnography

Netnography, also termed digital ethnography or virtual ethnography (Murthy, 2008), is ethnography adapted to the study of online communities. As a method, “netnography is faster, simpler, and less expensive than traditional ethnography, and more naturalistic and unobtrusive than focus groups or interviews. It provides information on the symbolism, meanings, and consumption patterns of consumer groups” (Kozinets, 2002, p. 61).

Kozinets, who created the term “netnography”, is a pioneer in netnography and his netnographical framework is quoted and utilized by multiple researchers. Maclaran et al (2003 p. 153) for example says that “Netnography is a term first coined by Kozinets (1997) to cover the use of online ethnography in consumer research”. Furthermore, Langer & Beckman (2005) utilize Kozinets’ netnography framework to understand virtual communities as well as additional benefits of the approach such as the greater control over the participants.

Although netnography as a method to observe online consumer behavior has been existed for over 10 years, "... the potential of netnographical methods in market research and analysis is still largely undeveloped...compelling insights into consumer ethnography online remains relatively undeveloped by markets, when set against the increasing penetration of the Internet into consumers' daily lives" (Xun et al., 2010). Murthy (2008) argues that a balanced integration of physical and digital ethnography can provide a larger and more exciting array of methods and enables the research to demarginalize respondents' voice. Kozinets(2010) himself also notes that netnography could be combined with other types of research approaches in order to generate further insight from online communities and social medias.

### 2.2.2 Netnography's research process

Kozinets' framework provides a clear structure for further research and reflection. However, to assure that a more holistic perspective of netnography is provided, some additional interpretation from other researchers is needed. Below thereby follows Kozinets (2002) framework for netnography, with some minor additions from other authors clearly marked:

1. **Entrée:** Define the market research question/s and learn about the participants, groups and forums one tries to understand.
2. **Data collection and analysis:** Observations regarding a community and its members, meanings and interactions as well as a direct copy of computer-mediated communications from the online community members.
3. **Providing trustworthy interpretation:** Conventional procedures must be followed to ensure that the research is trustworthy. For example, "... the marketing researcher must be conscious that they are analyzing the content of an online community's communitive acts rather than the complete set of observed acts" (Kozinets, 2002, p. 68). Langer & Beckman (2005) further adds empathies on analysis at this step by calling it analysis and interpretation. In addition Maclaran et. al. (2003) also pushes the importance of building trust towards the community one is studying.
4. **Research ethics:** Kozinets highlights that a key difference between ethnography and netnography is in ethics. For example, is a forum a public site and what is informed consent? Maclaran et. al. (2003) further adds that criticism towards researchers who utilize discussion lists without seeking permission has occurred, further illustrating the importance of this aspect.
5. **Member checks:** In order to solicit the comments of the people who have been studied all final research findings are presented to them.



### 2.2.3 Strengths and weaknesses of netnography

By integrating other researchers' opinions on netnography, Xun et al. (2010) in their research analyzed the strengths and weaknesses of netnography.

**Strengths:** They suggested that netnography is “a combination of the Internet's substantial information-carrying capacity and the removal of individual social cues, means that the researcher is provided with “a goldmine of information”. More specifically: (1) Greater accessibility to a broader cohort of respondents; (2) Greater continuity in research; (3) More economically viable and timesaving than conventional techniques; (4) Greater capacity and flexibility for observation and analysis; (5) The reflective quality of online discourse.

**Weaknesses:** In the same study, Xun et al.(2010) also mentioned several weaknesses of netnography. They propose that there may be “as many uncertainties and challenges as opportunities” for netnography as with other new methods. More specifically: (1) Respondent authenticity and instability of the user base; (2) An underdeveloped analytical toolkit. As a relatively new method, netnography has fewer and less developed analytical tools in the market; (3) Potentially poor quality of textual discourse; (4) Ethical sensitivity.

Besides the weaknesses listed above, Xun et al (2010, p. 22) also point out that “netnography, like ethnography, often tends to focus upon one, or a small number of subjects. The extent to which generalization is possible has been called into question”. However, some researchers argue that it is necessary and important to focus on limited amount of objects in order to develop an in-depth observation and understanding. Kozinets (2006, p. 279-280) defends that “I offer a detailed unpacking of a single early post from a single Star Trek fan to demonstrate the nuanced cultural understanding and interpretive subtlety and depth required for netnography to reveal holistic cultural realities”. Catterall and Maclaran (2002, p.228) also propose that netnographical researches need to limit the research scope to only certain online content because of the huge amount of information needed to be managed.

## 2.3 Text mining

### 2.3.1 Overview

Data mining is “... a process of discovering useful or actionable knowledge in large-scale data” (Tan et al., 2006), which belongs to “Knowledge Discovery from Data” process (Han et al., 2011), a typical approach of generating knowledge from raw data. Key elements of KDD are: data preprocessing, data mining and data post processing (Gundecha and Liu, 2012). Data mining isn't an

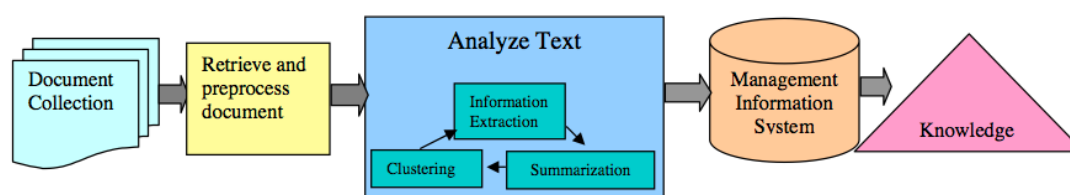
isolated technique but rather an integrated one which relates to different disciplines such as statistics, pattern recognition, visualization, machine learning, database system, association rule mining, feature selection, visual analytics, instance selection and so on (Han et al., 2011; Witten et al., 2011).

Text mining, as one type of data mining techniques, is also known as “text data mining” or “knowledge discovery from textual databases”(Hearst, 1999). It “refers to the process of extracting interesting and non-trivial patterns or knowledge from text documents” (Tan, 1999). While netnography is a qualitative approach to gain insight from online media, text mining is a quantitative approach to extract insight from textual data (Witten, 2005). It aims to generate attributes of freely formed text automatically including word frequency, summary, key features category and so on (Weiss et al., 2010). Text mining is “closely linked to the relatively new fields of natural language processing (NLP) and computational linguistics which develop and implement computer software programs for the purpose of generating, analyzing and manipulating electronically-stored texts” (Ranfagni et al., 2014, p. 9).

Text mining for marketing intelligence can be divided by 3 categories (Glance, 2005): early alerting (informing subscribers if a rare but critical condition occurs), buzz tracking (following trends within discussion and understanding which new topics are evolving) and sentiment mining (extracting and measuring positive vs. negative attitudes). For example, Rickman and Cosenza (2007) adapt weblog-textmining to fashion forecasting; Archak et al. (2007) use text mining to mine customers’ reviews to derive the pricing power of product features; Lee and Bradlow (2012) mine user generated and crowdsourced content to design ranking systems for hotels in travel search engines.

### 2.3.2 Text mining’s research process

Fan, Weiguo, et al (2006) provided a generic process model for text mining. It starts from document collection, and then the text mining tool will retrieve and preprocess the documents by format and character sets check. After that is the analysis of text data to extract information. The result is placed in the management information system and knowledge is created as the output.



**Figure 3.** Text mining generic process model. Source: Fan, Weiguo, et al (2006)

### 2.3.3 Strengths and weaknesses of text mining

JISC (2012) summaries the strengths and weaknesses of text mining in their report. According to their research, the strengths of text mining are: (1) Efficiency: Text mining can drastically shorten time required for research; (2) Unlocking “hidden” information and developing new knowledge; (3) Exploring new horizons; (4) Improved research and evidence base: researchers can utilize broad source of data which “provide a significantly enhanced research resource”(JISC,2012); (5) Improving research process and quality; (6) Broader benefits: They include “cost savings and productivity gains; innovative new service development; and new business models”(JISC,2012).

The weaknesses/challenges of text mining are summarized in the report as: (1) Legal uncertainty, orphaned works and attribution requirements: Before the text mining research is conducted, the permission of the copyright holder is required; (2) Entry costs: The cost of establishing tools and training researchers can be significant; (3) Noise in text mining results: Text mining of documents itself could create errors, false connections could be identified and real connections could be missed; (4) Document formats: Many documents’ format, such as images or charts, could limit the amount of text which could be mined; (5) Information silos and corpora specific solutions: “Corpora of documents or individual orphaned documents for which text mining agreements have not been made must be excluded, leading to inaccessible silos of information and limiting the effectiveness of text mining” (JISC,2012); (6) Lack of transparency: Text mining’s process is perceived by many researchers as a black box where the txt is the input and the new knowledge is the output; (7) Lack of support, infrastructure and technical knowledge: Text mining requires high specialized knowledge and skill, which “creates additional annotated copies of corpora and large information repositories” (JISC,2012); (8) Lack of critical mass: There is a lack of discipline-specific buzz or exemplars.

## 2.4 Mixed methods research

### 2.4.1 Debate between qualitative and quantitative research

The debate between qualitative and quantitative research paradigms has had a long history. The advocates of both sides view their own paradigms as the ideal paradigm for research.

Quantitative purists see their own paradigm in a way that is close to “... a positivist philosophy” (Burke and Anthony, 2004), meaning that social science researchers should treat social phenomena the same way as how natural scientists treat natural phenomena. Also, they propose that the researchers are divided from the entities which are subject to the observation. Furthermore, they

contend that social science should be, and can be, objective, time-free and context-free (Nagel, 1986) and that researchers are able to discover valid and reliable causes of social phenomena.

On the other hand, qualitative purists believe in the advantages of constructivism, hermeneutics, humanism, idealism (Lincoln and Guba, 2011; Schwandt, 2000) and so on. They argue that societal realities are constructed by members in the society and that time-free and context-free generalization is impossible and unnecessary. They also contend that researchers cannot be value-neutral and that causes cannot be totally differentiated from effects (Lincoln and Guba, 2011).

The debate between qualitative and quantitative research leads to the avocation of “incompatibility thesis” (Howe, 1988), which claims that researchers can’t and shouldn’t mix these two paradigms. Guba, a qualitative researcher, claims “accommodation between paradigms is impossible”(Guba, 1990, p.81). Different research orientations and values have been formed through the split of these two paradigms. The qualitative paradigm appreciates an in-depth, rich and observational data while the quantitative paradigm the value of concrete and “hard” data (Burke and Anthony, 2004).

#### **2.4.2 Similarities among different paradigms**

Regardless of the differences, both paradigms also share similarities. For instance, both employ empirical findings to answer research questions and both “describe their data, construct explanatory arguments from their data, and speculate about why the outcomes they observed happened as they did”(Sechrest and Sidani, 1995, p.78). Another similarity is that both paradigms “incorporate safeguards into their inquiries in order to minimize confirmation bias and other source s of invalidity (or lack of trustworthiness) that have the potential to exist in every research study”(Burke and Anthony, 2004, p.15). More fundamentally, all social science researches aim to create warranted statement regarding people and the context they live in (Biesta and Burbules, 2003). Hence, “the objectives, scope, and nature of inquiry are consistent across methods and across paradigms” (Dzurec and Abraham, 1993, p. 75).

#### **2.4.3 Need for mixed methods research**

Today’s research landscape is increasingly complex and dynamic. Consequently, despite the divisiveness of qualitative and quantitative research paradigms, there is a large room and strong need for mixed methods research. More and more researchers require multidisciplinary approach to complement and balance different research methods.

Mixed methods research is defined by Burke and Anthony (2004, p. 17) as “... the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods,

approaches, concepts or language into a single study”. It is the “third research movement, a movement that moves past the paradigm wars by offering a logical and practical alternative”. The logic of inquiry contains induction, deduction and abduction.

By anchoring in the middle ground of qualitative and quantitative research, the mixed methods research can utilize the strengths and reduce the weaknesses of both paradigms (Burke and Anthony, 2004). It can also provide methodologies that are close to the dynamic challenges that researchers face in practice and help the communication and collaboration among different researchers. Howe (1988) contends that although in many cases research processes are linked to specific paradigms, the link is actually neither sacrosanct nor needed. Burke and Anthony (2004, p. 15) propose that “... taking a non-purist or compatibilist or mixed position allows researchers to mix and match design components that offer the best chance of answering their specific research questions”. In the same article, they further demonstrate the advantages of mixed methods research as: “mixed methods research also is an attempt to legitimate the use of multiple approaches in answering research questions, rather than restricting or constraining researchers’ choices (i.e., it rejects dogmatism). It is an expansive and creative form of research, not a limiting form of research. It is inclusive, pluralistic, and complementary, and it suggests that researchers take an eclectic approach to method selection and the thinking about and conduct of research. What is most fundamental is the research question-research methods should follow research questions in a way that offers the best chance to obtain useful answers. Many research questions and combinations of questions are best and most fully answered through mixed research solutions” (Burke and Anthony, 2004, p. 17-18). Also, according to Greene et al (1989), there exist 5 main purposes for conducting a mixed methods research:

Purpose	Explanation
Triangulation	Seeks convergence, corroboration, correspondence of results from different methods
Complementarity	Seeks elaboration, enhancement, illustration, clarification of the results from one method with the results from the other method
Development	Seeks to use the results from one method to help develop or inform the other method, where development is broadly construed to include sampling and implementation, as well as measurement decisions
Initiation	Seeks the discovery of paradox and contradiction, new perspectives of frameworks, the recasting of questions or results from one method with questions or results from the other method
Expansion	Seeks to extend the breadth and range of inquiry by using different methods for different inquiry components

**Figure 4.** Purposes for Mixed Research. Source: Based on Greene, Caracelli and Graham (1989)

#### 2.4.4 Interdisciplinary commercial market research innovation

Regarding the application of mixed methods research, commercial market researches go further than the academia. In the recent decade, there has been a remarkable shift in commercial market research's clients' requirement from generating consumer knowledge to providing insights for both business tactics and strategy navigation (Culkin et al., 1999).

Xun et al. (2010) contend in their study that “as the marketing paradigm shifts from purely product-focused analyses, centering on concepts such as the 4Ps, in an attempt to understand and address the more experiential considerations of consumers, so methodological requirements have also shifted. Traditional normative and quantitative approaches have been complemented by the adoption of more qualitative, ethnographic methods from the social sciences designed to shed light in particular upon consumer experiences by drawing upon postmodern principles”. Accordingly, the clients' emerging demand of consumer insight is motivating market research practitioners to seek for creative and insightful ways to innovate their research methods and toolsets. The need for more innovative research methodologies and theoretical frameworks has been recognized by more and more commercial market research practitioners (Maclaran et al., 2002).

Different disciplines provide a wide selection of market research methods and theoretical frameworks for market research practitioners to choose from. It is becoming more and more common that researchers adopt and integrate different methodologies for the same project (Maclaran et al., 2002) in which way the benefits of different methodologies can complement each other and create synergy effect. This approach, combining different methods for specific research questions, can be described by a French term “Bricoleur” (Maclaran et al., 2002). This term is explained by Levi-Strauss (1996, p. 17) as “... jack of all trades or a kind of do-it-yourself person”. Weinstein and Weinstein (1991, p. 161) also defines it as the person who is “practical and gets the job done” and “works with his (or her) hands and uses devious means compared to those of the craftsman”. Denzin and Lincoln (1994, p.2) mentions that bricoleur is the person who “reads widely and is knowledgeable about the many interpretive paradigms” which could be utilized in a specific research.

Towards the application of “Bricoleur” approach in commercial market research, some academic scholars argue that it is not commensurable across different paradigms (Denzin and Lincoln, 1994) and thereby risks not scientifically sound. However, when it comes to the commercial market research, the “commensurability” is not equally emphasized as in academia. Many commercial market researchers believe that the difference between the principle of adopting theories for

academic purpose and for commercial purpose is quite big (Ereaut, 2002). Rather than limiting their approaches by certain paradigms, commercial market researchers often constructively break the boundaries of paradigms and creatively integrate methodologies from different disciplines in order to acquire fresh, holistic and multi-faceted insight innovatively and pragmatically (Gordon, 1999; Barker et al., 2001).

