Supplementing consumer insights at Electrolux by mining social media: An exploratory case study

Master thesis within Business Administration

Author: Amit Chaudhary
Tutor: Veronica Gustafsson

Jönköping University

August 2011
Supplementing consumer insights at Electrolux by mining social media: an exploratory case study

Amit Chaudhary

Veronica Gustafsson

August, 2011

Social media, text mining, consumer insights, analytic coding, mixed methods

The aim of this thesis is to explore the possibility of text mining social media, for consumer insights from an organizational perspective.

An exploratory, single case embedded case study with inductive approach and partially mixed, concurrent, dominant status mixed method research design. The case study contains three different studies to try to triangulate the research findings and support research objective of using social media for consumer insights for new products, new ideas and helping research and development process of any organization.

Text mining is a useful, novel, flexible and an unobtrusive method to harness the hidden information in social media. By text-mining social media, an organization can find consumer insights from a large data set and this initiative requires an understanding of social media and its building blocks. In addition, a consumer focused product development approach not only drives social media mining but also enriched by using consumer insights from social media.

Text mining is a relatively new subject and focus on developing better analytical tool kits would promote the use of this novel method. The researchers in the field of consumer driven new product development can use social media as additional evidence in their research.

The consumer insights gained from the text mining of social media within a workable ethical policy are positive implications for any organization. Unlike conventional marketing research methods text mining is social media is cost and time effective.

This thesis attempts to use innovatively text-mining tools, which appear, in the field of computer sciences to mine social media for gaining better understanding of consumers thereby enriching the field of marketing research, a cross-industry effort. The ability of consumers to spread the electronic word of mouth (eWOM) using social media is no secret and organizations should now consider social media as a source to supplement if not replace the insights captured using conventional marketing research methods.

Social media, Web 2.0, Consumer generated content, Text mining, Mixed methods design, Consumer insights, Marketing research, Case study, Analytic coding, Hermeneutics, Asynchronous, Emergent strategy
At the very outset, I would like to thank Prof. Veronica Gustafsson, my academic mentor without her help this thesis would not be complete. I am genuinely grateful for all your guidance and support on this thesis. It has been an honor and privilege learning from you.

I would like to thank Prof. Anna Blombäck, my program manager for her support.

I would like to convey my gratitude to Anton Lundberg, Director, Global Consumer Insight, Electrolux and Sohna Wikman, Consultant for their constant guidance on this thesis.

Amit Chaudhary
August 15, 2011
Jonkoping
Table of Contents
1. Introduction ......................................................................................................................... 1
  1.1 The Specter of Social Media ............................................................................................. 4
  1.2 Impact of internet growth on market research ............................................................... 6
  1.3 Consumer Generated Content: Issues ............................................................................ 7
  1.4 Opportunity ...................................................................................................................... 8
  1.5 Electrolux ....................................................................................................................... 11
    1.5.1 Research Proposal .................................................................................................... 12
    1.5.2 Project description and expectations ...................................................................... 12
    1.5.3 Silverbakk ................................................................................................................ 13
  1.6 Research Question ......................................................................................................... 13
  1.7 Purpose .......................................................................................................................... 13
  1.8 Significance .................................................................................................................... 14
  1.9 Perspective ...................................................................................................................... 14
  1.10 Target Group ................................................................................................................ 14
  1.11 Delimitations ................................................................................................................ 14
  1.12 Overview of the thesis ................................................................................................. 14
  1.13 Definitions .................................................................................................................... 16
2. Selected Methods ............................................................................................................... 17
  2.1 Research Approach ....................................................................................................... 17
  2.2 Research Method .......................................................................................................... 18
  2.3 Research Strategy ......................................................................................................... 20
    2.3.1 Case Study Design ................................................................................................. 22
  2.4 Data Collection ............................................................................................................. 23
  2.5 Data Interpretation ....................................................................................................... 26
  2.6 Research Quality ......................................................................................................... 27
    2.6.1 Validity and Generalization .................................................................................. 29
    2.6.2 Reliability .............................................................................................................. 30
3. Theoretical Framework ...................................................................................................... 31
1. Introduction

In this chapter, the reader will find in-depth overview of the research topic in order to get an idea of the significance of this research. Furthermore, the chapter contains the problem discussion, research questions along with the purpose of the thesis and the perspective. In addition, the chapter will cover the delimitations of the topics. The chapter ends with an overview of the thesis and the definitions used in the thesis.

“Markets are conversations” is a widely discussed statement because historically the marketplace was not only a location where people met to trade goods, but also a place where they talked about their needs and problems. In doing so, people connected to each other. These classical marketplaces rarely exist today, as most people shop in specialized stores where they interact only with sellers. The Internet is advancement in mass media that recreated such “old” marketplaces on a large scale. It hosts and provides access to virtual marketplaces, where consumers can once again easily connect to each other. There is little doubt that the Internet has also changed the way consumers communicate. An increasing number of consumers actively gather online and communicate in web forums, blogs and various kinds of user generated content platforms. They exchange personal experiences and opinions about products and their usage and talk about opportunities for solving product-related problems (need-information). Some even develop product modifications and innovations, which they post online and share with other consumers (solution-information). This makes social media interactive platforms where highly involved consumers exchange existing needs, wishes, experiences, motivations, attitudes and perceptions towards products and brands (Bartl et al., 2009).

![Figure 1.1: Levels of consumer insights](image)

In Fig 1.1 above consumers information need (Level 1, Level 2) can implicitly and explicitly be derived from the consumer dialogue, innovative users either as product prototypes or as solution present solution information (Level 3, Level 4). As a short excursion, I briefly want to demonstrate how the Reef Central Community dealing with aquariums and aquaculture helped in gaining insights for a chemical corporation. More than 200,000 members exchange their experiences and expertise on aquarium chemistry, pump systems, filtration equipment, water
contamination, fish diseases, vermin control or water quality systems and procedures (Bartl et al., 2009). Regularly more than 2,000 members are online at the same time and present their most recent solutions to the peer group. The left part of figure 1.2 shows a rebuilt TV set equipped with a heating circuit and lighting system in order to kill coliform bacteria. All system components activated by the former TV button controls (Bartl et al., 2009).