#### **2.4.5 Factors and guidelines of mixing research methods**

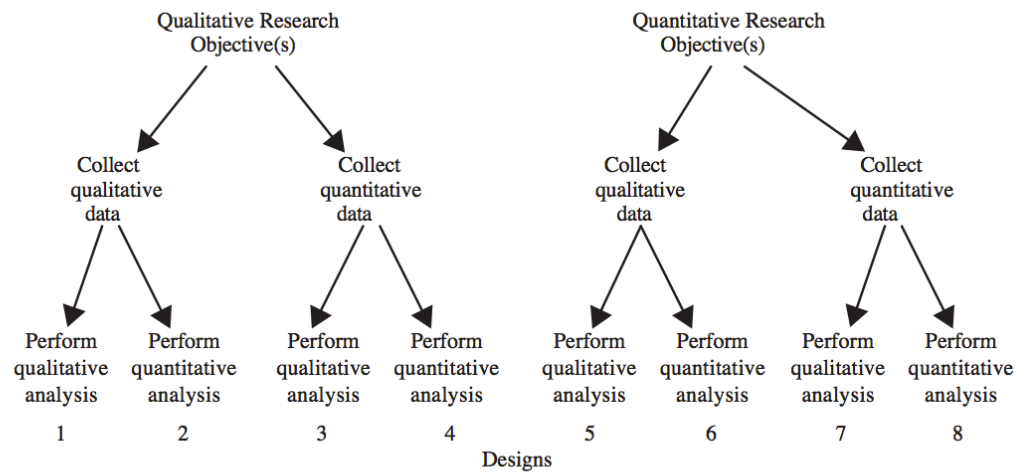
Burke and Anthony (2004, p.18-20) present the factors and guidelines of how researchers mix qualitative and quantitative research methods:

1. Consider all of the relevant characteristics of quantitative and qualitative research. For instance, “the major characteristics of traditional quantitative research are a focus on deduction, confirmation, theory/hypothesis testing, explanation, prediction, standardized data collection, and statistical analysis. The major characteristics of traditional qualitative research are induction, discovery, exploration, theory/ hypothesis generation, the researcher as the primary ‘instrument’ of data collection, and qualitative analysis” (Burke and Anthony, 2004, p.18).
2. Gain an understanding of the strengths and weaknesses of both qualitative and quantitative research paradigms. This understanding “puts a researcher in a position to mix or combine strategies and to use what Johnson and Turner (2003) call the fundamental principle of mixed research”. This principle requires researchers to “collect multiple data using different strategies, approaches, and methods in such a way that the resulting mixture or combination is likely to result in complementary strengths and nonoverlapping weaknesses” (Burke and Anthony, 2004, p.18). This principle can help the justification of mixed methods research.
3. Develop a mixed methods research typology. After considering many other typologies and multiple dimensions which researchers should take into consideration when conducting a mixed methods research, Burke and Anthony (2004, p.20) develop their own mixed methods research typologies. “The majority of mixed methods research designs can be developed from the two major types of mixed methods research: mixed-model (mixing qualitative and quantitative approaches within or across the stages of the research process) and mixed-method (the inclusion of a quantitative phase and a qualitative phase in an overall research study)”.

Mixed-model design: The qualitative and quantitative approaches are mixed either within or across different stages of the process. In within-stage mixed model study, qualitative and quantitative approaches are mixed within one or multiple stages of the research, while in across-stage mixed



model research, the two methods are mixed across at least two stages. Six across-stage mixed model designs are demonstrated below.



**Figure 5.** Monomethod designs (Designs 1 and 8) and mixed-model designs(Design 2, 3, 4, 5, 6 and 7)

Source: Burke and Anthony (2004)

Mixed-method design: To develop a mixed-method design, two decisions must be made: (1) Whether operate largely within one dominant paradigm or not, and (2) Whether conduct the phases sequentially or concurrently. Nine mixed method designs are shown below.

		Time Order Decision	
		Concurrent	Sequential
Paradigm Emphasis Decision	Equal Status	QUAL + QUAN	QUAL → QUAN QUAN → QUAL
	Dominant Status	QUAL + quan QUAN + qual	QUAL → quan qual → QUAN QUAN → qual quan → QUAL

“qual” stands for qualitative and “quan” stands for quantitative. “→”stands for sequential and capital letters stands for high priority or weight and low case letters stands for lower priority or weight.

**Figure 6.** Mixed-method design matrix with mixed-method research designs shown in the four cells

Source: Burke and Anthony (2004)

4. Develop a mixed methods research process model. Burke and Anthony (2004, p.21) proposed a mixed methods research process model which comprises 8 steps. They are: (1) Determine the research question; (2) Determine whether a mixed design is appropriate; (3) Select the typology; (4)



Collect the data; (5) Analyze the data; (6) Interpret the data; (7) Legitimate the data; (8) Draw conclusions (if warranted) and write the final report.

According to the authors, this process model “... incorporates Onwuegbuzie and Teddlie’s (2003) seven-stage conceptualization of the mixed methods data analysis process” which includes: (1) data reduction, (2) data display, (3) data transformation, (4) data correlation, (5) data consolidation, (6) data comparison, and (7) data integration.

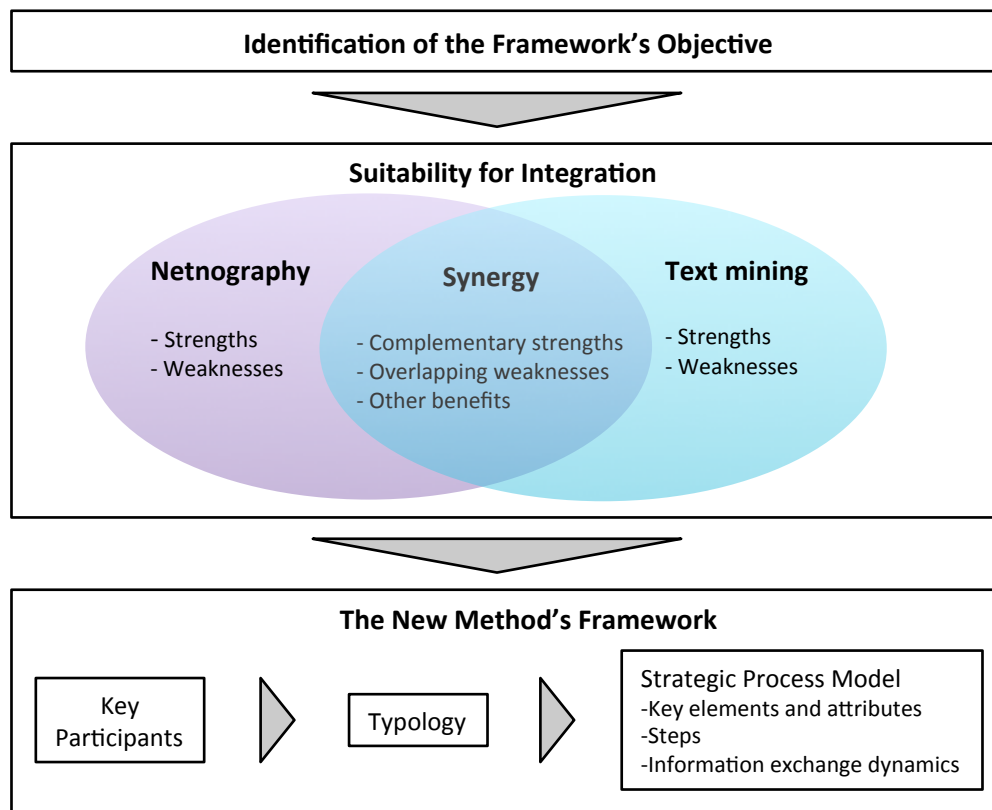
## **2.5 Research model: a synthesis of literature**

Based on the interpretation of literature and theories presented earlier in the chapter, a theoretical research model is constructed (see Figure7). It aims to serve as an analytical framework and theoretical foundation for the analysis part in chapter 5. The research model contains three building blocks: (1) Identification of the framework’s objective; (2) Suitability for integration and (3) The new method’s framework. The rational of this model is:

First of all, to develop a new research method, the author believes that it is critical to understand the method’s purpose which serves as the navigator of the entire method development. Without a clear purpose, one cannot judge and evaluate if the new method meets the expectation. Therefore it naturally becomes the starting point of the entire research model and guides the method development process.

After that, the suitability for integrating the two methods is explored by the analysis of strengths and weaknesses of both methods and the potential synergy created by integrating them. This is in line with the first and second steps of Burke and Anthony (2004, p.20)’s four-step guidelines of how researchers mix qualitative and quantitative research methods as introduced in chapter 2.4.5. The two steps are: (1) Consider all of the relevant characteristics of quantitative and qualitative research and (2) Gain an understanding of the strengths and weaknesses of both qualitative and quantitative research paradigms.

The last building block - the new method’s framework - contains key participant, typology and strategic process model. It is mainly based on the third and fourth steps of Burke and Anthony’s mixed methods research development guidelines: (3) Develop a mixed methods research typology and (4) Develop a mixed methods research process model. However, Burke and Anthony’s guidelines don’t mention key participants of the research method, which the author believes is necessary and important to highlight because in commercial market research the participants could be more diverse and their relationship could be more dynamic when compared to academic research.



**Figure 7.** Research Model

## Chapter 3 Methodology

### 3.1 Research strategy

This section will introduce this study's epistemological and ontological considerations.

#### 3.1.1 Epistemological considerations: hermeneutic

Epistemology is a branch of philosophy exploring the nature and scope of knowledge (Merriam, 1913). It aims to answer what the nature of knowledge is and how it can be obtained, and the extent to which knowledge relevant to any given subject or entity can be obtained. Epistemological approaches can be divided by positivistic approach and hermeneutic approach. The former approach aims to create explanation with a focus of cause and effect while the latter approach targets

understanding with considerations of certain contexts (Andersson, 1979). This study's goal is to create a profound understanding and build a model for the way consumer insight is generated from online social media through utilizing netnography and text mining. Therefore it makes sense to emphasize on "how" instead of exploring the cause and effect or testing certain hypothesis. Therefore the epistemological approach this study employs is a hermeneutic approach.

### **3.1.2 Ontological considerations: constructionism**

Ontology in social science aims to answer "... whether social entities can and should be considered objective entities that have a reality external to social factors, or whether they can and should be considered as social constructions built up from the perceptions and actors of social actors" (Bryman and Bell, 2007, p. 22). Ontology can be categorized by objectivism and constructionism. Objectivism believes that phenomenon and meanings in a society can exist independently from the members of the society, while constructionism believes that social phenomenon and meanings are interlinked with and constructed by the members of the society (Bryman and Bell, 2007, p. 22-23).

This study's objective is to develop a methodology for market research, an object created, constructed and influenced by people, and thus the social construction of reality should be taken into consideration. Therefore, the thesis employs constructionism as the ontological approach which see the world as been constructed by the society's members. It indicates that through changing the practitioners understanding of the research object, one is actually shaping the research object. It also implies that rules and structures of the methodology are not given but constructed and cocreated by different actors (Bryman and Bell, 2007).

### **3.1.3 Abduction**

There are 3 approaches regarding how knowledge is generated. The first is inductive approach, meaning that the knowledge creation starts from the observation of facts in reality. The second is deductive, meaning that the knowledge creation starts from theories. The third is abductive, meaning that the knowledge generation starts from the combination of the existing theories and the empirical reality (Abnor and Bjerke, 2009). Abductive approach believes that theories and empirics are often connected and interact and therefore inductive and deductive approaches are often intertwined in reality and can be combined for knowledge generation.

Based on the analysis above as well as the selection of hermeneutic and constructionist approaches, this research adopts an abductive approach to the generation of theory and knowledge. The research starts from an incomplete set of both observations and explanations. Therefore existing theories and

the observation from the real world function in the “light of each other” (Abnor and Bjerke, 2009, p 92) across the research’s whole journey.

#### **3.1.4 Qualitative approach**

The purpose is to build a new model for consumer insight generation, the descriptive and explorative nature of which decides that it would be very difficult to use a quantitative approach. On one hand, quantitative approach focuses on quantitative and precise relationship between different variables and has problems of capturing the meanings and contexts behind the variables. On the other hand, quantitative approach aims at being objectivist (Bryman and Bell, 2007), which can be problematic when it comes to creating concepts and models.

Qualitative approach, however, is more adequate for this research not only due to its in-depth understanding of the meanings and contexts but also because it allows the research to study actors in natural settings (Bryman and Bell, 2007), providing unexpected or unknown information, knowledge and perspectives which is critical for explorative research like this study.

### **3.2 Research design**

This part introduces and summarizes the design of the research, which is the detailed action to fulfill the research goal and execute the research strategy.

#### **3.2.1 Design of theoretical study**

This section introduces the way theoretical data are acquired and selected.

It begins with the literature collection and review based on certain key concepts that the researcher knew before the study started. Through this step, more key concepts emerged. This step helps the identification and understanding of all the key concepts of the research filed. It is simultaneously conducted with the initial empirical investigation.

The next step is to further explore the key concepts identified from the first step and to gain an in-depth understanding of established theories and knowledge around the key concepts.

The last step of the literature review is to mix, integrate, and synthesize the critical theories to create a theoretical framework for the analysis.

The output of the literature review is the theoretical framework that will be the base for the analysis of the empirical observation.

### 3.2.2 Design of empirical study

Since the study adopt a hermeneutic, constructionist and qualitative research strategy, interview approach for empirical study is naturally employed as the main method for the empirical study. Interview is a more suitable tool in a way that it is “... inherently more flexible [than questionnaires]” (Gillham, 2005, p.3), answers open questions and creates rich knowledge in certain context.

This research studies multiple key actors and their representatives. The reasons for choosing a multiple-actor strategy are twofold. On one hand, it enables the research to acquire a holistic picture and profound knowledge of the research objects. On the other hand, it “... allow the researchers to compare and contrast the findings deriving from each of the cases. This in turn encourages researches to consider what is unique and what is common across cases, and frequently promotes theoretical reflection on the findings” (Bryman and Bell, 2007, p.64).

More specifically, the key actors of consumer insight that have been investigated are:

- (1) Market research practitioners who could provide their practice of integrating netnography and text mining methods in commercial market research.
- (2) Academic scholars in the field of consumer research who could add academic knowledge, rigour and inspiration to the new research method development.
- (3) Consultant with experience in client utilization of consumer insights who could bring insight into the new research method from the consumer insight’s actual user’s perspective.

Semi-structured interviews were conducted for all interviews for the reason that that it enables researchers to cover specific issues but still can “... vary the sequence of the questions... ask further questions in response to ... significant replies” (Bryman & Bell, 2007, p. 213) with interviewees throughout the process of the interview, which allows researchers to extend, follow up and twist the questions and obtain unexpected new knowledge and insight from interviewees, a key contribution to explorative studies.

The reasons why unstructured interview is not chosen are that it is very time-consuming comparing to a semi-structured interview and that it may not cover the key issues that the research aims at focus on. On the other hand, the reason why structured interview is not chosen is that it limits the issues to the researcher’s knowledge scope and therefore is not adequate for explorative questions.

### 3.2.3 Selection of institutions and representatives

This section will provide more information on the specific interview subjects.

#### 3.2.3.1 Market research practitioner representative – Company A

Due to the requirement of confidentiality from the company, its name and the interviewees' names are all pseudonyms. Company A is a consulting and market research company. Their clients range from medium-sized enterprises to public sectors and multinational companies. It has its head office in Stockholm and multiple offices around the globe. Three representatives was interviewed for this research:

Name	Title	Date	Duration
Researcher X	Director, Company A	2014-2-25	107 minutes
Researcher Y	Consultant & Business Anthropologist, Company A	2014-2-10	99 minutes
		2014-2-11	101 minutes
		2014-3-14	55 minutes
Researcher Z	Partner, Director Consumer, Markets & Innovation, Company A	2014-2-12	50 minutes

**Table 1.** Interviewees at a consulting and market research company Company A

Researcher X is a director at Company A. As a big data expert and project manager, he has rich experience and has played leadership role in multiple market research projects with the mixed methods research approach the thesis studies.

Researcher Y is a consultant and business anthropologist at Company A. He has performed multiple netnographical researches together with other researchers with quantitative approaches.

Researcher Z is a partner at Company A with specialization in Consumer, Marketing & Innovation. She has participated in several research projects tha involve the methods of netnography and text mining. She acts as the interface between the research team and the clients.

#### 3.2.3.2 Academic scholar representatives – Stockholm School of Economics and IULM University

To broaden the perspective multiple subjects within academia were interviewed. Two interviewees are professors from Department of Marketing and Strategy at the Stockholm School of Economics, a leading business school in Europe and one interviewee is a Ph.D. researcher at IULM University,

Italy and visiting researcher at Stockholm School of Economics in the field of corporate communication and marketing. The three interviewees were:

Name	Title	Date	Duration
Paola Peretti	Ph.D. Corporate Communication, Major in Marketing at IULM University, Visiting Researcher at Stockholm School of Economics	2014-03-12	30 minutes
Magnus Söderlund	Professor, Department of Marketing and Strategy, Stockholm School of Economics	2014-03-28	60 minutes
Hans Kjellberg	Associate Professor, Department of Marketing and Strategy, Stockholm School of Economics	2014-03-31	36 minutes

**Table 2.** Interviewees from academia

Paola Peretti is a visiting researcher at Stockholm School of Economics. Paola's research field is within customer experience in digital marketing and consumer brand engagement dynamics for luxury brands in mature market. During her PhD period in IULM University, she, together with other researchers, conducted a research combining netnography and text mining which was presented at International Marketing Trends Conference 2014.

Magnus Söderlund is a professor of Department of Marketing and Strategy at Stockholm School of Economics. His research comprises consumer behavior, and his research interest lies in consumers' reactions to marketing stimuli, especially topics such as customer satisfaction, loyalty, emotions, justice perceptions, attitudes, and intentions.

Hans Kjellberg is an associate professor of Department of Marketing and Strategy at the Stockholm School of Economics. His research field focuses on activities that shape individual economic exchanges, activities through which rules and norms for such exchanges are shaped and activities through which such exchanges are represented as markets. His research approach is inspired by sociology and could be characterized as constructivist market studies.

### **3.2.3.3 Insight user representative – Springfellow International**

Springfellow International is a Swedish concept agency that delivers innovation strategies and concepts for international companies. The interviewee from this company is:

Name	Title	Date	Duration
Lars Iversen	Senior Business Advisor and Partner, Springfellow International	2014-03-13	84 minutes

**Table 3.** Interviewees from insight user representative

As a senior business advisor and partner at Springfellow International, Lars has 12 to 15 years experience working with technology, from mobile telephone operations to media and e-commerce platforms as well as ways of utilizing techniques and has during the last 4 years working with disruptive technologies and how big industries need to adapt. He has for example recently provided advice to Electrolux, a Swedish multinational household and professional appliances manufacturer, regarding market-driven innovation decision making for new products, concepts or new product categories. As such Lars is a suitable interview subject to enlarge the understanding for the area and to gain understanding into client perspective on consumer insights.

### **3.2.4 Empirical data collection and analysis**

Interview guides are utilized in all interviews with a simplified version shared with interviewees before the interviews. The questions of the interviews were created based on theories, literatures and questions emerging from the on-going interviews. However, the interview guides are just a guideline for “fairly specific topics” (Bryman and Bell, 2007, p. 484) to cover. In practice the interviews are more flexible and open than what is in the guide. For example, when interviewees showed a strong interest in issues that were not part of the schedule, the author still allowed and sometimes encouraged them to dig in. This enables the creation of new knowledge and insight. Meanwhile, the interview questions have often been updated on an ongoing basis based on the new insight and knowledge gained during previous interviews. This involves changing the structure, fixing the terminologies, polishing the expressions or improving the questions’ intelligibility.

In order to help interviewees keep a good mental status, they were informed before the interviews that:

- They can choose the time and location of the interviews that suited them best.
- They will have the opportunity of fixing the information from the interviews, which will be published publicly in the thesis so that they don’t need to be too cautious.
- They don’t need to limit the conversation inside the exact questions from the interviewer and instead can add new topics, insight and issues that are relevant and important.

Interviews last from 30 min to 100 min long. The working language is English. The entire conversations have been recorded so as to make sure the content can be accessed over and over again. The interviews have also been transcribed into textual documents for further analysis.



Most interviewees were interviewed once, however certain key interviewees received more than one interviews. By the revisiting of the key interviewees, interview problems were corrected and more in-depth information was gathered.

After the transcription of the interviews, the next phase was to analyze the text data. The analysis began by identifying patterns, themes and segments in the data which later was connected, and contrasted against, to each other. Furthermore, the empirics were also analyzed through and contrasted towards existing theory to generate further insight.

Finally, when summaries were established, interviewees received the summary of the interviews respectively through emails and gave feedbacks. This step not only allowed the interviewees to have the chance to correct the preciseness of the content, but also could create the opportunity of potentially gaining more in-depth insight from the interviewees.

### 3.3 Limitations

As a researcher working on this thesis alone, the restriction of limited resource, e.g. time and network, is always an issue the author needs to deal with. There are mainly 2 limitations during the conduction of the study due to the resource issue.

**Input from more insight users:** To acquire the input from consumer insight's user side, the study conducted an interview with a senior advisor who has experience in marketing and new product development in multiple industries. However, the conclusion would be more holistic if more representatives from insight user side could have been involved, such as marketing manager, advertising agency etc.

**Input from consumer:** The research aims to investigate the main actors or "stakeholders" of consumer insight and has covered the insight user side, the research practitioner side and the academia side. However, due to the limitation of the time, the research didn't investigate the consumer side and therefore lacks the insight from this field. If time allows, interviews towards consumers who generate online contents such as bloggers, virtual world players would make the research more holistic and robust.

### 3.4 Research quality

According to Bryman and Bell (2007) there are four ways of judging the quality of quantitative research: Credibility, Transferability, Dependability and Confirmability.

### **3.4.1 Credibility**

Bryman and Bell (2007) highlight that respondent validation, multiple accounts of social reality and that the interview was done in the correct way are important aspects for credibility. For the thesis responds were enabled to validate their interviews, a broad selection of interviewees was selected and interviews were performed in accordance with academic standards for semi-structured interviews.

### **3.4.2 Transferability**

Transferability regards if the qualitative information can be transferred to other milieu where the key factor is to give rich accounts enabling others to make this judgment (Bryman and Bell, 2007). To assure transferability details regarding the interview subjects have been presented in this section. Furthermore, the empirics section is “stand alone”, allowing a greater holistic understanding for each interviewee’s input.

### **3.4.3 Dependability**

The term is similar to “reliability” in a quantitative research context. According to Bryman and Bell (2007) notes and records is to be kept and be available for evaluation to assure dependability. This has been followed by the author and all interviews have been recording in order to increase dependability.

### **3.4.4 Confirmability**

That the researcher acts in good faith and is as objective as possible is the focus of confirmability according to Bryman and Bell (2007). For the thesis interviews were recorded, conducted in a neutral way and encouragement was given only when the subjects initial response to a question had been given. Interpretation also strived for an objective stance when analyzing the material.

## Chapter 4 Empirical findings

*This chapter will present empirical findings from semi-structured interviews, which will be the empirical foundation for the analysis in chapter 5.*

### 4.1 Interview with Researcher X

Researcher X introduces the generic process how a research that mixes netnography and text mining methods is conducted. First, the researchers define the research area and set up the search criteria based on which the data from social media is then collected. After the data collection they conduct multiple specific analysis that combines text mining and netnography. Some examples of the analysis are theme analysis, tribe analysis and timeline analysis. In the end the information is integrated in a way that gives the customer insights that they can act on. Due to the sensitive nature of the information provided, only extracts essential to the study will be presented here.

Researcher X provides his opinion of the advantages of this mixed method:

- *“Unobtrusive. We can do research without our subjects even being aware of it.*
- *We can go back in time and see how things changed over time. This is not the case with surveys and interview, which measure things at a specific point in time.*
- *For high-interest topics, there is a lot of information, sometimes millions of pieces of information, which would not be feasible to collect in other ways*
- *There is so much information and value in the pieces of information that people leave online. This is a way to get access to millions of people’s daily thoughts, views, habits, etc.*
- *It is emergent/bottom-up, meaning that we find the patterns from individual pieces. For example, we can pick up themes and trends that no individual is aware about. If a lot of people start writing about red hats, we see that there is a red hat trend even if no single person is aware that their preference for red hats is part of a bigger pattern.*
- *It is a relatively easy and cheap way to get access to data from a lot of different places. For example, we do a brand scan on 12 markets. This would be very expensive in the form of a survey, but it is much easier to get material and do analysis with blog analysis.”*

On the question of how insight is generated from the integration of netnography and text mining, Researcher X answers that netnography gives the deep understanding of the content from social media but it lacks the overview. Text mining, on the other hand, is *“a way to get an overview so that*

*you know what to focus on and know what topics are out there ... This way, we know what is representative and what is not since we start with the overview and then successively zoom in. If, by contrast, you just pick a random quote, you have no idea where it fits into the overall discussion.”* He believes that the integration of the two methods “give overview and at the same time enables drilling down. When you have these two things, it almost always leads to some useful insights. You know what the forest looks like and if you want to you can zoom in and look at a specific tree. This gives a lot of information about the forest. If you only do qualitative research, you just see a lot of trees but it is hard to know much about the forest.”

Regarding the detail of how text mining can be conducted in this method, Researcher X gives an example. *To analyze and understand a theme in social media content, firstly “we cluster words that are often used by the same bloggers or in the same blog posts. This can be done in different ways ... but the purpose is the same for all of them: to see what words cluster together into themes. This gives insights into what people are talking about within the given area. It gives an overview of the discussion landscape ... end up with a map of the discussion landscape and its topics”. Then, “based on the quantitative results, we select themes that look interesting. We then read individual posts that contain these themes to pick up nuances, illustrative quotes, and to get a better understanding of what the themes are about and how people are thinking related to them. If we would only have done the qualitative part without the quantitative part, we would not know what is relevant. We would just pick up things that seem interesting. The quantitative part gives an overview to start from. But if we only had the quantitative part, we would not get the depth or understand the nuances.”*

When asked how netnography and text mining interact with each other, Researcher X answers: *“Usually we start with quantitative analysis to get an overview. We then do qualitative analysis to get depth and insights. Based on these insights, we then do quantitative analysis to quantify to insights. ”*

Finally, Researcher X also notes that this type of research is often inductive and the knowledge emerges through the research process. *“Since it is not hypothesis driven, we tend to pick up topics that we didn’t know existed. Depending on the purpose of the project, this might add useful insights that can be translated into actions in the customer’s organization”.*

## **4.2 Interview with Researcher Y**

Three interviews were conducted with Researcher Y.

#### 4.2.1 First interview

Researcher Y first clarifies that netnography, if you do it in a proper way, is very scientific. Your toolkit is your theory. You interpret what you see based on theories.

Researcher Y says that Netnography began to emerge with the development of online communities such as second life where people live their lives online. Ethnographers and anthropologists started to think that it would be interesting to study this. When doing so the scientists and informants sit in front of the computers with the limitation of that is that you don't meet your informants and hence have to accept that the informants reveal themselves online and thus *"You could never get the whole context."* However, *"... if you argue that it is artificial online, then what is not artificial? A face-to-face interview is not artificial? What is natural about this? Why won't social media be the truth? It is still a way people express themselves. "*

Researcher Y believes that netnography also often lacks interaction with informants, which is an important part of ethnography because you make judgements and can double check if it is true or biased. Researcher Y continues: *"Without the interaction it is easy for researcher to form a judgement before data gathering and then gather the material accordingly. Ethnography and netnography are not about supporting anything, but about finding something you would never expect. "*

However Researcher Y believes that if you acknowledge these limitations and admit them, it is ok as long as: *"... you don't trick yourself saying that netnography is ethnography and is a representation of what is actually going on. It is wrong. You have to say that netnography is a representation of what is going on online with the limitation of the lack of interaction."*

Researcher Y continues by describing a work experience: *"During one project I worked with, there were 3 researchers involved. Two of us read the blogs, looked at the pictures and used theories to understand what we have seen. And then we have a big data expert who checked all the forums with big data analytics. And sometimes we told that person that 'You have to look for this word. What does this word in connection with this because we see people talking about it a lot.'"*

Researcher Y continues on the topic of big data: *"Big data doesn't seem to be enough to find the unexpected. You still have to have an idea when you do big data. You still have to know what you try to prove. When people talk about big data, they think that you just put a lot of data together and you mine them and then suddenly insight appears. It doesn't work like that. You can't do good big data work without some netnographical work as well because there are so many things in human life that you can't just put into numbers. You can ask the question 'are you male or female' but you can't get*

*a good answer if you only ask the question 'how do you experience as a male or female'. It requires a long answer and analysis, especially in the field you are not familiar with. That is the advantage of netnography which enable people to find something unexpected. I was in a project where I talked with clients about what I have observed by talking with people and the clients said that that was the opposite of what they got from the survey. Then we realized that the survey misses something. Usually people apply netnography when they feel that other methods are not getting anywhere and don't give the result they are hoping. They have the material but can't understand it. "*

According to Researcher Y typical research involving both netnography and data mining works like: *"I together with other researchers look at a lot of materials, go around key words to find what is interesting. Sometimes we talk with the data mining researcher and ask him to analyse certain key words with data mining methods such as word clouds."*

He also describes the contrast between commercial and academic research as: *"Commercial research, comparing to academic research, is usually quicker, more efficient, pragmatic and creative. A lot of academic process is slow because one needs to engage in a theoretical and methodological discussions that are not called for in commercial research...Traditional ethnography calls for the researcher to be submerged in the field between 12-24 months...Therefore the commercial research is an overall quicker process, however, not in any way less qualitative in its results."*

Researcher Y then continues regarding his experience of a mixed methods research approach: *"In research using netnography and big data, it always goes back and forth. That is one difference between netnography and survey because you have so much material that you can draw so many conclusions without "death by survey" due to the wrong question. You can always go back to big data to ask another question. If you do netnography together with survey instead of big data, then you can't go back to the survey to redesign the questions because the survey would have already been finished. But big data can begin to collect data and analyse it the same time as the netnography goes along and interact with it."*

Finally, regarding the output of netnography, Researcher Y says: *"Netnography research's result is very vivid and sometimes can evoke clients' 'aha' moment because they feel they get to know the person. It is very hard to ignore."*

#### **4.2.2 Second interview**

*Regarding the pros and cons of text mining with big data, Researcher Y believes that "Big data can demonstrate the correlation but not the cause and effect. But that is fine. You can combine big data with netnography."*

Regarding representativeness of netnography Researcher Y says: *"... netnography never talks about representativeness. It doesn't claim that by understanding 3 people we understand the whole group. This is not what netnography aims at answering. It tries to research what is hard to be counted such as status and identity. Netnography is not representation. Netnography is articulating human behaviour that can stand as examples or deep insight and meaning. If netnography is why, big data is how and what. If there is an anthropological study by 5 well-trained anthropologists with a clear research question and method, they would most likely get the same result, meaning that it is a misconception to say that it is only one person's interpretation. It is actually a theoretical view that the researcher applies in the field. It is repeatable. You also see changes over time and why the change occurs. Big data also can reveal patterns evolve through time from quantitative approach, so these two tools can reveal the changes through time in two dimensions."*

On comparing netnography to quantitative methodologies: *"... netnography can be more subjective due to its nature and it leaves space for other quantitative approaches to complement. In the best world, if you really want to do a good analysis, you use both especially in a commercial setting. Netnography gives you in a very nice way what questions you want to ask in the quantitative material. That is how big data and netnography should work together. Big data would identify something that netnography can't and netnography would pick it up, try to figure out why and maybe bring back new questions for big data. It is two sides of the same coin."* He also says: *"From my perspective big data is more a support and maybe from big data side, netnography is more a support."*

#### **4.2.3 Third interview**

According to Researcher Y, netnography is about that you need to go back and forth like a pendulum to find knowledge. One needs to be clear that you go back and forth. The tools are defining your research and your research tools. The difference between ethnography and netnography is different analog and digital fields. Big data is about finding patterns in huge amounts of data to help you ask the right questions. A benefit with this tool compared to a survey is that you can go back and *"ask questions as you go along"*.

The interview now moves into various elements of the researches that integrate netnography and text mining such as tribes. According to Researcher Y tribes, and to understand what they talk about in the tribe, are both netnographical and quantitative. Ritual, another element, can help answer the why. People won't say that they are doing a ritual, they do it unconsciously. Ritual is about doing things together in a very structured way. To have a ritual you need artifacts. Rituals, themes and artifacts



can be seen as a sub-theme of tribe. Tribe in turn can be seen as a form of segmentation. Ritual, theme, artifacts, icons, status are examples but there are more such as gender, gatherings and boundaries such as norms, conventions and social rules.

On the question if both text mining analysis and netnographical analysis can study all elements Researcher Y says that gender analysis may be more netnographical, gatherings are more text mining analysis, and norms and boundaries are more netnographical. Artifacts and icons may be more text mining. Symbols are both. Since so much can fit into the framework, choosing the most symbolic such as tribe may be wise for the framework. *“Some things you have to answer with netnography, some things you have to answer with this, but then there are also things that go both ways”*.

### 4.3 Interview with Researcher Z

Researcher Z introduces that Company A begins to combine netnography and big data techniques around 2008 and the main purpose to combine the two methods at that time was that *“Researcher X had this method, and I saw the possibility for that method because I have done this sort of research before so we thought it would be interesting to explore the combination”*.

Researcher Z affirms that there are many commercial research companies that use similar approaches and more and more coming everyday. For example there is a research company in London which uses this method to analyse fashion trend.

Researcher Z believes that Company A’s advantage in this method is that it has deep knowledge in consumer theory. Second, it depends on the people who do the analysis, not just the methodology itself that makes the difference. It is more about finding the right analysts.

When asked if there is any established framework for researchers to follow, she answers: *“I am not sure you should ever have a very formulated process. It depends on what you are doing. If you are following trends, then perhaps the step-by-step process could be useful. But in the strategic projects, it’s much more about reading data. It is a human factor thing. I don’t think you could ever have a computer doing it. It’s very dynamic and intuitive. If you are just to discover new things coming up in certain fields, then you can rely on technology. I am sure you can do that. But if you want to interpret what this means for certain companies, then you need the competence of people.”*

She describes her role in the research as to *“pick up things that’re interesting for companies and to lift them to a strategic level. An interface between the research and the clients”*.