![Figure 1.2: Innovative problem solutions in social media](image)

The right part of figure 1.2 shows a natural filter system developed by a community member. It is made of materials like acrylic glass, coral sand, activated carbon coal, glass sand and two 35-watt pumps. This short excursion exemplifies that online consumer conversations can be a valuable source of information (Bartl et al., 2009). In his book The New Influencers, Gillin (2007) points out that “Conventional marketing wisdom has long held that an unsatisfied customer tells ten people. That is out of date. In the new age of social media, he or she has the tools to tell 10 million” (p. 4) consumers virtually overnight. Gillin illustrates this potential power by recounting the story of Vincent Ferrari, a blogger who posted an audio recording of his encounter with an AOL customer service representative (Mangold & Faulds, 2009). The representative’s persistent attempts to convince Ferrari not to cancel his account offended listeners’ sensibilities to the extent that approximately 300,000 of them requested to download the audio file. The story went “viral” as it was picked up by thousands of other bloggers and websites. It eventually drew the attention of such mainstream media as The New York Post, The New York Times, and NBC. AOL’s management was embarrassed, to say the least. In a sense, this role of social media enabling customers to talk to one another is an extension of traditional word-of-mouth communication. However, as the Vincent Ferrari story illustrates, the uniqueness lies in the magnitude of the communication. Instead of telling a few friends, consumers now have the ability to tell hundreds or thousands of other people with a few keystrokes (Mangold & Faulds, 2009).

In Fig 1.3 below, I have tried to capture the essence of Vincent Ferrari story in a 2011 perspective considering the fact that growth of social media over the past years has been exponential. The reason behind this interest in social media is twin fold. Organizations are recognizing the increasing importance of Social media and of consumers who are active in online communities. Almquist and Roberts (2000) find that the major factor influencing positive brand equity for one brand over another is consumer advocacy. Social media is context in which consumers often partake in discussions whose goals include attempts to inform and influence
fellow consumers about products and brands (Kozinets 1999; Muniz and O’Guinn 2001). Secondly, the advent of networked computing is opening up opportunities for Organizations to study the tastes, desires, and other needs of consumers who interact in social media (Kozinets 2002).

**Figure 1.3: Impact of social media**

*Observation* The emergence of Internet-based social media has made it possible for one person to communicate with hundreds or even thousands of other people about products and the companies that provide them. Secondly, these consumer-to-consumer communications greatly magnified in the social media are a source of consumer insights for such as but not limited to innovative ideas to support new product research, consumers needs for new products and improving customer service (Mangold & Faulds, 2009). The question from an organizational perspective has changed from why use social media to “*How can consumer insights from social media harnessed for the benefit of the organization?*” In the next section, I define social media, and the need for consumer insights for market research function of an organization.
1.1 The Specter of Social Media

In order to understand social media, it is critical to establish a clear definition of social media so the nature of the content targeted is clear. Any definition of social media should include a description of Web 2.0 and user generated content. Web 2.0 is a term coined by the Web analyst Tim O’Reilly and used to describe the more collaborative use of web technologies (O’Reilly, 2007, p.19). Web 2.0 sites allow its users to create, collaborate, share and publish their own content such as video, text and audio files. Websites such as Dictionary.com and MSN are spaces where the user simply consumed content without contributing to its creation, and are therefore technologically and culturally different to Web 2.0 sites. These sites are ostensibly different to other websites, which actively seek user participation in order to create new content such as Wikipedia and the various blog platforms such as Wordpress and Blogger. User generated content, which is produced using Web 2.0 as a technological platform, can be described as the creation of online content by the users of particular social media platforms. The Organization for Economic Cooperation and Development (OECD) defined user generated content as needing to be placed on a website, to have demonstrated a degree of creativity and, finally, to not have been professionally created (OECD, 2007). Social Media may be defined as the combination of Web 2.0 technologies, and the resulting emergence of user generated content, gives rise to 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 & Haenlein, 2010). As of July 2011, the online social networking application Facebook registered more than 750 million active users. To put that number in perspective, if Facebook was a country it would be the third most populated country in the world! At the same time, every minute, 10 hours of content uploaded to the video sharing platform YouTube. In addition, the image-hosting site Flickr provided access to over 3 billion photographs, making the world-famous Louvre Museum’s collection of 300,000 objects seem tiny in comparison. According to Forrester Research, 75% of Internet surfers used “Social Media” in the second quarter of 2008 by joining social networks, reading blogs, or contributing reviews to shopping sites; this represents a significant rise from 56% in 2007. The growth is not limited to teenagers, either; members of Generation X, now 35-44 years old, increasingly populate the ranks of joiners, spectators, and critics. It is therefore reasonable to say that Social Media represent a revolutionary new trend that should be of interest to companies operating in online space or any space, for that matter. Yet, not overly many firms seem to act comfortably in a world where consumers can speak so freely with each other and businesses have increasingly less control over the information available about them in cyberspace. Today, if an Internet user types the name of any leading brand into the Google search, what comes up among the top five results typically includes not only the corporate webpage, but also the corresponding entry in the online encyclopedia Wikipedia. Here, for example, customers can read that the 2007 model of Hasbro’s Easy-Bake Oven may lead to serious burns on children’s hands and fingers due to a poorly-designed oven door, and that the Firestone Tire and Rubber Company has been accused of using child labor in its Liberian rubber factory. Historically, companies were able to control the information available about them through strategically placed press announcements and good public relations managers. Today, however, firms increasingly relegated to the sidelines as mere observers, having neither the knowledge nor the chance or, sometimes, even the right to alter publicly posted comments provided by their customers. Wikipedia, for example, expressly forbids the participation of firms in its online community. Such an evolution may not be surprising. After all, the Internet started out as nothing more than a giant Bulletin Board System (BBS) that allowed users to exchange
software, data, messages, and news with each other. However, social media also offer an unprecedented opportunity to increase business responsiveness and agility (Kaplan & Haenlin, 2010). For example, recent surveys reveal that 32% of the nearly 250 million bloggers worldwide regularly give opinions on products and brands, 71% of active Internet users read blogs, and 70% of consumers trust opinions posted online by other consumers. Thus, social media is a vast source of business-relevant opinions. A central challenge in leveraging the information present in social media is the enormous scale of the problem. The data of interest to a particular business is in the vast and largely irrelevant, output of millions of bloggers and other online content producers. Consequently, effectively exploiting these data requires development of new, technology driven methods of analysis (Colbaugh & Glass, 2011).