When asked what type of research projects at Company A usually use this methodology, Researcher Z answers: *“Branding projects, innovation projects, to understand how people talk about things, to*



*understand the business environment around certain themes etc. We have used it for many many different projects. ”*

Regarding the feedback from clients who employed this methodology, Researcher Z says:

*“It is still a little bit scary. They don’t know what they will get out of it. The success of the project is much more about using the proper methodology that matches the problem you aim to solve for clients than the methodology itself. When used in the proper situation, it is very successful. Some of the clients think it is a little bit cool to do it because they can show internally that they are working with progressive methods and can find something interesting and so on. That is the advantage. But the drawback is that you also create a lot of uncertainties for clients because you could never guarantee clients what kind of result they would get. You can’t predict the result. That is the scariness for clients. You need to establish a lot of trust with clients for this kind of projects...They don’t understand it. They understand the result but they don’t understand the methodology. You can’t expect them to do that. ”*

When asked if it’s possible to invite clients to participate in the research process to help them understand more and they also can express their needs better, Researcher Z answers: *“It depends on the clients, situation and case. It is about matching clients with situations. You need to find clients that can find that much of time to engage in the research process. You need to have a very good relationship with the clients if you want to do this kind of co-creation and have clients that are really committed and want to put time into the projects.”*

Researcher Z also introduces the advantage of the method: *“It gives you a very cost-efficient market analysis. You can go into several markets and make interesting portrait of consumers. It is a very good tool but not always the best and only tool you use. It depends on the purpose.”*

She also mentions that at the end of the research Company A sometimes holds workshop with clients for idea generation or concept creation. During that workshop, she also uses other materials such as trends material, quantitative material, internal material from clients and so on. She like to mix different methods and material because it is more interesting.

In the end, Researcher Z introduces the process of the research: *“It is a very interactive process. We usually go back and forth a little bit at least. Sometimes when you find something interesting in the material and you will ask your colleagues if they can dig deeper around it a little bit and then they go back again. It happens in every project. It is also my role to point out the issue because I am the interface between Company A and the clients so I can see what is interesting to go back. ”*

## 4.4 Interview with Lars Iversen

The interview with Lars focuses on how clients-the consumer insight's users-handle market research, how they view insights and what role insights play as a part of the organization and innovation process. Also, due to his role as a subject matter expert he was shown an early draft of the framework based on previous interviews towards which he could provide his perspective and insight.

Based on his previous client experiences, Lars believes that insights need to be segmented. You cannot use the same product in Sweden and China. Consumer insight is both ethnographical studies where research is performed externally through various companies, through a fairly new group, as well as quantitative studies bought from third parties, both generic and specifically ordered research. This is then turned into insights. Insights then need to be transformed into benefit areas which then results in mapping out territories where you are relevant and from there you will build a deeper understanding in certain areas. So you need to transform insights into strategy.

Lars believes that comparing to small firms, bigger organizations are enabled to do bigger broader studies, both for the short term and long-term decisions.

Lars does not seem to agree on the negative side of ethnography once the author presents it. Lars says that: *"The combination of both quantitative and qualitative studies will always, you know, be two building blocks, two views of the same view, so if you don't have the other one you will not have the full picture"*

Lars does not think ethnographic studies are mainly to make discoveries, instead they are to verify what you find from competitors or qualitative studies. They verify the hypothesis. Large organizations have already done their homework. So ethnography builds upon basic studies. And the basic studies can be built upon both quantitative, ethnographic, market, competitive and technology studies. So ethnography method is utilized to identify specific issues rather than the general image. His previous client has routine data gathering but sometimes also conducts strategic data gathering. However, all data bought is a part of the strategy.

Regarding the attitude of companies towards new research methods, Lars affirms that large organization can use multiple approaches such as getting highly available data as well as studies where you do surveys, knock on peoples' doors. But new methodologies may not necessarily be appreciated. It would need to provide additional value. Some organizations can be risk reluctant regarding new methodology. So a new methodology needs to be similar to the old ways with a recognizable result.

Lars puts emphases on that the value is in the consumer insight. If you have the wrong consumer insight all parts of the organization will fail, from marketing to how to work with innovation. Consumer insight is the most valuable item in the entire organization, more so than engineers' ability to come up with technological innovations. So the insight and processes to get the consumer insights is critical to prevent the organization from failing. Regarding the role insights play in the innovation process Lars says: *"It is the most important part of the innovation process. You base everything you do on specific consumer insights"*

Lars points at the difference between the insight finding that consumers finding something to be a problem and an insight finding that consumers will pay for having the problem solved. The wrong insight can be worse than no insight. You will have an insight and then check if what you are making solves what the challenge is/goes into the discovered consumer benefit and then go back to for example the focus group. It is a continuous check with the consumer. This can avoid you making a product people do not want to buy. This is not the same as having a product and then trying to find a market to sell it in: *"You will not be stronger than the consumer insight"*.

On the question about if there is time pressure the answer is: *"Of course"*.

On the question if a market methodology could deliver similar value in less time would be interesting: *"Of course"*.

The discussion moves on to online data. Structurally from an internal point of view a company being discussed does not use social media data, but they will buy this data. Online data is treated as signals: *"They would not fit the standards, because you will not be able to put them into the same kind of framework as the traditional data you have, so it would be more of a guideline..."*

But Lars thinks it is still valuable to for example ideation and concept development. And in the later stages of development you do lots of ideation and having signals is crucial.

On the question of what stops it from being the foundation: *"More or less I would say, I think it is it lacks methodology, it would be unstructured ... the results you will get will be important but they will not be methodology wise significant ... valid and significant"*

The client wants to be safe so *"if the results are in a different format they will question that"*.

On the topic of other benefits Lars continues: *"The amount of data is obviously bigger. From a big data perspective it all makes sense. You just need to find a way to find ways of securing the data into formats that are acceptable for large organizations so it fits traditional ways of doing it"*

On the question if the quantitative and netnography methodologies can be integrated Lars believes this is possible and could increase accuracy since the more knowledge you have the more accuracy. However it is new ground and traditionally schooled people may not understand it. However it is an opportunity since it is an industry wide challenge.

On the question of cooperation between client and the organization providing the research Lars responds that they cooperate but it is project based and the client is a part of the project. They need to see the data with their own eyes and ears. So the interaction/cooperation is throughout the research. However, some smaller companies lack the muscle to cooperate but this is not good for the small company. However, if the small company had more resources they would of course cooperate.

At this point of the interview the early draft process model of the method that integrates netnography and text mining developed for the thesis is shown and explained to receive input and perspective from Lars.

Lars' first impression is that the framework will give top-level knowledge but not depth, and that more in depth could be good. Lars considers that instead the framework should be constructed differently in order to segment and in steps reaches additional depth. It has to be an ongoing process where multiple processes narrowing down instead of one large process. That you need a lens/layer and then you have a new layer/lens. So you go deeper, deeper and deeper. And this would utilize the value of online where you have lots of data and can go for the long tail of big data. Lars also believes in focusing on streams. It is like a loop. Identify go further, further identify and then go further. So you generate more and more specific information where you always can go back to the "mothership". What you need is not more and more data, but more and more specific data. It needs to be clearer where it's going, what the purpose is. What people want to buy is the insight and hence the insight should be at the center of the model. Another piece missing, he points out, is the creative part of it made by talented people and the interpretation where cooperation happens with the client.

Lars also mentions the power of big data: instead of generating assumptions from a big group you can actually get actual data. In today's world you can pinpoint reasons and behavior. *"That is the disruptiveness of big data. It's that suddenly you actually have data, but every methodology used in the old world is based on that you do not have data and now that you actually have data you still treat it as you do not have data ... Data is not a product, data is not an end in itself ... The amount of data is not interesting if it does not end up here ... so how do I get insights well that is going from generic data to specific data."*

Lars believes a good consumer insight is fresh, crisp and provides a useful understanding. Having the right data can still result in the wrong insight. What the relationship is between big data and the insight is a key question.

#### 4.5 Interview with Paola Peretti

The research area that Paola applied netnography to is consumer brand engagement dynamics, aiming to understand the dynamics between the brand engagement drivers and how this can influence consumer behaviors. Paola described the reason she employed this mixed methods research as: *“This is a new topic with multidisciplinary approach. It is an explorative study around a topic that is totally unknown. Both me and my supervisor feel the netnography is the only research method possible for this kind of research. One of the main reasons is that you can combine qualitative and quantitative information together and because netnography is the most time-efficient way to get the relevant result. The second main reason is that it is really effective. If you combine the time, the relevant information and the phenomenon you are able to understand, it is perfect, the best and only one. So time-efficiency and the high quality of the information are the two key advantages of netnography. It is the most effective research you can do that balances the cost, time and quality to understand the dynamics of different variables and drivers.”*

When asked her view of “quality” of research, Paola explains as: *“It is about you can combine information in multiple places, you can use it in local level and global level. Especially for luxury brands this is very important.”*

Paola comments on the effect of netnography as: *“I was very skeptical in the beginning. I was not very sure because I am a quantitative researcher. But I quickly realize that you go in-depth into the data and you can learn a lot from different perspectives. I think the rigor of this methodology is good enough if you follow the right process. To do it properly is very important. I use Kozinets’s methodology. I think this is the best one.”*

Despite the benefits of netnography, Paola also sees flaws in the current netnography methodology and created a way to compensate it. *“Comparing to his framework, I added a text mining analysis. That is the only part (I changed). I decided to double check the result so I did text mining and used the software called Tlab. I did a 3D representation of my clusters and found that, although I used two different methodologies, the results were the same. I did it as a test to double check the result.”*

When asked what is the purpose of using text mining, Paola answers: *“That is a good question. When you deal with a lot of content in the netnographical research, it is easy to make mistakes. It depends on the languages. Sometimes I use Italian and French and in these two languages when you*

*take a word outside the context you can easily make mistakes. The software can help double check if everything was correct.”*

When asked if it is because in the classic netnographical research researchers may interpret certain phrases or words wrong, Paola answers: *“Exactly.”*

Paola also shows examples how the research could go wrong. *“It could even be that the sentiment of the content is wrong. Sometimes you have positive sentences in French and Italian but the truth is that it is ironic and negative. You can judge the sentiment when you read between lines but you can make mistakes. So text mining is like a quantitative study with words. It is just to double check. When you do the interpretation, you do it on a personal level and the machine is about putting together the words. In the end the results of both researches are the same. My interpretation is one thing, but number is number. If at the end of the day you don’t see anything completely different from your (netnographical research) result, then you can feel fine that it is good. The combination of netnography and text mining can help you to double check your result. It perhaps is not necessary but I just want to do it to double check.”*

When asked if Tlab helps her to do sentiment analysis to see if the emotion in the text is positive or negative, Paola replies: *“Exactly. I think the text mining is not enough, but the combination with netnography is really good.”*

When asked if there is another area of research where text mining can be combined with netnography besides sentiment analysis, Paola answers: *“So many! I think almost everything. That is the problem. You have to decide based on the time frame, the resource and the research question. Usually one part is sentiment and another is behavior. I didn’t have time because my research was done but consumer behavior is another area where you can combine the two. There are a lot of advanced tools online that you can track data, for example analyzing the consumer journey online. This is the part that netnography is lacking. The combination of netnography and data analytics is completely new. The future is here. If you really want to understand the customer journey, we have to change the traditional approaches of doing research. I am happy to be part of the revolution. There are a lot of opportunities but you have to see the opportunities. Sometimes when you do research you have to be brave. Your work is much harder than other topics but you will learn and achieve something that will give value for future research. That is really good, hard but good.”*

When asked how to guarantee that the information you obtain is representative by using netnography, Paola replies: *“That is the thing. I feel that you can’t completely. That is why I did the text mining. I*

*have done netnographical research for two years and I think this is the reality right now but maybe there will be something coming in the future. It is really changing fast.”*

#### **4.6 Interview with Magnus Söderlund**

When asked what is a good consumer insight, Magnus explains: *“It will capture several things within the frame of the same study. It will capture the causes of something, and the effect of those causes but, if possible, it will also capture the consequences of the effect. So if you can have the causes, the effect and the consequences within the same frame, it would be nice. In addition to this,... the typical study will only capture what is happening inside the mind of consumers. If the study can also capture the real behavioral data, it will be really nice because there is really a call for researchers to move outside the brains of people and to access also what those mental states mean in terms of behavior. So a really interesting study will be able to capture both those psychological reactions and some sort of behavior and match them for the same individual. This is not impossible and it has been done many times. But it is kind of rare anyway.”*

When asked what role consumer insight plays in companies’ decision making, Magnus answers: *“This is debated within marketing circles. I think most people would agree that it is very important to understand what is going on in consumer’s minds when you want to fine-tune your offers when the offers already exist. This is what most people would agree and this is the basis of marketing research for big industries. But some people, I tend to agree with them, think if you really want to surprise the market, come up with something dramatically new, a real innovation, then you most likely rather less ask consumers questions because most people can’t come up with dramatically different ways of doing things.”*

*“Nowadays more and more people are talking about so called ‘big data’. The big data doesn’t necessarily need to provide answers to questions. It can be the observations to people’s behavior. By clever means of analysis, it might be possible to see patterns in those behaviors and from these patterns you may be able to generate new offers to the market. It is possible to capture data from customers which somehow later on could be platforms for various innovation. But it is not necessary that the data needs to be the answer to questions. It might be other things such as putting together large volume of behavioral data to reveal some sort of unsatisfied needs.*

When asked his opinion of ethnographical research, Magnus says that *“Academia has a long tradition of observing people in their natural habitat. Anthropologists and ethnographers always favor that you should stay close to the people you are observing and see what is going on, not*



*necessarily by asking questions but by being there. So it is definitely an academic approach but at the moment other ways of collecting data might be more popular.”*

When asked if this is the kind of methodology innovation accepted by academia or if it is only commercial try out, Magnus says: *“The majority of academia may not prefer that particular method and I’m not sure if it’s intensively used in practice either. Nobody can promise that it reveals the tangible suggestion for solutions. But maybe that could happen and it’s definitely a method worth using.”*

Magnus also presents his view on the main difference when evaluating the marketing research methodology for commercial usage comparing to academic research usage. *“The purpose is different because the practitioners want to sell more and thus methods need to be tuned to it. Academicians wouldn’t necessarily include various profit related variables such as sales, cost or profitability. Academicians would be more interested in things happening early on in the chain of events. Practitioners may not be very concerned about explaining things in terms of theoretical variables and more concerned with what works, how can I improve sales etc. More and more firms say that they’re interested in big data. But they only have behavior data but no psychology behind it. So the real challenge in my mind is to connect these two things for the same individuals within the frame of the same study.”*

When asked if it is necessary to integrate different methodologies, Magnus answers: *“Definitely. Although everyone has a favorite method, everyone knows that every method has some imperfections. It’s just that it takes some time to master one particular method and most people who have invested in one particular method may not have enough time for developing skills for other methods. Usually in academic world researchers use one particular method. But what is coming more and more in academia is that a study should consist of several smaller studies. This is very obvious if you look at academic journals now. Because a typical research will contain study one, two, three and even four done by several people with slightly different questions, allowing researchers to see if different data collections point to the same directions. But what happens also, but more rare, is that they use very different methods for the same study. When they do this and when everything points to the same direction, then your result somehow becomes more convincing given those imperfections in every single method. If you have two very different methods and both have established some sort of patterns, then it is impossible or more difficult to argue that this is just the artifact of the method. Typically in academia it will be like this: they begin with a qualitative study- interview, participant observation or ethnography- and then they begin to see some patterns. The next step is questionnaires, standardized questions towards many people, focusing on some of the things we*



*discovered in the first step. And then the picture will get more detailed. Sometimes it needs one more step, although it is rare, for experiment to isolate causes from effects. When you have all these three side by side and all the results point out that x really seems to be the result of y, then it becomes much more convincing. So far this doesn't happen a lot but it doesn't happen sometimes. I think most researchers would agree that this is a sound practice but they are investing so much time to master one specific method so most people would feel more comfortable with one specific method."*

On the question on how he perceives consumer insight generated from online channels he responds: *"It's definitely interesting. It has something to do with its ability of capturing behaviors. Because now you can collect data on what people did digitally: they click this or that, or did they stay this long, did they buy, when did they leave etc. These are online behaviors that can easily be captured as it occurs, which is kind of interesting because standard questionnaires wouldn't capture this. "*

Magnus then answers the question regarding how he views the consumer information contained in social medias: *"What we think we know is that probably the most important information source for consumers make important buying decisions is their relatives or friends or other people in their social environment. What is apparently becoming more and more important is that you also lie on advices from strangers online who are talking about all sorts of things, rate things or spread digital word of mouth."*

Regarding the pros and cons of text mining technique in terms of consumer insight generation he says: *"Data mining sometimes can reveal correlations that are not expected or unknown. It is an uncontrolled outcome of a process. Most researchers would not like uncontrolled research like this but quite a number of things we have learned from researches occur in that way. You have a lot of data and you begin to explore things. Somehow you find things which you didn't have in your mind. It happens very often in academia even though people don't talk about that. It perhaps is even more likely to happen in practice where you can collect huge databases and have people sitting full time exploring patterns."*

Since one of the main characteristics of big data is that it mainly explores the correlation not the cause and effect inside data, Magnus is asked if he thinks this can be a good foundation for companies' decision making and he says: *"Definitely. One of the foundations for causality is correlation anyway. So we need to look for correlations. We can never be sure that there is a causal relationship but we can be more sure that there might be such a relationship given that the correlation exists. in my mind correlations are far from useless. They are very useful. It's just that it is time-consuming and involves a lot of trial and error just to sit there hoping to find something. This*

*is not typically how academician would work. They would have theories in mind and access these theories and then collect data. It takes some effort to find something in big data if you don't know what you are looking for. But somehow firms seem to be able to find these patterns if they know what they are looking for."*

On the question if the understanding of correlation more for companies' short-term tactic decision making than long-term strategic decisions he says: *"Of course even the retailers not only the academicians would like to understand the causes because it would help discover even more patterns. But if it is not possible to do, you can still capture the correlations. For the short-term tactics such as promotions it is enough. Also, people in the retailing sector have a tendency to be practical, hands-on and would like things to happen now rather than sit and think about the causes of things. "*

He is then asked regarding one of the main critiques about big data which is that it lacks the thickness of information and can't provide rich insight and how he sees this critique: *"That is often true. But several sources are needed. Big data could establish certain patterns and then you can follow up with some other types of methods."*

Regarding the pros and cons of netnography, Magnus says: *"The pros is that you can capture the conversation as it occurs, which would be more or less impossible in real life. This is definitely interesting. Another advantage is that somehow you can decide where you should look for the conversations, not on the Internet in general but in certain places where certain conversations are more likely. The negative aspect is that you don't know who are talking with what kind of role. Is that someone who just likes to talk or is that a real customer? Who is it?"*

Regarding if it is hard to make sure the research is representative: *"Yes. How do you know whether those people active in the Internet are representative of those who happen to not be there?"*

If he thinks this representative issue of netnography can be complemented by big data if the data volume is big enough: *"Definitely it can be complemented. It can be complemented with anything. If you use very different methods and all highlight the same pattern, then this is a very nice position if you want to tell the world that you have found something. "*

Regarding if there is a definite need to combine netnography with something else: *"Yes I think so. More and more theses are now compressing two studies within the framework of the same thesis. So it would be perfectly possible to have part one with netnography and when you find something you can take the next step, e.g. survey, experiment and so on. "*

## 4.7 Interview with Hans Kjellberg

Hans believes that a good consumer insight depends on the context, which can be seen in the following quote: *“There is a radical difference between what is a good consumer insight in established markets, that are already operational, where consumers know the offerings and are skilled at evaluating them, and on the other hand, innovative solutions that require a sort of changed mind-set in consumers. Then a good consumer insight is not necessarily to know what the consumer thinks, but know how the consumer can be changed. So two very different types of insights”*

The type of insight that would suite the innovative market has to be very pragmatic in character, how you can get consumers to change their way of doing. In contrast to more established markets where the cognitive evaluation between different offers instead of behavioural. So the established market insight is more rationalized, something that is not as possible in innovative markets since seeing the results of adopting a certain solution is not as clear.

The rational style does have limitations as well since people are not rational, however it is clearer in innovative markets since the information needed for this rational approach is not even available. Thereby longitudinal processual approaches, following people, is needed. It is unclear how cross-sectional studies can get this: *“We have to do it step by step, we have to follow them, we have to see how they deal with new situations...”*

Hans also makes it clear that there is a gap between intentions and behaviours for consumers. People believe they would act differently from how they actually act. A good consumer insight is able to see beyond what consumers say they will do and thereby credibly say how they actually will behave.

On the question whether netnography fits such new situations Hans responds that: *“I think so, I think that kind of method takes us, potentially takes us, where we want to be to understand a new situation but I think the use of those methods have not perhaps, I mean they have not been exploited to their full extent because I think, there is also a tendency to focus on the symbolic side of consumption in those studies, focusing more on what people say and what people try to communicate through consumption rather than how they actually consume, how they combine novelty with routine...”*

Hans focuses on that it is not about surface but about solving everyday real problems. Ethnography combined with the understanding that people are actually doing things can take us a long way. The decision part has been covered through psychological and organizational research, where the psychological approach is more quantitative in nature. However, sometimes there are not that many decisions made, they were for example made by someone else somewhere else. So decisions should not be given too much weight. It is at times important, but the focus on decision-making makes us

not focus on other things. It is hard to catch the moment consumers make decisions. The view of decisions as a process where you take a decision and then act is not necessarily true: *“It can be that you act and by acting you take a decision”*.

A cross-disciplinary approach can be better at capturing that facet of consumption. E.g. a focus group takes you into a very specific setting which in turn relies on that you are able to communicate what you will actually do. Survey instruments also have the same challenge in capturing the actual act of consumption. *“The obvious downside of ethnography is that you need to do it in real time”*

It is also time consuming and you need to invest in understanding the context well. Also, the outcome is uncertain since you do not know if you will catch a change.

Another approach is the historical approach. So creative use of historical methods, available data, can help us find changes, such as a time series analysis. The problem with historical material is that the quality is not necessarily very good and you may have difficulty of getting to a close level of understanding of the individual consumer since the material may just not be there. This can lead you to accounts that were made for other purposes but may have the information you need. For example study photos of public spaces and see what people carry their items in (consumer logistics). Interesting observations could be made even though the material was not originally intended for this purpose, Hans explains: *“So creative use of historical data that is available but was not necessarily collected for the purpose at hand but says something about consumers and consumer purchases that is one way of doing it”*.

On the question about reliability Hans regards this kind of method as reliable but of course one has to be ready to discuss and reason regarding reliability, but it is likely a rich source of information: *“...present day archaeology or something like that”*

On the question if it is similar to qualitative content analysis Hans believes that in a sense it is, if it is formalized and treated quantitative, but it can be treated qualitative. It is a combination of content analysis and historical approach. Also, you can do this analysis without historical sensitivity by using a single snapshot. It is a historical aspect of ethnography in a sense.

So a researcher can combine methodologies. However there are many rules and some disciplines are more conservative than others. Methodology can affect your ability to be published. For example if a field prefer quantitative methodology then this combined approach may make it difficult to get published. But this does not necessarily hold for more practical applications, like marketing research: *“Then it is usefulness that is the key. There is no given formula for usefulness”*.

On the question if a large data pool in the multi-approach solves the problems Hans indicates yes but it depends on how it is used over time. For example you can generate a better image today using netnography than you could 15 years ago, since people live their lives more in the digital world. Methods able to capture these aspects of our lives thereby become more relevant.

The author raises that a common critique of representativeness of netnography. So another trend is to make netnography more representative through a quantitative scanning approach to see the common pattern, direction, evolvement over time. Question if it is necessary to combine quantitative approach with netnography. Hans does not feel it is necessary since it is about usefulness. If the audience needs more objective information it is useful, then of course the quantitative aspects become important, but then we are back to if a market exists or not. In an existing market we can make estimations but it is much more difficult if we want to change things to point somewhere else.

On the basic idea of combining different methods Hans think it is a good starting point. However the different methods often require different skills and mind-sets so the combination is not necessarily easy. However, this may be more of an academic than business problem.

A weak spot of netnography is that it studies what people say. However, if consuming is to an increasing extent living their lives and consuming digitally then of course these methods become increasingly important to capture what they do and the methods where you follow the consumer in real life become less important. But online is bound to become a partial image since people sooner or later need to be IRL, but you are not just capturing what people say so it is not as bad as capturing simply what they say and you can get a richer than that, but it depends on how you use the material, if you only for example look at forums you will only capture what they say.

## Chapter 5 Analysis

*This chapter will present the analysis of empirical findings in chapter 4. Based on the research model presented in the end of chapter 3, the analysis is conducted to provide a deeper understanding of the nature of netnography and text mining and develop the mixed methods research typology and strategic process model.*

### 5.1 Insight – the objective of market research

Before the exploration and creation of any new research method, it is crucial to understand the ultimate goal of market research. Market research, which is often in line with marketing research, is

defined as “... the function that links the consumer, customer, and public to the marketer through information-information used to identify and define marketing opportunities and problems; generate, refine, and evaluate marketing actions; monitor marketing performance; and improve understanding of marketing as a process” (AMA, 2004). This definition clearly shows the importance of consumer understanding in market research. Culkin et al. (1999) also notes that there has been a remarkable shift in commercial market research’s clients’ requirement from generating knowledge of consumers to providing insights. Furthermore, Lars puts emphases on in his interview that the value is in the consumer insight. He believes that if a company has the wrong consumer insight all parts of the organization will fail and thus consumer insight is the most valuable item in the entire organization. He says that “*[Consumer insight] is the most important part of the innovation process. You base everything you do on specific consumer insights*”. Thereby, based on both literature and empirics the conclusion is that the information collection and the understanding of consumer, or consumer insight, is the objective of market research which the market research method development should target. This objective will always be kept at the core during the following new market research method framework/model development.

## 5.2 Suitability for integrating netnography and text mining methods

As introduced in chapter 2, Burke and Anthony (2004, p. 18-20) presented the factors and guidelines of how researchers mix methods. The early stage is to gain an understanding of the strengths and weaknesses of different research methods. It requires researchers to “... collect multiple data using different strategies, approaches, and methods in such a way that the resulting mixture or combination is likely to result in complementary strengths and nonoverlapping weaknesses” (Burke and Anthony, 2004, p.18). This understanding is also what Johnson and Turner (2003) call the fundamental principle of mixed research. Based on this theory, this section will analyze the potential to combine text mining and netnography.

### 5.2.1 Strengths of netnography

In chapter 2, literature’s view on netnography’s strengths includes: (1) Greater continuity in research; (2) Greater accessibility to a broader cohort of respondents; (3) More economically viable and timesaving than conventional techniques; (4) Greater capacity and flexibility for observation and analysis; (5) The reflective quality of online discourse.

Support for the strengths identified above can also be found through the interviews. For example, regarding the continuity advantage of netnography, Researcher X says that “*We can go back in time and see how things changed over time. This is not the case with surveys and interview ...*”. Magnus

mentions *“The pros are that you can capture the conversation as it occurs, which would be more or less impossible in real life. This is definitely interesting”*. Furthermore, regarding the effectiveness of netnography, Paola believes that *“time-efficiency and the high quality of the information are the two key advantages of netnography. It is the most effective research you can do that balances the cost, time and quality to understand the dynamics of different variables and drivers”*. She comments that *“I was very skeptical in the beginning. ... But I quickly realize that you go in-depth into the data and you can learn a lot from different perspectives”*. Also, regarding the flexibility for observation and analysis, Hans believes that *“The obvious downside of ethnography is that you need to do it in real time”*. This can be avoided by netnography. Researcher X mentions that *“For high-interest topics, there is a lot of information, sometimes millions of pieces of information, which would not be feasible to collect in other ways”*. Magnus notes that *“Another advantage is that somehow you can decide where you should look for the conversations, not on the Internet in general but in certain places where certain conversations are more likely”*. Moreover, regarding the time-efficient and economical strength of netnography, Lars believes that there is time pressure in market research in the innovation division he was working for and that it would of course be interesting if a market methodology could deliver similar value in less time. Researcher Y also believes that the commercial research needs a quicker process than many academic processes and that netnography has advantage of time efficiency comparing to traditional ethnography which calls for the researcher to be submerged in the field between 12-24 months. Furthermore, Researcher X says that *“it is a relatively easy and cheap way to get access to data from a lot of different places”*.

The thickness is a valuable character of insight, which is identified by Hans who believes that a good consumer insight is able to see beyond what consumers say they will do and thereby credibly say how they actually will behave. Besides the strengths listed in chapter 3, the thick and in-depth understanding as well as unexpected finding, which is consistent to the advantage of ethnography, is another advantage mentioned in the interviews. Researcher Y notes that *“You can’t do good big data work without some netnographical work as well because there are so many things in human life that you can’t just put into numbers. ... you can’t get a good answer if you only ask the question how do you experience as a male or female. It requires a long answer and analysis, especially in the field you are not familiar with. That is the advantage of netnography which enable people to find something unexpected.”* He also believes that *“usually people apply netnography when they feel that other methods are not getting anywhere and don’t give the result they are hoping. They have the material but can’t understand it”* and that *“netnography is articulating human behavior that can stand as examples or deep insight and meaning”*. Furthermore, Researcher X believes that *“since it*



*is not hypothesis driven, we tend to pick up topics that we didn't know existed*". Also, an additional strength mentioned by interviewees is the intuitive and easy-to-understand output of netnography. Researcher Y commends that *"netnography research's result is very vivid and sometimes can evoke clients' AHA moment because they feel they get to know the person. It is very hard to ignore"*. Finally, another strength noted by Researcher X is the unobtrusive nature of netnography which can allow researchers to observe the target group in the natural context. He believes that *"We can do research without our subjects even being aware of it"*.

### 5.2.2 Weaknesses of netnography

Literature's view on netnography's weaknesses include: (1) Respondent authenticity and instability of the user base; (2) An underdeveloped analytical toolkit; (3) Potentially poor quality of textual discourse; (4) Ethical sensitivity; (5) Generalization issue

These suggested weaknesses are also supported by the interviews. Regarding the authenticity and instability issue, Researcher Y believes that *"The limitation of that [netnography] is that you don't meet your informants, then you have to accept that the informants reveal themselves online"*. Magnus notes that *"The negative aspect [of netnography] is that you don't know who are talking with what kind of role. Is that someone who just likes to talk or is that a real customer? Who is it?"*. Moreover, regarding the underdeveloped toolkit, Paola says that in the classic netnographical research researchers may interpret certain phrases or words wrong when dealing with a lot of content and therefore needs to be complemented by other tools. Hans also believes that a weak spot of netnography is that it studies what people say instead of what they are actually thinking about. Furthermore, regarding the generalization issue, Researcher Y believes that *"You could never get the whole context"* since *"netnography is a representation of what is going on online with the limitation of the lack of interaction"*. Paola also comments that *"That [representativeness issue of netnography] is the thing. I feel that you can't completely"*. Researcher X also says that *"if we would only have done the qualitative [netnographical] part without the quantitative part, we would not know what is relevant. We would just pick up things that seem interesting"*.

Another weakness that was identified through the empirics is the lack of interaction in netnography. It is not impossible for researchers to interact with the informants in netnography research, but due to the nature of the virtual interaction, it would be often less efficient and effective to conduct interaction. Regarding this issue, Researcher Y says that *"Another limitation of netnography is that it often lacks interaction with the informants, which is very important in ethnography because you make judgment and you need the interaction to double check if it is true or biased. Without the*



*interaction it is easy for researcher to form a judgment before data gathering and then gather the material accordingly.”* Also, comparing to quantitative methods, netnography could be more subjective because it relies on the researcher’s interpretation and therefore can potentially lead to mistake. Paola mentioned that *“Sometimes I use Italian and French and in these two languages when you take a word outside the context you can easily make mistakes”*.

### 5.2.3 Strengths of text mining

The literature’s view on netnography’s strengths include: (1) Efficiency; (2) Unlocking “hidden” information and developing new knowledge; (3) Exploring new horizons; (4) Improved research and evidence base; (5) Improving research process and quality; (6) Broader benefits such as cost saving, productivity etc.

The interviews have also lent support to these strengths. Regarding the unlocking of “hidden” information, development of new knowledge and exploration of new horizons, Researcher X mentions that *“It is emergent/bottom-up, meaning that we find the patterns from individual pieces. For example, we can pick up themes and trends that no individual is aware about”*. Magnus also mentioned that *“By clever means of analysis, it might be possible to see patterns in those behaviors and from these patterns you may be able to generate new offers to the market ... Data mining sometimes can reveal correlations that are not expected or unknown. ... Somehow you find things which you didn’t have in your mind. ... It perhaps is even more likely to happen in practice where you can collect huge databases and have people sitting full time exploring patterns”*. In addition, regarding the cost and time efficiency, Researcher X notes that *“It is a relatively easy and cheap way to get access to data from a lot of different places ... we do a brand scan on 12 markets. This would be very expensive in the form of a survey, but it is much easier to get material and do analysis with blog analysis”*. Furthermore, regarding the improved research and evidence base, process and quality, Paola says that *“I decided to double check the result so I did text mining ... and found that, although I used two different methodologies, the results were the same.”* and that *“My interpretation is one thing, but number is number. If at the end of the day you don’t see anything completely different from your [netnographical research] result, then you can feel fine that it is good”*. Researcher X also believes that *“For high-interest topics, there is a lot of information, sometimes millions of pieces of information, which would not be feasible to collect in other ways”*.