**Consumer Inputs: An analogy to understand the future**

Considering marketing managers’ information requirement, an analogy between a manager navigating his company and driving a car is in Fig 1.4 below. The car has two main domains of activity: internal systems controlled through the brake, accelerator, etc. and feedback given through the instrument panel; in the external domain we are coping with changes as we navigate our way to the destination, and we cope with the external by adjustments to the internal controls. The “tableau de bord”, literally the “dashboard” of a car in English, allows the driver regularly to monitor the function of various performances. In the same way, a manager’s dashboard information system regularly presents key performance measures and highlights any problems. It works as a reporting system that focuses on key control parameters, which could trigger immediate managerial action. Obviously, parameters appearing on the dashboard indicate the internal operational state (Xu, X & Kaye, G 1995).

![Figure 1.4: An Analogy of Consumer Input](image)

This internal information is vital for controlling the operation, but cannot determine the direction of navigation. External information is of strategic importance, since strategic decisions are primarily long term with a balance towards external focus, whereas operational decisions are primarily short term and have an internal focus (Xu, X & Kaye, G 1995). Reid reveals that companies often do not collect environmental intelligence, although management claims
environment knowledge to be something they ought to know about, and in many cases, they do not obtain the types of data they claim to prize. A common problem was that companies frequently lack appropriate structures or organizational format to accommodate data based strategic planning (Reid, D 1989). McNichol’s survey of company presidents and marketing director’s reports that, despite different personal backgrounds, organizational cultures and structures, a remarkably consistent view emerged that consumer insight needed a future, not a past focus (McNichol, J 1993). Some executives shared:

- **Stop using the rear-view mirror to drive the car. Consumer insights should be the eyes and ears of our business and help us plan for tomorrow, not yesterday.**
- **Giving us a 50-page report full of historical data only shows us where we went wrong. I would rather have a two-page sheet that told me what our customers are saying and what our competitors are going to do.**

**Observation** Estimated balance between external monitoring and scanning and internal checking is probably 80:20. While in the car we spend more than 80 per cent of the time scanning the environment, the manager and the collective organization probably spend less than 20 per cent on the external perspective. Marketing managers calling for additional external information reflects a failure of existing systems in supporting managers’ information requirements (Xu, X & Kaye, G 1995). The question from an organizational perspective becomes; *“How can an organization monitor the external environment such as social media for consumer insights?”* The next section pertains to the impact of the growth of internet on performing market research.

### 1.2 Impact of internet growth on market research

Before we consider new techniques, methods, processes and technologies, it is important to look back at how the practice of finding customer inputs has evolved in recent years. In the 1990s, the structure of data feeds for research was straightforward: there was one bucket for company data, retailer data, syndicated marketing and sales data, and syndicated media data. Then there was a second bucket for survey research, which came in a number of shapes and sizes. Custom survey research conducted mostly by phone or in malls; “traditional” qualitative research included primarily focus groups and individual in-depth interviews; syndicated survey research studies rounded out the offerings. In the years after, growth of Internet access—more to the point, the expanded access offered by the availability of increased bandwidth—began to reshape many industries, including marketing research. Marketing research suddenly grew from a two- to a four-bucket practice. One new bucket contained mountains of company and syndicated digital data pulled from Web sites and mobile and social media, all of it feeding the analytical left-brain (Micu et al., 2011).

Another new bucket developed from unprompted consumer feedback—data that were not just answers to researchers’ questions. It came from listening, search analysis, ethnographies, virtual shopping, neuroscience, biometrics, eye tracking, metaphor elicitation, emotion mining, behavioral economics and more—all of it feeding the creative right brain. In addition, the survey research bucket did not stay still—online surveys replaced much of what done by phone or in malls; online access panels, custom online panels, and hosted online communities flourished; do-it-yourself surveys sprang up. New online capabilities (such as virtual shopping and online ethnography) emerged (See Figure 1.5). What were only data feeds in the 1990s became broader and richer information feeds, with video, pictures, emotions, eye movement, facial tracking,
body and brain responses, and more. With so much information has developed a powerful mandate to synthesize all this information—to tell stories that can influence business (Micu et al., 2011).

Figure 1.5: The four buckets of data

*Observation:* The volume of available information is growing rapidly, driving the need for synthesis. However, without a systematic procedure to identify, select and analyze large volumes of consumer conversations on the Internet, researchers confront an information overload (Micu et al., 2011). The question from an organizational perspective becomes: “How can unprompted consumer feedback synthesized for the benefit of the organization?” In the next section I try to share the problems of the consumer generated content in social media.

1.3 Consumer Generated Content: Issues

One can think of consumer-generated content in venues such as forums and blogs as an online channel for word of mouth, which is one of the marketing operationalizations of the somewhat broader concept of social interaction. Numerous academic papers, industry market research, and a large body of anecdotal evidence point to the significant effect of word of mouth on consumer behavior and, in turn, on sales (e.g., Eliashberg et al., 2000; Reichheld and Teal 1996). Online word of mouth, often known as “internet word of mouth” or “word of mouse,” enables
consumers to communicate quickly with relative ease. Numerous cyberspaces such as chat rooms, product review websites, blogs, and brand communities invite and encourage consumers to post their ideas, views, and reviews. The level of activity in these channels of communication has grown exponentially in recent years. In 2008, there were approximately 1.6 million blog postings per day, about double the number in 2007 (Sifry 2008). Consumers making product and brand choices are increasingly turning to computer-mediated communication, for information on which to base their decisions. Besides perusing advertising and corporate Web sites, consumers are using newsgroups, chat rooms, e-mail list servers, personal World Wide Web pages, and other online formats to share ideas, build communities, and contact fellow consumers, seen as more objective information sources. Motion pictures, sports, music, automobiles, fast food, toys, consumer electronics, computers and peripherals, software, cigars, beer, coffee, and many other products and services discussed in online communities whose importance is being increasingly recognized by contemporary marketers (Kozinets, 2002). This discussion in social media enables massive production of free form and interactive data. The data available via social media can give us consumer insights that were not previously possible in both scale and extent. This digital media can transcend the physical world boundaries to study consumers and help measure popular sentiment about a product or brand without explicit surveys (Barbier & Liu, 2011).