Interpreted from the empirics, another benefit of text mining is that it can give an overview of the research object. Researcher X proposes that *“[text mining] gives an overview of the discussion landscape”* and that *“If we would only have done the qualitative part without the quantitative part,*

*we would not know what is relevant. We would just pick up things that seem interesting. The quantitative part gives an overview to start from".* In addition, due to the time efficiency, Text mining's advantage in revealing correlation instead of causality can also be seen as strength, especially in terms of commercial research practice for the purpose of short-term tactics decision-making. Magnus says *"One of the foundations for causality is correlation anyway. So we need to look for correlations. We can never be sure that there is a causal relationship but we can be more sure that there might be such a relationship given that the correlation exists"* and that *"For the short-term tactics such as promotions [correlation] is enough. Also, people in the retailing sector have a tendency to be practical, hands-on and would like things to happen now rather than sit and think about the causes of things"*.

#### 5.2.4 Weaknesses of text mining

Weaknesses of text mining are presented in the literature as: (1) Legal uncertainty, orphaned works and attribution requirements; (2) Entry costs; (3) Noise in text mining results (4) Document formats; (5) Information silos and corpora specific solutions; (6) Lack of transparency; (7) Lack of support, infrastructure and technical knowledge; (8) Lack of critical mass.

They are strengthened by the findings from the interviews. For example, regarding the legal uncertainty, Researcher X says that *"privacy is a very important topic. We only analyze publicly available information that the individual has chosen to publish online. We do not, for example, download private Facebook information"*. Furthermore, regarding the noise in text mining results, Researcher Y believes that text mining's result is not as clear as many people think: *"When people talk about big data, they think that you just put a lot of data together and you mine them and then suddenly insight appears. It doesn't work like that"*. Magnus also mentions that *"It takes some effort to find something in big data if you don't know what you are looking for"*.

Another weakness of text mining that hasn't been mentioned in the literature is the uncontrolled outcome. Magnus says that *"It is an uncontrolled outcome of a process"*. Researcher Y believes that *"Big data [text mining] doesn't seem to be enough to find the unexpected. You still have to have an idea when you do big data. You still have to know what you try to prove"*. Also, the lack of thick understanding of research object is another weakness that text mining suffers. Magnus believes that one of the main critiques about big data/ text mining is that it lacks the thickness of information and can't provide rich insight: *"More and more firms say that they are interested in big data. But they only have behavior data but no psychology behind it. So the real challenge in my mind is to connect these two things for the same individuals within the frame of the same study"*.

### 5.2.5 Synergy created by integrating netnography and text mining

The table below aggregates the strengths and weaknesses identified from the previous section. This is what Johnson and Turner (2003) call the fundamental principle of mixed research.

	Netnography	Text mining
Strengths	<ul style="list-style-type: none"> <li>• Greater continuity in research</li> <li>• Greater accessibility to a broader cohort of respondents</li> <li>• More economically viable and time-saving than conventional techniques</li> <li>• Greater capacity and flexibility for observation and analysis</li> <li>• The reflective quality of online discourse</li> </ul>	<ul style="list-style-type: none"> <li>• Cost and time efficiency</li> <li>• Exploring new horizons, unlocking “hidden” information and developing new knowledge</li> <li>• Improved research and evidence base, process and quality</li> <li>• Overview creation of the research object</li> <li>• Demonstration of a correlation of factors</li> </ul>
Weaknesses	<ul style="list-style-type: none"> <li>• Respondent authenticity and instability of the user base</li> <li>• An underdeveloped analytical toolkit</li> <li>• Potentially poor quality of textual discourse</li> <li>• Ethical sensitivity</li> <li>• Generalization issue</li> <li>• Lack of interaction with research objects</li> <li>• Potential mistake caused by subjective interpretation</li> </ul>	<ul style="list-style-type: none"> <li>• Lack of thick understanding</li> <li>• Uncontrolled outcome</li> <li>• Legal uncertainty, orphaned works and attribution requirements</li> <li>• High entry costs</li> <li>• Noise in text mining results</li> <li>• Document formats</li> <li>• Information silos and corpora specific solutions.</li> <li>• Lack of transparency</li> <li>• Lack of support, infrastructure and technical knowledge</li> <li>• Lack of critical mass</li> </ul>

**Table 4.** Strengths and weaknesses of netnography and text mining

After understanding the strengths and weaknesses of the two methods, the next step, according to Burke and Anthony (2004), is to analyze if there are complementary strengths and overlapping weaknesses. In table 4, the text marked green represents the weaknesses that can be compensated by the other method while the text marked red represents the overlapping weaknesses of the two methods.

#### 5.2.5.1 Complementary strengths

Through carefully comparing and matching the strengths and weaknesses of both methods, one can observe the opportunity for weaknesses of one method to be compensated by the strengths of the other method if the two methods are combined.

##### 1. Compensation of weaknesses of netnography

**Respondent authenticity and instability of the user base and potentially poor quality of textual discourse:** This issue cannot be completely avoided due to the nature of online data, but it can partly be compensated by text mining because text mining is to analyze the whole landscape of the

database and to create an overview of the research object, meaning that the result as a whole won't be fundamentally influenced by textual discourse with poor quality.

**Underdeveloped analytical toolkit:** As a well-developed powerful analytical tool, text mining can enrich netnography's analytics. Especially if a mixed method framework could be established to guide the process and strengthen netnography's analytical power.

**Generalization issue:** Comparing to netnography's sampling approach of data selection, text mining usually processes the whole text data set and thus, especially when applied to large data sets, has a strong advantage for the generalization issue that netnography suffers. This is supported by Researcher X: *"[text mining is] a way to get an overview so that you know what to focus on and know what topics are out there ... This way, we know what is representative and what is not since we start with the overview and then successively zoom in. If, by contrast, you just pick a random quote, you have no idea where it fits into the overall discussion."*

**Potential mistake caused by subjective interpretation:** Thanks to the nature of quantitative approach, text mining relies less on subjective interpretation than netnography, which reduces the potential mistake brought by it. That is why Paola says that *"When you deal with a lot of content in the netnographical research, it is easy to make mistakes"* and that *"I decided to double check the result so I did text mining"*. Researcher Y also affirms that comparing to quantitative methodologies, netnography can be more subjective due to its nature and it leaves space for other quantitative approaches to complement.

## **2. Compensation of weaknesses of text mining**

**Lack of thick understanding:** The thick understanding is a core advantage of netnography comparing to quantitative methods and therefore netnography can create deeper insight and explanations to compensate text mining. This is verified by Researcher Y who believes that *"You can't do good big data work without some netnographical work as well because there are so many things in human life that you can't just put into numbers"*.

**Uncontrolled outcome:** Although netnographical research doesn't start with an hypothesis and structured questions and therefore the result is also open, the researchers will likely get useful insight of the informants from the netnographical study and therefore the outcome is more controllable than text mining which may risk no finding informative insight.

**Noise in text mining results:** Thanks to the researchers' intervention, netnography can compensate this weakness by interpreting the result, identifying the noise and selecting useful information from the noise.

**Document formats:** Non-textual information, such as video, picture and sound, can be barriers for text mining and leads to ineffectiveness. Netnography can help overcome the obstacle because it is conducted by the researchers' interpretation and hence can unlock insight from non-textual information.

#### **5.2.5.2 Overlapping weaknesses:**

Two overlapping weaknesses between netnography and text mining were identified. They are:

**Ethical sensitivity and legal uncertainty:** Weakness shared by netnography and text mining since both involve data from online content which may lead to ethical and even legal issue. It affects any research method utilizing online content, especially social media data.

**Lack of interaction with research objects:** For text mining, the lack of interaction is very clear and for netnography there is a debate of passive vs. active observation of participants (Murthy, 2008) but the main stream seems agree on the passive and unobtrusive observation. For example, Kozinets (2002, p.65) believes that "the uniquely unobtrusive nature of the method is the source of much of its attractiveness and its contentiousness".

#### **5.2.5.3 Other benefits**

Besides the considerable compensation potential and limited overlapping weaknesses, there are additional synergies created through the integration:

**The shared database:** The data collected for one method can be utilized for the other and therefore is cost and time-efficient.

**Inter-verification:** Netnography and text mining can work as a "double check" for each other. This is supported both by Paola who says that "*I did [text mining] as a test to double check the result [of netnography]*" and by Magnus who believes that "*When they do this and when everything points to the same direction, then your result somehow becomes more convincing given those imperfections in every single method. If you have two very different methods and both have established some sort of patterns, then it is impossible or more difficult to argue that this is just the artifact of the method*".

**Multi-faceted views to demonstrate the full picture:** This is what Lars describes: "*The combination of both quantitative and qualitative studies will always, you know, be two building blocks, two view of the same view, so if you don't have the other one you will not have the full*

*picture*". Researcher X also holds a similar opinion: "[netnography and text mining] give overview and at the same time enables drilling down. When you have these two things, it almost always leads to some useful insights. You know what the forest looks like and if you want to you can zoom in and look at a specific tree".

#### **5.2.5.4 Conclusion: a clear suitability for integration**

To summarize, there is an obvious compensation and synergy created by integrating the two methods, and the overlapping weaknesses of the two methods, on the other hand, are quite limited. Therefore there is a clear suitability for integration of netnography and text mining in order to generate consumer insight.

### **5.3 Development of a mixed methods research framework**

The previous section has analyzed the suitability of integration the two methods. The next step is therefore to analyze how they can be integrated. Based on existing theory, empirical findings and logical deduction, this section aims to propose a framework for a mixed methods research approach that integrates netnography and text mining, including the identification of key participants, method typology, key blocks of the framework, main process and interaction between netnography and text mining. To simplify the expression, the new method that integrates netnography and text mining is shortened into **NTMI** (Netnography and Text Mining Integration) method.

#### **5.3.1 Key participants**

There are 3 key participants in this research framework.

**Researcher:** Either working alone or as a team, the researcher(s) needs to have the knowledge and ability of conducting both netnographical research and text mining research. Due to the disciplinary gap between the two methods, it is more common for researchers with different disciplines to corporate on the mixed methods research. An example is the Company A's research where multiple researchers with different backgrounds work together on the projects involving netnography and text mining.

**Insight user:** It represents the final user of the output of the research. In academia it is often the same group as the researchers, but in commercial market research, the insight users are often constituted by parties outside the research group such as companies, institutions or the public.

**Netizens:** It represents the object of the research: online users who generate content through online social media.

## 5.3.2 Typology

### 5.3.2.1 Mixed-model or mixed-method design

Burke and Anthony (2004) propose that the majority of mixed methods research designs can be developed from two main types of mixed methods research: mixed-model design and mixed-method design. This sector aims to develop the new method's typology. Mixed-model design mixes qualitative and quantitative approaches within or across the stages of the research process, while mixed-method design includes a quantitative phase and a qualitative phase in an overall research. Based on the two designs' definition, theory and empirical findings, it's clear that NTMI method belongs to the mixed-model design. The reasons are twofold:

Firstly, netnography and text mining share the same data source, which is a typical mixed-model design character, while in mixed-method design different methods often have different database respectively. For instance, Researcher X and Paola said that they use netnography and text mining to analyze the same data source rather than collecting different data for different methods.

Secondly, netnography and text mining interact throughout the research process, which is another character of mixed-model design. Mixed-method design usually doesn't contain the interaction during the research process. The interacting nature of integrating netnography and text mining methods can be observed through interviews with Researcher X, Researcher Y and Paola, which will be demonstrated in chapter 5.3.3.3.

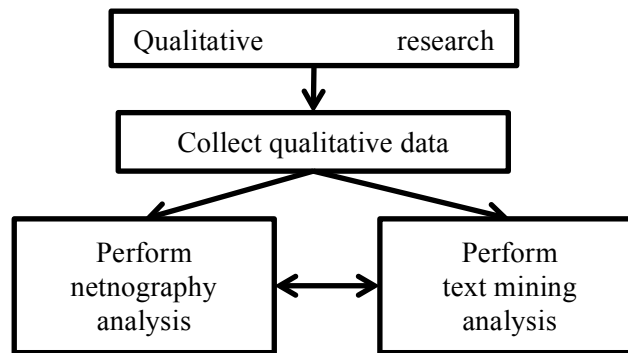
### 5.3.2.2 Within or across the stages of the research

Mixed-model design includes both within-stage mixed-model and across-stage mixed-model. In within-stage mixed-model design, the qualitative and quantitative approaches are mixed within one or more stages of the research. In across-stage mixed-model design, qualitative and quantitative approaches are mixed across one or more stages of the research. The NTMI method is both within and across the stages of the research:

**Across the stages:** The method begins with the identification of qualitative research objectives-consumer insight-and then collect qualitative data-content generated through online social media. The analysis phase after the data collection is constituted by both netnographical analysis and text mining analysis.

**Within the stages:** The netnographical and text mining analysis are not isolated. Instead they involve a lot of cooperation and interactions within the stages along the entire process.

The within-and-across-stage model of the method is demonstrated through the figure below.



**Figure 8.** Within-and-across-stage model illustration

### 5.3.3 Strategic process model

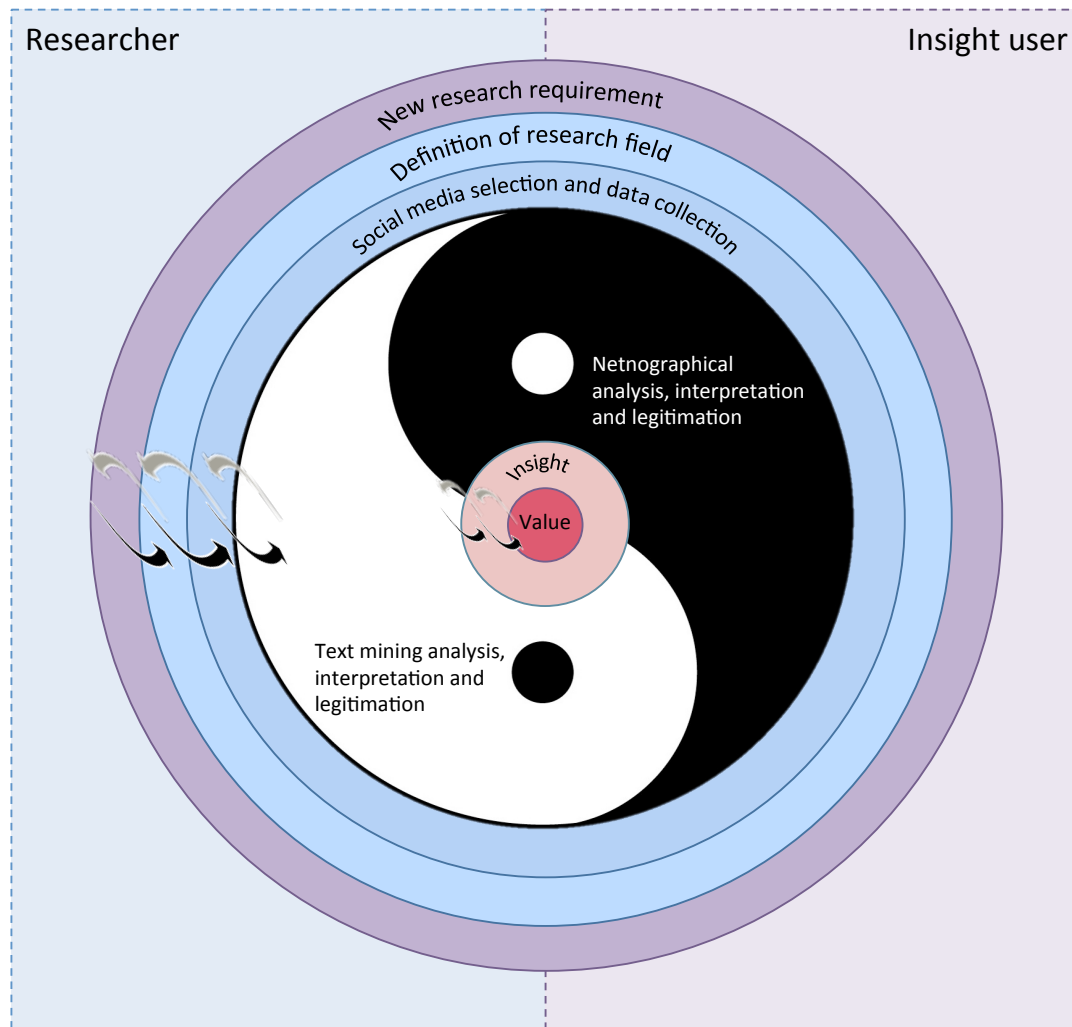
The previous section presents the abstract structure of NTMI method through the identification of its typology. This section aims to build a strategic process model for NTMI method from a more executive perspective. However, note that this model serves more as a guideline than a rigid paradigm. The reason is well explained by Researcher Z: *“I am not sure you should ever have a very formulated process. It depends on what you are doing. If you are following trends, then perhaps the step-by-step process could be useful. But in the strategic projects, it’s much more about reading data. It is a human factor thing. I don’t think you could ever have a computer doing it. It’s very dynamic and intuitive”*.

The model is structured as a funnel with multiple layers. To demonstrate the full picture of the funnel, it will be described from both the cross section angle and the vertical section angle. The key elements and attributes of this model will be analyzed in both sections, which will then be followed by an analysis of the model’s information exchange dynamics. Finally, please note that for pedagogical reasons the model will be shown before the analysis that lead to it.

#### 5.3.3.1 From the cross section angle

From the view of cross section, the model looks like the figure shown below.





**Figure 9.** Cross section illustration of the model

The key attributes of the model that can be shown in this cross section are:

**1. Cocreation between researcher and insight user:** From the cross section view, the model is structured in a research box shared by researcher and insight user, meaning that in an ideal scenario the two parties cooperate and cocreate the research through the research process. This cocreation will be further analyzed in chapter 5.3.3.3.

**2. Value-centered model:** Inside the research box looks like an “onion” with multiple layers from the surface to the core. The reason for choosing onion model instead of the more common linear or circular process diagram is that both linear and circular process diagrams are process-driven or insight-driven. However, the market research requires the research model to be value-driven. For example, Lars believes that *“Insights then need to be transformed into benefit areas which then results in mapping out territories where you are relevant and from there you will build a deeper understanding in certain areas. So you need to transform insights into strategy”*. Although insight is the core goal of the research process, the core goal of the insight is value. As a result the core goal of the framework is

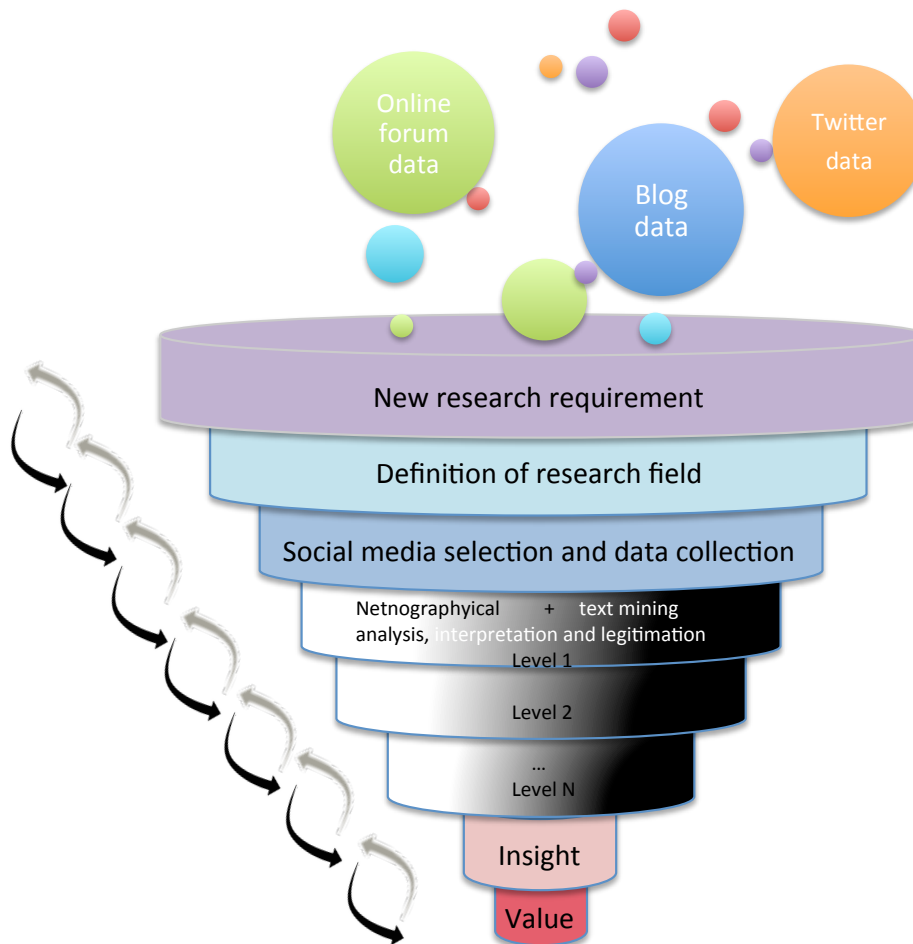
an insight which results in value. Therefore, the value is set at the core of the model with all other steps and elements targeting and going towards it.

**3. Research process:** With Burke and Anthony (2004)'s Mixed Methods Research Process Model as the foundation, the proposed research process in NTMI method model also integrates the process of netnography (Kozinets, 2002) and text mining (Fan, et al., 2006). Burke and Anthony (2004)'s Mixed Methods Research Process Model is constituted by 7 steps: (1) determine the research question; (2) determine whether a mixed design is appropriate; (3) select the mixed- method or mixed-model research design; (4) collect the data; (5) analyze the data; (6) interpret the data; (7) legitimate the data; and (8) draw conclusions (if warranted) and write the final report.

However, in commercial market research, a new research is often initiated by the new research requirement from the client side. Also, it often starts with a research direction and field instead of a clear question. Therefore, the NTMI method model starts from new research requirement and definition of research field. Both netnographical research and text mining research begin with the selection of online communities and data collection which can be shared between the two methods because they use the same data source. Therefore the next step after the definition of research field is the selection of social medias and data collection. When the data has been collected, the netnographical and text mining analysis will be performed with a continuous interaction between the two methods which will be introduced more in chapter 5.3.3.3. Data integration and legitimation characterize the next stage, where the netnographical and text mining analysis are integrated into a whole and the trustworthiness of the data and interpretations are assessed. Then insight will then emerge which will, in an ideal scenario, be turned into value. The end of the research process might be the start of another round's research.

### **5.3.3.2 From the vertical section**

From the view of vertical section, the model looks like the figure shown below.



**Figure 10.** Vertical illustration of the model

In the vertical section, the core elements and processes are consistent with the cross section. However, this angle demonstrates two more attributes of NTMI method model.

**1. Narrowed focus and filtered information:** As the process goes forward, the focus gets narrower and the information filter is more precise. This is in order to arrive at a value adding insight from a large collection of data. This is supported by Lars who believes that a market research has to be an ongoing process where multiple processes narrowing down instead of one large process and the research going deeper, which would utilize the value of online where you have lots of data and can go for the long tail of big data. Hans also believes that *“We have to do [market research] step by step, we have to follow [consumers], we have to see how they deal with new situations...”*

**2. Back and forth interaction:** Netnography and text mining interact back and forth across multiple stages. This attribute will be described in detail in the chapter 5.3.3.3.

### 5.3.3.3 Information exchange dynamics

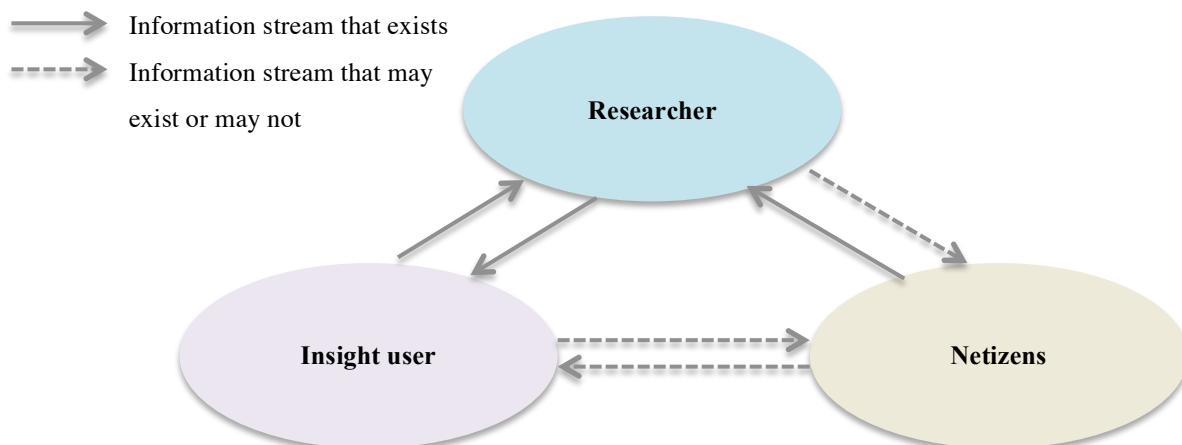
#### 1. Information exchange among three parties (researcher, insight user and netizens)

**(1) Interaction between researcher and insight user:** This interaction is bidirectional with information flow coming from and going towards each side. It would be very beneficial if the user and the researcher could cooperate not only at the beginning and the end of the project but communicate and “co-create” during the whole journey of the research to make the result close to the need and expectation of the user as much as possible through different stages of the research. For example, Lars affirms that the company’s innovation department assigns colleagues to research agencies to work together on the consumer research projects because they need to see the data with their own eyes and ears. So the interaction/cooperation is throughout the research. Researcher Z also says that this cooperation sometimes happens and is beneficial. However, the cooperation is not always the case. Researcher Z mentions that it also depends on the clients’ will, time and dedication. Lars also believes that some smaller companies lack the muscle to cooperate but this is not good for them.

**(2) Interaction between researcher and netizens:** This interaction can be unidirectional with information from netizens to the researcher when the researcher only conducts passive observation without intervention. It can also be bidirectional when the researcher interacts with the netizens, which is less common in netnography.

**(3) Interaction between insight user and netizens:** This interaction can be bidirectional when the insight user participants in the research’s process and interact with the netizens, unidirectional when the insight user participants but doesn’t interact, or nonexistent when the insight user doesn’t participant at all.

The interaction dynamics can be demonstrated through the model below.



**Figure 11.** The interaction dynamics between the three parties

## 2. Information exchange among different layers

Burke and Anthony (2004, p.21) affirm that one of the mixed methods research model's attributes is "several arrows leading from later steps to earlier steps indicating that mixed research involves a cyclical, recursive, and interactional process". In consistency with this logic, although the NTMI method model has a clear process, different stages of the process are not always linear or unidirectional. Except the new research requirement step, the rest of the stages could vary in different orders. The information flow through the research process follows a "cyclical, recursive and interactional" pattern. The arrows in both figure 9 and figure 10 represent the information flow pattern.

Empirical findings also support this attribute. For example, Lars believes that the market research process is like a loop with continuous identification while going further. It should generate more and more specific information where you always can go back to the "mother ship".

## 3. Information exchange between netnographical and text mining analysis

The core attributes of the information exchange between netnography and text mining in NTMI method were found to be:

**Recursive and dynamic interaction:** As analyzed in chapter 5.3.2.2, netnography and text mining hold "dialogues" throughout the analysis. The analysis and interpretation can begin from netnography, text mining or simultaneously and go back and forth between the two methods. Regarding the "back and forth" information flow, Researcher Z mentions that "[NTMI research] is a very interactive process. We usually go back and forth a little bit at least. Sometimes when you find something interesting in the material and you will ask your colleagues if they can dig deeper around it a little bit and then they go back again. It happens in every project". Researcher Y confirms that "In research using netnography and big data, it always go back and forth". Regarding the sequence of the two methods, Researcher X says that "Usually we start with quantitative analysis [text mining] to get an overview. We then do qualitative analysis [netnography] to get depth and insights. Based on these insights, we then do quantitative analysis [text mining] to quantify to insights". However, Researcher Y believes that "big data can begin to collect data and analyze it the same time as the netnography goes along and interact with it". Paola's approach, on the other hand, is to first conduct netnographical research and then use text mining to verify it. Thus the interaction is very dynamic with no fixed pattern.

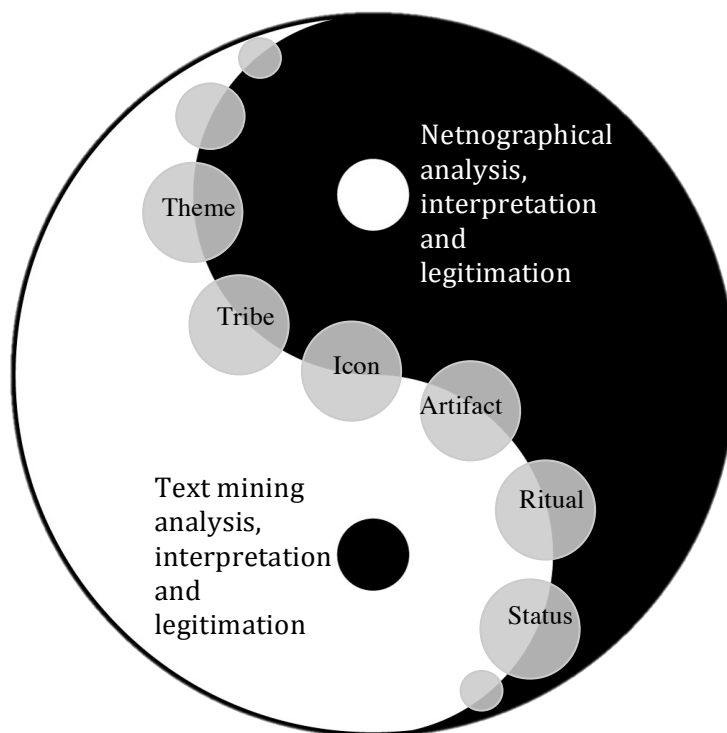
**Interaction through elements:** Netnography and text mining's interaction is embodied in elements such as theme, tribe, icon, artifact, ritual, status etc. Some elements lie more on one method and some lie more on the other. This is supported by Researcher Y: *"Ritual, theme, artifacts, icons, status are examples but there are more such as gender, gatherings and boundaries..... gender may be more netnography, gatherings are more content analysis, norms and boundaries are more netnography. Norms and boundaries may be to abstract for content analysis, but potentially there may be tools. Artifacts and icons may be more content analysis. Symbols are both"*. Researcher X also describes an example how netnography and text mining interact through theme analysis: *"We cluster words that are often used by the same bloggers or in the same blog posts. This can be done in different ways (factor analysis, cluster analysis, structured word cloud) ... end up with a map of the discussion landscape and its topics ... Based on the quantitative results, we select themes that look interesting. We then read individual posts that contain these themes to pick up nuances, illustrative quotes, and to get a better understanding of what the themes are about and how people are thinking related to them"*.

**Inter-transformation:** During the interaction between the two methods, the output of one method becomes the input of the other. One cannot separate the results of the two methods because they transform into each other throughout the process. Through this way, netnography and text mining inspire, inform and complement each other and create a synergy that is greater than the sum of the two separately. This can be illustrated through Researcher Y's description of a typical NTMI research: *"I together with other researchers look at a lot of materials, go around key words to find what is interesting. Sometimes we talk with the data mining researcher and ask him to analyze certain key words with data mining methods such as word clouds ... You can always go back to big data to ask another question. ....Big data can begin to collect data and analyze it the same time as the netnography goes along and interact with it"*. He also believes that *"Netnography gives you in a very nice way what questions you want to ask in the quantitative material. That is how big data and netnography should work together. Big data would identify something that netnography can't and netnography would pick it up, try to figure out why and maybe bring back new questions for big data"*. This argument clearly demonstrates that insights from the two methods transform into each other during the research process.

**Unity of opposites:** Netnography and text mining reveal textual data from online social media from two completely different facets. One focuses on the thick description and interpretation, while the other focuses on the general pattern and landscape. From this perspective, they are "opposites". However, the two methods together create a united and holistic picture of the consumer insight

contained in the social media, just as what Researcher Y believes: *“these two tools can reveal the changes through time in two dimensions...It is two sides of the same coin”*.

To visualize this information exchange dynamics between netnography and text mining, a Taiji symbol is employed with Yin representing netnography and Yang representing text mining. Taiji is one type of Chinese dialectics where Yin and Yang oppose, connect, inter-transform and together constitute the whole world. This metaphor illustrates that the netnographical and text mining approaches contrast, complement, validate, interact with and transform into each other throughout the insight generation process.



**Figure 12.** Visualization of the information exchange between netnographical and text mining analysis

## Chapter 6 Conclusion

*This chapter summaries the findings of the thesis and through this fulfill the purpose and answer the questions raised earlier.*

The purpose of this thesis was to (1) enhance the understanding of the nature of netnography and text mining; (2) explore the suitability for the two methods' integration based on the analysis of strengths, weaknesses of and potential synergy between both methods; and (3) as an extension develop a research method framework that integrate the two methods and serve as a guideline to generate consumer insights from online social media in commercial market research context.