However, it is extremely difficult to gain useful information from social media data due to its unique challenges listed below:

I. First, social media data sets are large; consider the 750 million Facebook users as an example. Without automated information processing for analyzing social media, social network data analytics becomes an unattainable in any reasonable amount of time (Barbier & Liu, 2011).

II. Second, social media data sets can be noisy. For example, spam blogs or “splogs” are abundant in the blogosphere, as well as excessive trivial tweets in Twitter (Barbier & Liu, 2011).

III. Third, data from online social media is dynamic, frequent changes and updates over short periods are not only common but also, an important dimension to consider in dealing with social media data. In addition, Wikis created, modified while friend networks ebb, flow and new blogs routinely published. Other data sets may contain some of the challenges present in social media but usually not all at once. For example, the set of traditional web pages create a data set that is a large and noisy but, compared to social media data, is not nearly as dynamic (Barbier & Liu, 2011).

Observation The sheer volume of the data available in social media makes it difficult to identify, collect and analyze. The question from an organizational perspective becomes: “How can an organization collect relevant data from social media without manual intervention?” In the next section I try to converge the four questions and share the opportunity these questions provide for doing this research.

1.4 Opportunity

In the previous sections I have tried to establish not only the potential usefulness of available consumer insights in social media but also raise questions which Organizations may ask when prompted to use social media for consumer insights. In order to share the possible benefits arising from using social media for extracting consumer insights I compare conventional
methods of extracting consumer insights with social media analysis. In Fig 1.6, the comparison of conventional methods to social media analysis reflects the edge social media has. Not only can

![Figure 1.6: Comparison with conventional methods](image)

a larger sample size be targeted but the approach requires minimal prompting for response as the degree of interaction is minimal. Based on previous sections and the comparison with conventional methods above I have listed below the primary reason for performing this research:

**Q. How can consumer insights from social media harnessed for the benefit of the organization?**

*Sub Q. How can an organization monitor the external environment such as social media for consumer insights?*

*Sub Q. How can the unprompted consumer feedback synthesized for the benefit of the organization?*

*Sub Q. How can an organization collect relevant data without manual intervention from social media?*

These four questions prompted me to consider mining social media for consumer insights as my research topic because not only is the theme current and relevant, but also provides an opportunity to innovatively approach a dilemma faced by many organizations.
Based on my previous experience and keen interest in the field of data mining I did some literature review and found in the field of computer sciences the technology to mine information on the internet exists and called “text-mining”. The use of text-mining techniques to derive insights from user-generated content primarily originated in the computer science literature (Akiva et al., 2008; Dave, Lawrence, and Pennock 2003; Feldman et al., 2007; Glance et al., 2005; Hu and Liu 2004; Liu, Hu, and Cheng 2005; Malouf, Davidson, and Sherman 2006). Text mining (sometimes-called knowledge discovery in text) refers to the process of extracting useful, meaningful, nontrivial information from unstructured text (Dörre, Gerstl, and Seiffert 1999; Feldman et al., 1998; Feldman and Sanger 2006). This thesis is an attempt to find out if text mining can support marketing professionals by capturing consumer insights from social media. The theme of my research is that “asynchronous and unobtrusive social media mining can provide organizations consumer insights about such as but not limited to their needs, product feedback and new ideas to support the development of new product.” In Fig 1.6 above, I have shared the thought behind this research where by applying data mining software to social media information can be retrieved which can then be analyzed for consumer insights. Organizations can try to explore the social media using text mining software and a structured analysis of the captured data.

One phenomenon especially appropriate for text mining social media is that of electronic word-of-mouth, or (eWOM.) eWOM is a modified, online, extension of traditional word-of-mouth. It is often used in the literature interchangeably as “word-of-mouse”, “word-online”, “online word-of-mouth”, as well as being associated with “user generated media (UGM)”, or “user generated content (UGC)”. Hennig-Thurau et. al., (2004,) define eWOM as “any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet”. However, this constrains eWOM as a static concept, leaving its potential as an information exchange process
unexplored. Although eWOM is also defined as “peer consumers’ statements made online”, this study, rather than adopting a constrained interpretation of the term, takes eWOM as a dynamic and ongoing information exchange process (Xun, 2010).

In addition, I would like to bring up for discussion the difference in text mining and netnography and the reason behind my selecting text mining. Netnography is a qualitative, interpretive research methodology that uses Internet-optimized ethnographic research techniques to study the social context in online communities. Marketing professor Robert Kozinets created it in 1998 (Kozinets 1998, 2002). The word “Netnography” is a linguistic blend of two words: “Internet” and “Ethnography”. It is also known as multimedia cyber-anthropology or virtual ethnography (Bartlett et al., 2009). Netnography differs from text mining as after identification of the research terms, in netnography the online community identification is the next step to perform the research while in text mining after the research terms have been identified the text mining software searches for them across the selected social media without specifying online communities.

![Figure 1.8: Spectrum of types of netnography](image)

Text mining and netnography have a similarity based on the level of participation of the researcher in an online community. In Fig 1.7, the spectrum of different kinds of netnography shows that a researcher may use netnography without being involved in any online community, a method similar to text mining. Hence, observational netnography and text mining are both unobtrusive and follow similar steps but differ on the identification of online communities in netnography. Ultimately applying data mining to social media is about understanding data about people online, which is at the heart of netnography research (Barbier & Liu, 2011). To conclude, scope of master’s thesis along with the text-mining tool available at Electrolux made me select text mining as my method to search social media for consumer insights. In the next section, I share my collaboration with Electrolux and the software used for text mining.

### 1.5 Electrolux

The Electrolux Group is a Swedish appliance maker. As of 2010 the 2nd largest home appliance manufacturer in the world after Whirlpool, its products sell under a variety of brand names including its own and are primarily white goods and vacuum cleaners. The company also makes appliances for professional use. Forbes Magazine says Electrolux is one of the top 5 companies in consumer durable goods, worldwide, and named it to its list of 130 Global High Performers in 2010. Electrolux products include refrigerators, dishwashers, washing machines, vacuum cleaners, cookers and air-conditioners sold under esteemed brands such as Electrolux, AEG,

1.5.1 Research Proposal
I started applying in Q4 2010 to various organizations to do my master’s thesis and Electrolux was one of them. Anton Lundberg, Director Global Consumer Insights, showed interest in my research proposal about using social media for consumer insights and in the first week of December 2010 I had my first presentation which was appreciated. My research proposal was accepted and that once I signed the non-disclosure agreement from Electrolux and I would have my first meeting at Electrolux, Stockholm in second week of December 2010.

1.5.2 Project description and expectations
During the kick off meeting with Anton on December 16, 2010, I had an overview of the expectations from the thesis. According to Anton, consumer insight is at the core of all product development at Electrolux.

The key action points of the meeting were:

I. Focus area of research to find consumer insights from social media
II. Text mining software to be approved and licensed by Electrolux
III. A consultant to supervise the progress along with Anton
IV. Information about Electrolux to be taken from annual reports and corporate website
V. The output from research may be used for new product development input
VI. The text mining of social media to run for at least a few months to capture adequate data

Going further, I applied a project management approach to the text-mining project such as:

I. Identify the core project team, which would comprise of the academic supervisor, project owner, mentor from Electrolux, academic guide and me.
II. Creation of project distribution list to ensure all stakeholders in the project are updated at the same time by sending a single mail
III. An initial status report to the core team after the first two weeks to evaluate progress and then final report
IV. Monthly call with the Electrolux mentor to share problems, if any
V. Measure the effort by capturing number of hours spent on interpreting the data

Any new initiative has not only support from the actors driving but may also meet with resistance within the organization. The purpose of following a project methodology, Fig 1.8, was to counter any such resistance. Strong implementation leads to client satisfaction and garners internal support. Carefully achieved early success if coupled with evangelizing the success can build support and generate commitment to the initiative. Hence, the idea behind managing the research like a project and capturing the number of hours spent on analysis to share the degree of difficulty. This may help planning and allocation of resources for any future social media-mining project Electrolux may want to execute.
1.5.3 Silverbakk

Electrolux is currently using Silverbakk, third party software for its social media mining needs. The Public Relations team at Stockholm for searching news articles or mention of Electrolux in media primarily uses this software. A Swedish company called Patch6 based out of Stockholm makes Silverbakk. For the purpose of this research can mine the entire social media landscape for the selected terms and provides a report with the number of hits along with the text.

1.6 Research Question

Based on the opportunity description in the previous section and coupled with the work of Henry Mintzberg on emergent strategy, Robert Burgelman’s autonomous strategic processes of adapting organizations along with concept of consumer driven product development process shared in detail in Chapter 3 I have defined the research questions:

Research question: Can text mining social media result in consumer insights for an organization

Sub Q: Study 1, text mining social media for consumer insights, search terms type 1

Sub Q: Study 2, text mining social media for consumer insights, search terms type 2

Sub Q: Study 3, text mining social media for consumer insights, search terms type 3

1.7 Purpose

This research has an exploratory purpose, which backed by the ambition to explore social media for consumer insights from an organizational perspective. The purpose of this thesis is to

I. Refer to the growing relevance, reach and importance of social media
II. Show how organizations can use text mining software and consistent analysis methods to capture consumer insights in multiple areas of interest
III. Briefly mention areas for future research and improvement from an organizational perspective
1.8 Significance

Harvesting information from the data rich environment of online social media in all of its forms is a topic studied by many groups. Market researchers, psychologists, sociologist, ethnographers, businesses, and politicians all can gain useful insights into user behavior via a variety of social networks by applying mining techniques to online social media. The perspective mining social media brings can yield information from online social networks that may not be obvious or attainable otherwise (Barbier & Liu, 2011). In addition, the importance of a dynamic relationship between marketing research and R&D is well known. Mining social media can help market researchers share consumer perspective with R&D which is driven by technological perspective. Thus, text mining social media reiterates the marketing-R&D relationship to the reader. Another significance of social media mining is use of consumer insights in a consumer driven product development approach.

1.9 Perspective

Web 2.0 provides gathering places for Internet users in social-network sites, blogs, forums, and chat rooms. These assembly points leave footprints in the form of colossal amounts of textual data (Micu et al., 2011). This research done from the perspective of an organization that has an interest in gathering consumer insights from this consumer generated textual data is a starting point for improvements in marketing research and social media strategy formation. Secondly, after studying for masters in innovation, an underlying perspective is to put theory to practice in a real life corporate scenario.

1.10 Target Group

Not only organizations interested in mining social media but also organizations interested in providing social media monitoring services can use the findings as a reference. Social media mining is a young and growing field and this research can be a starting point for other students at both under-graduate and graduate level, doing research in text mining the social media.

1.11 Delimitations

This thesis explores the option of using text-mining technique in social media to capture consumer conversations and then interpret them into consumer insights. This thesis is not going to explore:

I. Text mining software’s and their features or perform any comparison
II. The cost aspect of text mining software
III. Industry standards in text mining or related search
IV. The results of the text mining project for Electrolux

1.12 Overview of the thesis

In Chapter 1 (Fig 1.11) an introduction to the thesis and overview of the research topic given, followed by Chapter 2 where I discussed the selected research methods that I used to build up the base for the thesis. Chapter 2 also contains reference to the research methods that one could use when conducting research studies, while sharing giving alternatives and justifying why the particular approach selected.
Chapter 3 defines and elaborates the basic concepts, models and theories used in the thesis. The chosen concepts defined in this chapter are the ones that link directly to the research questions. Furthermore, in this chapter previous research conducted in this area shared also. Social media, text mining and emergent strategy along with autonomous adaptation process as the main concept in the thesis clearly elaborated and it acts as a key for the understanding of the rest of the thesis.

Fig 1.10 Overview of the thesis

As the directions and fabrications of the thesis elaborated in Chapters 2 and 3, in Chapter 4 short description of each case study in this research along with the findings shared.

The main aim of Chapter 5 is to analyze the findings with relation to the concepts and models introduced in Chapter 3. Chapter 6 has the final words about the research done. Finally, in Chapter 7 the suggestions for the further research shared.
1.13 Definitions

Social Media

Social Media 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 & Haenlein, 2010).

Web 2.0

Web 2.0 is a term that was first used in 2004 to describe a new way in which software developers and end-users started to utilize the World Wide Web; that is, as a platform whereby content and applications are no longer created and published by individuals, but instead are continuously modified by all users in a participatory and collaborative fashion. While applications such as personal web pages, Encyclopedia Britannica Online, and the idea of content publishing belong to the era of Web 1.0, that was replaced by blogs, wikis, and collaborative projects in Web 2.0 (Kaplan & Haenlein, 2010).