In relating to the suitability for a mixed methods research approach based on the analysis of strengths and weaknesses of netnography and text mining, the conclusion is that the two methods seem highly suitable for a mixed methods research approach due to the compensation and synergy created by integrating the two methods being noticeable, with only limited overlapping weaknesses.

The development of NTMI method framework includes the method's key participants, typology and the process model.

Key participants wise, researcher, insight user and netizens are the three key players in the NTMI method. The connection among the three players can be bidirectional, unidirectional or nonexistent.

Typology wise, due to both methods sharing the same data source, interacting throughout the research process and being mixed both at and between research process stages, the typology is mixed-model design both within and across stage.

The NTMI method process model consists a funnel-like process with multiple steps: 1) New research requirement; (2) Definition of research field; (3) Social media selection and data collection; (4) Netnographical and text mining analysis, interpretation and validation; (5) Insight; (6) Value. Each step in the process can result in moving to the next or previous step and the information is filtered and narrowed down towards insight which results in value. During this process netnography and text mining interact through elements, transform into each other and holistically unite as two opposites.



## Chapter 7 Discussion

*This chapter discusses the thesis's managerial implications, critical reflections of the study and provides suggestions for further research.*

### 7.1 Managerial implications

Both opportunities and challenges are raised from a managerial point of view due to the findings in the thesis.

#### 7.1.1 Opportunities

Firstly, for commercial market research institutes, the thesis provides thorough research targeting the large volume of online social media data by utilizing two methods that complement each other well. This can as a result enable organizations to make more informed decisions and their innovative processes to produce more value through more accurate targeting enabled by the improved insight generation. As an extension, the research here also provides an understanding that can capture more value out of big data. Furthermore, by providing a systematic process based on multiple points of views and theories the research provides the increased credibility to the mixed methods research approach. This in turn improves the applicability of the methods due to increased ability to sell and accept the methods and the results they generate.

Secondly, for insight users, the thesis can help them understand the pros and cons of netnography and text mining, why they should integrate and how the insight is produced through the integration through demonstrating the elements, attributes and dynamics of the mixed methods research. The thesis can also help them gain a deeper insight into the role they play in the research through the identification of the key participants and their relationships.

Finally, for academia, although the thesis's main purpose is to explore this method in commercial context, but the result could still shed light on academic researches that target online social media in the research field such as business, politics, media, education, sociology and anthropology.

#### 7.1.2 Challenges

Despite the opportunities the findings offer there are also two key challenges identified which should be further discussed relating to acceptance and skill requirements.

### **7.1.2.1 Open research for open data acceptance**

The research has provided clear indications that a mixed methods research approach can have benefits that make the sum greater than its parts. Furthermore, it has demonstrated how the large amounts of data available online can be analyzed from more than one direction, even for the same element. However, as also seen in the empirics, there is innovation resistance towards the new approaches among the organizations that can benefit from them. Due to this the author would like to highlight the importance of continuous research that categorizes, standardize and spread information about these methods and also the importance for companies to embrace the innovative research methods adapting to new research challenges in the digital age. If this is not performed, the methods risk being underestimated due to a lack of market confidence resulting from the innovation resistance.

### **7.1.2.2 Evolution of competencies and collaboration**

One should note that the approach analyzed in this thesis also requires a mixed set of competencies. In other words, to be able to fully harness the benefits of the mixed methods research approach an individual must either possess multiple market research skills or be a part of a well collaborating team. The author would like to highlight that this model has been focusing on integrating two methods yet in theory more methods, with even greater skill discrepancies, could be combined. Due to the continuously increasing amount and dimensions of social media data, the author believes that the importance of mixed methods research approaches is likely to continue growing. This in turn will result in an increase in the requirements of individual researchers with multidisciplinary skills as well as the collaborations within and among teams. There is thereby, an interesting connection between the increase of collaborative technologies partly resulting in the social media these methods aim to analyze and the need to utilize them for the research itself.

### **7.1.2.3 Concern of overlapping weaknesses**

Although this study indicates that netnography and text mining can create significant synergy through the compensation of each other's weaknesses, one should not ignore the overlapping weaknesses which are identified by this study as primarily ethical sensitivity and the lack of interaction with research objects. The former could cause serious moral or legal problems, which is the reason company A only analyses "publicly available information that the individual has chosen to publish online". Meanwhile, the lack of interaction with research objects could lead to misunderstanding or shallow interpretation. Thereby "member check", proposed by Kozinets(2002), as the end phase of netnographical research could be valuable. Those two overlapping weaknesses deserve researchers' attention in practice.

## 7.2 Reflections on the study

Through reflecting on the study the author would like to discuss its generalizability, two key weaknesses of the approach researchers must understand as well as weaknesses in the study itself.

### 7.2.1 Generalizability

The author believes that the generalizability of the main findings of the study is applicable to commercial insight generation from social media in a web 2.0 context similar to Sweden. The reason is that the methods which are analyzed and the goal they aim to achieve are unlikely to have a strong cultural bias. However, they have a web 2.0 bias in the sense that they require certain forms of technologies to be utilized in a certain way. Thereby, in areas where users and/or utilization of these technologies, due to for example law or social concern, are restricted the author cannot say if the findings would apply.

### 7.2.2 Open data as a window

It should be noted that the mixed methods research approach studied in this thesis appear to have two clear weaknesses which both result from that open data, as any other research approach, is a window. As a window it allows the researcher to view into and analyse an extensive amount of information. However, it is still a window and one cannot see around the corners.

Firstly, for the NTMI method, you can only see what people say not what they do. This criticism was raised in the empirics but should be further noted since it is critical that any researcher utilizing the approach understands that this weakness means that one either combines further methods together or accepts that there is a blind spot. Since this blind spot could lead to faulty insights resulting in falsified value and as an ultimate result wasted organizational resources, it is a critical understanding.

Secondly, the approach focuses, as should be clear by now, on online social media. This however results in that further elements that could generate better or new insights can be missed. Such elements could for example be data only attainable offline such as facial expressions or how people communicate face to face in contrast to the textual interfaces social media often consists of.

A solution is that an extended mixed methods research approach including offline behavioural studies could provide a “360” degree view and hence tackle both these shortcomings. Due to this it will also be in the suggestions for further research.

### 7.2.3 Critical reflections

This study has opened up for new and exciting areas of research method within the field of market research. However, some critique can be raised towards the study. Reasoning concerning potential weaknesses and shortcomings regarding the manipulations, the sample and the execution of the experiment results in the following:

Firstly, the study has a limited amount of interviewees, which limits the full scope of knowledge that could be gathered.

Secondly, the interviews risk having a cultural bias since they are focused on northern Europe. These two weaknesses was a result of the limited resources available yet they do weaken the global applicability of the results.

Thirdly, despite multiple steps being taken to provide an accurate and objective interpretation of the empirics, misinterpretations could still be performed due to the disadvantage of qualitative approach.

Finally, even though the model is being tested by the author through a research project which employs the NTMI method in Company A as this thesis is being written, the full test is not yet finished. The real life test indicates that the model does work well, yet without multiple real life tests of the model it cannot be fully validated.

### 7.3 Suggestions for further research

Firstly, as demonstrated in the previous sector, future research could utilize the NTMI method model to real life data to test the validity and usability of the model.

Secondly, this thesis has focused on the commercial utilization of a specific mixed methods research approach in market research and further research could thereby explore how the approach fit more academic applications and in other fields such as sociology, politics etc.

Thirdly, the model presented here aims at increasing the understanding for the mixed methods research approach, yet additional research can be performed regarding the interaction between the methods, cultural affect on the model as well as further validation of the model through field studies.

Fourthly, the thesis has focused on two specific mixed methods research approaches adapted to social media data and further research could explore an offline-online blended mixed methods research approach.

Finally, interaction between more than two methods hasn't been covered by the thesis and would thereby be an additional area, possibly combined with the previous suggestion for further research.

## Bibliography

- Abnor, I. and Bjerke, B., 2009. *Methodology for creating business knowledge*. 3rd ed. London: Sage Publications.
- Ahlqvist, Toni, et al. "Social media roadmaps." Helsinki: Edita Prima Oy (2008).
- Andersson, S., 1979. *Positivism kontra Hermeneutik*. Göteborg: Bokförlaget Korpen.
- Archak, Nikolay, Anindya Ghose, and Panagiotis G. Ipeirotis. "Show me the money!: deriving the pricing power of product features by mining consumer reviews." *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007.
- Biesta, Gert, and Nicholas C. Burbules. *Pragmatism and educational research*. Lanham, MD: Rowman & Littlefield, 2003.
- Bryman, A. and Bell, E., 2007. *Business research methods*. 2nd ed. Oxford: Oxford University Press.
- Catterall, Miriam, and Pauline Maclaran. "Researching consumers in virtual worlds: A cyberspace odyssey." *Journal of Consumer Behaviour* 1.3 (2002): 228-237.
- Clever, N., Kirchner, A., Schray, D., & Schulte, M. (2008). *User-generated Content*. IEessay. Retrieved 11 August 2011 from <http://www.wi.uni-muenster.de/wi/studieren/io/ws08-09/essay/>
- Cristianini, Nello. "Are we there yet?." *Neural Networks* 23.4 (2010): 466-470.
- Culkin, Nigel, David Smith, and Jonathon Fletcher. "Meeting the information needs of marketing in the twenty-first century." *Marketing Intelligence & Planning* 17.1 (1999): 6-12.
- Daugherty, Terry, Matthew S. Eastin, and Laura Bright. "Exploring consumer motivations for creating user-generated content." *Journal of Interactive Advertising* 8.2 (2008): 16-25.
- Dzurec, Laura Cox, and Ivo L. Abraham. "The nature of inquiry: Linking quantitative and qualitative research." *Advances in Nursing Science* 16.1 (1993): 73-79.
- Ereaut, Gill. *Analysis and interpretation in qualitative market research*. Vol. 4. Sage, 2002.
- Fan, Weiguo, et al. "Tapping the power of text mining." *Communications of the ACM* 49.9 (2006): 76-82.
- G & C. Merriam Co. (1913). Noah Porter, ed. *Webster's Revised Unabridged Dictionary* (1913 ed.). G & C. Merriam Co. p. 501. Retrieved 29 January 2014.
- Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li. "Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content." *Marketing Science* 31.3 (2012): 493-520.
- Gillham, B. (2005). *Research interviewing: The range of techniques*. New York: Open University Press.
- Glance, Natalie, et al. "Deriving marketing intelligence from online discussion." *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. ACM, 2005.
- Guba, Egon G., ed. *The paradigm dialog*. Sage, 1990.

- Gundecha, Pritam, and Huan Liu. "Mining social media: A brief introduction." *Tutorials in Operations Research* 1.4 (2012).
- Han, Jiawei, Micheline Kamber, and Jian Pei. *Data mining: concepts and techniques*. Morgan kaufmann, 2006.
- Harrysson, Martin, Estelle Metayer, and Hugo Sarrazin. "How 'social intelligence' can guide decisions." *McKinsey Quarterly* 4 (2012): 81-89.
- Hearst, Marti A. "Untangling text data mining." *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. Association for Computational Linguistics, 1999.
- Howe, Kenneth R. "Against the quantitative-qualitative incompatibility thesis or dogmas die hard." *Educational researcher* 17.8 (1988): 10-16.
- JISC (2012) *The Value and Benefit of Text Mining to UK Further and Higher Education*.
- Johnson, R. Burke, and Anthony J. Onwuegbuzie. "Mixed methods research: A research paradigm whose time has come." *Educational researcher* 33.7 (2004): 14-26.
- Kaplan Andreas M., Haenlein Michael (2010). "Users of the world, unite! The challenges and opportunities of social media". *Business Horizons* 53 (1). p. 61.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53, 59-68. doi:10.1016/j.bushor.2009.09.003.
- Kietzmann, Jan H., et al. "Social media? Get serious! Understanding the functional building blocks of social media." *Business horizons* 54.3 (2011): 241-251.
- Kozinets, Robert V. *Netnography: Doing ethnographic research online*. Sage Publications, 2010.
- Lévi-Strauss, Claude. *The savage mind*. University of Chicago Press, 1966.
- Lincoln, Y. S., and N. K. Denzin. "Introduction: Entering the field of qualitative research." Denzin, NK and Lincoln, YS (1994) *The Handbook of Qualitative Research*, Sage Publications (1994).
- Lincoln, Yvonna S., Susan A. Lynham, and Egon G. Guba. "Paradigmatic controversies, contradictions, and emerging confluences, revisited." *The Sage handbook of qualitative research* (2011): 97-128.
- Lyman, Peter and Nina Wakeford. "Introduction: Going into the (Virtual) Field." *American Behavioral Scientist*. 43 (1999): 359-376.
- Maclaran, Pauline, and Miriam Catterall. "Analysing qualitative data: computer software and the market research practitioner." *Qualitative market research: An international Journal* 5.1 (2002): 28-39.
- Murthy, Dhiraj. "Digital Ethnography An Examination of the Use of New Technologies for Social Research." *Sociology* 42.5 (2008): 837-855.
- Nagel, Thomas. *The view from nowhere*. oxford university press, 1989.
- Nigel Morgan; Graham Jones; Ant Hodges. "Social Media". *The Complete Guide to Social Media From The Social Media Guys*. Retrieved 12 December 2012.

- Pete Comley(2008) Online research communities A user guide International Journal of Market Research Vol. 50 Issue 5: Web 2.0 Special Issue
- Ranfagni, Silvia, Simone Guercini, and Belinda Crawford Camiciottoli. "An interdisciplinary method for brand association research." *Management Decision* 52.4 (2014): 5-5.
- Rickman, Tracy Anna, and Robert M. Cosenza. "The changing digital dynamics of multichannel marketing: The feasibility of the weblog: text mining approach for fast fashion trending." *Journal of Fashion Marketing and Management* 11.4 (2007): 604-621.
- Russom, Philip. "Big data analytics." TDWI Best Practices Report, Fourth Quarter (2011).
- Schwandt, Thomas A. "Three epistemological stances for qualitative inquiry." *Handbook of qualitative research* 2 (2000): 189-213.
- Sechrest, Lee, and Souraya Sidani. "Quantitative and qualitative methods:: Is There an Alternative?." *Evaluation and program planning* 18.1 (1995): 77-87.
- Solutions Big Data IBM(2012)
- State of the media: The social media report 2012. Featured Insights, Global, Media + Entertainment. Nielsen. Retrieved 9 December 2012.
- Stone, Merlin, ed. *Consumer Insight: How to Use Data and Market Research to Get Closer*. Kogan Page Publishers, 2004.
- Tan, Ah-Hwee. "Text mining: The state of the art and the challenges." *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*. 1999.
- Tashakkori, Abbas, and Charles Teddlie. *Mixed methodology: Combining qualitative and quantitative approaches*. Vol. 46. Sage, 1998.
- Tucciarone, Kristy. "Vying for attention: How does advertising affect search and college choice?." *College and University* 83.1 (2007): 26.
- Weiss, Sholom M., et al. *Text mining: predictive methods for analyzing unstructured information*. Springer, 2010.
- Witten, Ian H., and Eibe Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- Witten, Ian H.(2005),Text mining, in *Practical Handbook of Internet Computing* Munindar Paul Singh, ed. Boca Raton: Chapman & Hall/CRC Press, 14-1-14-22.
- Xun, Jiyao, and Jonathan Reynolds. "Applying netnography to market research: The case of the online forum." *Journal of Targeting, Measurement and Analysis for Marketing* 18.1 (2010): 17-31.

## Appendix

### Example questions from interviews

#### For research practitioner representatives

##### Process of the research

- Could you describe the current step-by-step framework of how a research is conducted with the text mining and netnography methodology?

##### Text mining:

- How to select online medias based on the research purpose?
- How to generate insight using text mining?
- Is it possible to identify positive and negative attitude behind the words?

##### Netnography:

- How to utilize the information from the text mining for netnographical research?
- What is the approach of selecting target medias and representative object in this step? How do we know those quotes are representative of the group's voice?
- How does these two methods interact throughout the research?

##### General comments:

- What is your general opinion of this methodology? Pros and Cons?
- How do you think about the validity and reliability of this approach?
- What is missing? Anything that should be improved to better generate consumer insight? What is the direction this tool should go towards?
- Do you concern privacy issue during the online research?
- Since you are the expert in the field of innovation, how do you see the potential of applying text mining and netnography methodology to generate consumer insight for innovation? Pros and cons of this methodology?
- In your opinion, how to improve it to make it a better tool to serve the purpose of innovation or marketing?



**For insight user representative:**

- Could you please shortly introduce your title and role at your client?
- What types of consumer insight does your division utilize? Pros and cons?
- What kind of role does the consumer insight play in your division? How important is it?
- How do you get the insight? Internal research or external resource?
- If external, how does the outsourcing process go? Any interaction during the process?
- Have your division utilize online information to understand customers? Why or why not?
- Have your division utilize big data or netnography to understand customers?
- What is your opinion on these methods? Pros and cons? Opportunities and challenges?