User Generated Content

When Web 2.0 represents the ideological and technological foundation, User Generated Content (UGC) seen as the sum of all ways in which people make use of Social Media. The term, which achieved broad popularity in 2005, applied to describe the various forms of media content that are publicly available and created by end-users. According to the Organization for Economic Cooperation and Development (OECD, 2007), UGC needs to fulfill three basic requirements in order to be considered as such (Kaplan & Haenlein, 2010):

I. It needs to be published either on a publicly accessible website or on a social networking site accessible to a selected group of people;
II. It needs to show a certain amount of creative effort
III. It needs to have been created outside of professional routines and practices

Consumer Insights

Consumer insights are penetrating discoveries that can lead to specific opportunities. It also translates into a deep understanding of one’s customers and implies actionable use of this understanding. At Electrolux consumer insights are defined as a focused understanding of unfulfilled needs, problems, wants or desires. In order to maintain uniformity, for this thesis the definition used in the thesis is the same as that used by Electrolux.

Text Mining

A specific area of data mining is text mining or text data mining as the process of deriving high quality information from texts (unstructured data). The key idea behind text mining is finding new information in a data set that is hidden or latent (Barbier & Liu, 2011).
2 Selected Methods

This chapter introduces the main methods used together with the justification of their choice. The chapter is structured in the following sequence: first the reasoning approaches have been introduced followed by the choice of research design and research method with clear indication as to which approach has been used in the research and why. Afterwards, the data collection techniques discussed along with research validity and reliability.

Although research is important in both business and academic activities, there is no consensus in the literature on its definition. One reason for this is that research means different things to different people. However, from the many different definitions offered there appears to be agreement that (Amaratunga, 2002):

I. Research is a process of enquiry and investigation
II. It is systematic and methodical
III. Research increases knowledge

Buckley et al., (1975) suggest that an operational definition of research requires the satisfaction of the conditions that:

I. It be an orderly investigation of a defined problem
II. Appropriate scientific methods be used
III. Adequate and representative evidence be gathered
IV. Logical reasoning, uncolored by bias, be employed in drawing conclusions on the basis of the evidence
V. The researcher be able to demonstrate or prove the validity or reasonableness of their conclusions
VI. The cumulative results of research in a given area yield general principles or laws that be applied with confidence under similar conditions in the future

Based on the above I have structured this chapter on explaining my research approach then research method followed by research design, data collection, interpretation and analysis and lastly research reliability.

2.1 Research Approach

According to Easterby-Smith et al., (2002) there are three reasons behind selecting a research approach:

I. Research approach helps the researcher in taking an informed decision about the research design
II. Research approach helps the researcher in identifying the correct research strategy
III. Research approach helps the researcher in catering to any constraints

Based on these three recommendations I did literature review to find the research approaches and select the one, which is relevant and appropriate for this thesis.

Research literature states three different approaches used for an investigation: deductive, inductive, or a combinatory approach of the two abductive. Deductive approach is the testing of specific theory from a developed hypothesis, which then tested rigorously to confirm whether it
is accurate or needs modification. An inductive approach allows for the development of theory based on data collection and analysis (Saunders et al., 2007). An abductive approach is the combination of inductive and deductive investigation; from a specific case a preliminary theory is generated which is then tested again other cases (Patel and Davidson, 2003). Trochim (2006) identifies that the deductive approach starts with a theory about a specific topic, followed by making a hypothesis, tested by collecting data, which addresses it and finally ending up with a confirmation. On the other hand, the inductive approach starts with the observations, identifying the patterns, which are the source of building a hypothesis and at the end drawing some general conclusions and coming up with theories. According to Peirce (1955) the abductive reasoning is based on the notation that there are no priori hypotheses, presumptions and that making the conclusions include preferring one hypothesis over others. This can explain the facts, when there is no basis in previous knowledge that could justify this preference or any checking done (cited in Levin-Rozalis, 2004) (Gilani et al., 2010).

The decision to select an inductive approach for this thesis is an inspiration from the work of Sharan Merriam as she states in her book if the nature of the topic is new and not a lot of research is available in the area; then an inductive approach is more preferable (Merriam, 1998). Using an inductive study would allow for the development of hypotheses, concepts, and abstractions (as opposed to the testing of theory which is employed within a deductive study) (Merriam, 1998). An inductive study would allow us to build toward theory based upon our “observation and intuitive understandings gained in the field” (Merriam, 1998, p.7). Additionally, an inductive study rather than a deductive is beneficial because of the inability of a deductive approach in permitting alternative explanations of what is going on (Saunders et al., 2007). To conclude this section, not much research is available on the subject of text mining in social media and this research is an attempt to supplement the research area.

Going further, in the next section I share the choice of my research methods.

2.2 Research Method

Understanding the various types of research designs can be a daunting task for many beginning researchers, doctoral students, and others. For years, the choice has seemed to be dichotomous; one could choose either a quantitative design or a qualitative design (Leech et al., 2009). 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 (Johnson et al., 2004). Yet, there is a third viable choice, that of mixed methods (Leech et al., 2009).

Mixed methods research is the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study (Johnson et al., 2004). Its logic of inquiry includes the use of induction (or discovery of patterns), deduction (testing of theories and hypotheses), and abduction (uncovering and relying on the best of a set of explanations for understanding one’s results) (e.g., de Waal, 2001). 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 (Johnson et al., 2004).

In general, mixed methods research represents research that involves collecting, analyzing, and interpreting quantitative and qualitative data in a single study or in a series of studies that investigate the same underlying phenomenon. Moreover, mixed methods research falls on a continuum from not mixed (i.e. mono-method designs) to fully mixed methods, with partially mixed designs occupying regions somewhere between mono-method designs and fully mixed method designs (Onwuegbuzie and Johnson 2004). Specifically, mono-methods, at one end of the continuum, involve the exclusive use of either quantitative or qualitative research techniques in a study. Once a study combines quantitative and qualitative techniques to any degree, the study no longer is utilizing a mono-method design. At this level, the study is using a fully mixed design or a partially mixed design. Fully mixed methods designs represent the highest degree of mixing research methods and research paradigm characteristics. This class of mixed research involves using both qualitative and quantitative research within one or more of the following components in a single research study (Leech et al., 2009):

I. The research objective (e.g., researcher uses research objectives from both quantitative and qualitative research, such as the objective of both exploration and prediction)
II. Type of data and operations
III. Type of analysis
IV. Type of inference

The major difference between partially mixed methods and fully mixed methods is that whereas fully mixed methods involve the mixing of quantitative and qualitative techniques within one or more stages of the research process or across these stages, with partially mixed methods, the quantitative and qualitative phases not mixed within or across stages. Instead, with partially mixed methods, both the quantitative and qualitative elements conducted either concurrently or sequentially in their entirety before mixed at the data interpretation stage (Leech et al., 2009). In fig 2.1 below the typology of mixed method research is given. Based on the typology, the research method for this research is partially mixed, concurrent dominant status method as the qualitative data captured during text mining takes prominence in analysis of consumer insights. According to literature, three reasons to conduct mixed methods research are (Tashakkori et al., 2003):

I. Mixed method research can answer questions that other methods cannot
II. Mixed methods research provides stronger inferences
III. Mixed methods research provides a greater diversity of views

Online text mining involves searching for the defined text in the social media. The daily output of the search is a consolidated report which has the number of times the text has been identified in the social media (quantitative component) and along with the location of the text in social media the entire phrase is captured in the report for all the hits (qualitative component). Secondly, this online search is concurrent in nature, which results in data collection of both
Fig 2.1 Typology of mixed research

Qualitative and Quantitative data thereby making mixed methods the right method for this research. Also, the data analysis of quantitative and qualitative data done by using qualitative methods, which is like putting “meat on the bones” of “dry” quantitative findings. The reason I have selected the partially mixed method is that data collection of both quantitative and qualitative data is concurrent and while mixed at data interpretation. Lastly, the three reasons given above to support mixed methods research in the work of Tashakkori and Teddlie (2003) have influenced my decision for selecting mixed methods research. In the next section, I have outlined my research strategy.

2.3 Research Strategy

A case study investigates a contemporary phenomenon in depth and within its real life context when the boundaries between phenomenon and context are not evident clearly (Yin, 2003). The distinctive topics for applying the case study method arise from at least two situations. First and
most important (e.g., Shavelson and Townes, 2002, pp. 99-106), the case study method is pertinent when the research addresses either a descriptive question (what happened?) or an explanatory question (how or why did something happen?) and second, in order to illuminate a particular situation, to get a close (i.e., in-depth and first-hand) understanding of it. The case study method helps the researcher to make direct observations and collect data in natural settings, compared to relying on “derived” data (Bromley, 1986, p. 23) (Yin, 2003).

The use of a case study method is one of the most challenging research strategies when conducting an investigation, deriving from the need to gain a deeper understanding into complex social phenomena (Yin, 2003). While use of case studies is possible in an exploratory, descriptive and/or explanatory manner, Yin (2003) argues the choice of this strategic method depends upon two specific conditions. Firstly, the type of research question(s) being asked (this includes research questions that ask “how” and/or “why” that are tailored for an explanatory purpose) and second, the degree of control the scholar has over contemporary behavioral events. Yin (2003) outlines that the selection of research questions provides the greatest insight into differentiating which strategy is most appropriate for the specific study. When using an exploratory approach, Yin (2003) states that the use of “what” questions are used to “develop pertinent hypotheses and propositions for further inquiry”. On the other hand, “how” and “why” questions are more effective in conducting an explanatory study that can deal with answering questions regarding “operational links needing to be traced over time” (Yin, 2003).

The degree of control over and access to actual behavioral events can also help give clarity to which strategy to use. As outline by Yin (2003) “the case study is preferred in examining contemporary events, but when relevant behaviors cannot be manipulated”. Furthermore, the use of case studies when examining contemporary events allows for the addition of “direct observation of events being studied and interviews of the persons involved in the events” (Yin, 2003).

According to Yin, regardless of its source, case study evidence also can include both qualitative and quantitative data. Qualitative data may be considered non-numeric data—e.g., categorical information that can be systematically collected and presented; quantitative data can be considered numeric data—e.g., information based on the use of ordinal if not interval or ratio measures. Both types of data can be highly complex, demanding analytic techniques going well beyond simple tallies (Yin, 2003).

It is for the following reasons I have chosen a case study approach to help bring greater insight and depth to my research:

I. My research deals with illuminating a particular situation and getting a hands on understanding of use of text mining in social media, which according to Yin is an opportunity to use case study method
II. My research question deals with how to use text mining for getting consumer insights from social media which according to Yin, is suited for an exploratory case study
III. My research observes the contemporary event of interaction of online consumers in social media where the behavior of online consumers cannot be manipulated which according to Yin is best suited for a case study method
IV. My research has both quantitative and qualitative data that a case study can include as study evidence according to Yin
2.3.1 Case Study Design

The matrix below shows that single- and multiple-case studies reflect different design situations and that, within these two variants, there also can be unitary or multiple units of analysis. The resulting four types of designs for case studies are single-case (holistic) designs, single-case (embedded) designs, multiple-case (holistic) designs, and multiple-case (embedded) designs (Yin 2003).

![Fig 2.2 Basic types of designs for case studies](image)

A primary distinction in designing case studies is between single- and multiple-case designs. This means the need for a decision, prior to any data collection, on whether a single case or multiple cases used to address the research questions. The same case study may involve more than one unit of analysis. This occurs when attention is on a subunit or subunits. Irrespective of unit selection, the resulting design is an embedded case study design (see Figure 2.2). In contrast, if the case study examined only the global nature of an organization or of a program, a holistic design is in use (see Figure 2.2) (Yin 2003).

In this thesis, I am asking the question how text mining social media can help in capturing consumer insights, which is the primary case study. In order to support my research I am doing three different studies, which are:

*Study 1, text mining social media for consumer insights, search terms type 1*

*Study 2, text mining social media for consumer insights, search terms type 2*

*Study 3, text mining social media for consumer insights, search terms type 3*
In view of the literature review (Yin, 2003) these three studies take the form of three different units of analysis for one case study hence for my thesis it is a *single study embedded* design.

**Fig 2.3 Thesis structure**

Based on the three sections of research approach, research method and research strategy above, this thesis is an “**exploratory, single case embedded case study having an inductive approach, with partially mixed, concurrent, dominant status mixed methods design**”. (Fig 2.3)

**2.4 Data Collection**

There are two fundamental categories or types of data: primary data and secondary data. According to Crowther and Lancaster (2008) primary data does not exist until or unless it is generated through the research process and is often collected through techniques such as interviewing, observing, surveys etc. On the contrary, secondary data is information that already existed in some form but not necessarily collected for the particular research at hand. This thesis
deals with exploring the possibility of text mining social media to capture consumer insights from an organizational perspective. Data collection for the research was by mining the social media using the Silverbakk mining tool for specific pre-defined terms, which is primary data. Imperative to note here is that the research conducted has a higher dependence on primary data as the secondary data used in the research is only the organization strategy and product development process of Electrolux, both taken from Annual report, 2010. Data collection process for the three studies is as below in Fig 2.4:

**Fig 2.4 Data Collection process in Silverbakk**

**Study 1** Text mining social media for consumer insights, search terms type 1

**Key Search words**

**Dinner party** (Additional attributes must include one of the following: disappointed, frustrated, difficult, disaster, stressed)

**Dinner guests** (Additional attributes must include one of the following: disappointed, frustrated, difficult, disaster, stressed)

**Cooking** (Additional attributes must include one of the following: disappointed, frustrated, difficult, disaster, stressed)

**Search Duration** 07/05/2011– 07/08/2011

**Search in** Blogs, Micro blogs (includes Twitter), Videos, Images, Facebook,

**Silverbakk project name** Cooking 1

**Study 2** Text mining social media for consumer insights, search terms type 2

**Key Search words**

**Dinner party** (Additional attributes must include one of the following: I wish I had, if there was a way to, why don’t they have)
**Dinner guests** (Additional attributes must include one of the following: I wish I had, if there was a way to, why don’t they have)

**Cooking** (Additional attributes must include one of the following: I wish I had, if there was a way to, why don’t they have)

**Search Duration** 07/05/2011 – 07/08/2011

**Search in** Blogs, Micro blogs (includes Twitter), Videos, Images, Facebook,

**Silverbakk project name** Cooking 2

**Study 3 Text mining social media for consumer insights, search terms type 3**

**Key Search words** Microwave, Bake, Grill, Defrost, Barbeque

**Search Duration** 19/07/2011 – 07/08/2011

**Search in** Blogs, Micro blogs (includes Twitter), Videos, Images, Facebook,

**Silverbakk project name** Cooking 3

The selection of search terms done after a discussion between the external consultant, Silverbakk moderator and me. The consensus was to have a generic view and ensure the sample size was large to have a feel of the system, its output and search for insights such as but not limited to product features, functionality or any other consumer needs. In the next part, I discuss the data advantages and disadvantages from text mining.

According to Ghauri and Gronhaug (2010), primary data has a few disadvantages. The main disadvantage is that primary data can take a long time and can cost a lot to collect. In addition, it is difficult to get access: to find consumers, companies or target groups who are willing to cooperate and answer the questions. Text mining does not face the disadvantages of the conventional primary data. Not only is data collection using text mining cost effective but also is not time consuming, as the data collection is automatic by software from the social media. Secondly, the question of access does not arise as all the information is user generated on social media and is out in the open for collection and interpretation. This does gives rise to the debate of public data versus private data, which I have attempted to answer in section 2.6, research quality. The benefits of data collection using text mining social media are as below:

I. **Greater accessibility to a broader cohort of respondents**

Due to the exponential growth of social media, the number of online consumers is increasing and openness, anonymity and decontextualisation of the online environment makes it easier for researchers to text mine larger set of consumers than the conventional methods of marketing research for consumer insights (Xun, 2010).

II. **More economically viable and time-saving than conventional techniques**

Text mining is essentially costless compared with research requiring physical travel and face-to-face fieldwork costs, making studies that are more ambitious potentially more feasible (Xun, 2010).
III. Greater capacity and flexibility for observation and analysis

The availability of a digitally archived data trail greatly strengthens the possible breadth and depth of research by permanently documenting otherwise perishable information. Blogs, discussion forum threads and posts, expert review articles or the latest online video streams are generally archived and stored. The existence of digital “footprints” of historical data not only streamlines the process of transcribing field data, but also permits the building up of insights on an ongoing basis (Xun, 2010).

2.5 Data Interpretation

The metaphor frequently taught in graduate seminars is, data are like raw material, close to the sensory level of experience and observation that requires mining (Kozinets, 2010). Before I share the data interpretation and analysis approach, I would like to share the structure (Fig 2.5) below:

![Data Structure and Analysis](image)

**Fig 2.5 Data Structure and Analysis**

In the previous section of research methods, I have selected mixed method approach and the reason is that the data collected is both qualitative and quantitative. Secondly, induction is a form of logical reasoning in which individual observations built in order to make general observations about a phenomenon. Inductive data analysis approach used in this thesis is a way to manipulate the whole body of extracted consumer generated content from social media collected over the course of text mining (Kozinets, 2010). I have listed the analysis process below arranged in a sequence used in this thesis:

I. Coding

Data collected from text mining the social media such as blog postings, Facebook wall scrawls, Twitter tweets, videos, photographs coded where labels assigned to particular units of data. These codes label the data as belonging to being an example of some more general phenomenon. Categories for coding usually emerge through a close reading of data rather than imposed by prescribed categories (Kozinets, 2010).

II. Noting

Reflections on the data or other remarks are noted and this form of annotation also known as ‘memoing’ (Kozinets, 2010).
III. Abstracting and Comparing

Data sorted and sifted to identify similar phrases, shared sequences, relationships and distinct differences. This abstracting process builds the categorized codes into higher order or more general, conceptual constructs, patterns or processes. The comparing process looks at similarities and differences across data incidents (Kozinets, 2010).

IV. Checking and Refinement

The return to data is to isolate, check and refine the understanding of the patterns, processes, commonalities and differences (Kozinets, 2010).

V. Generalizing

This process includes the elaboration of a small set of generalizations that cover or explain the consistencies in the dataset (Kozinets, 2010).

VI. Theorizing

The last and the most important part of the analysis process include confronting the generalizations gathered from the data with a formalized body of knowledge (Kozinets, 2010).

I would like to bring to attention the idea of hermeneutics and especially hermeneutic circle, a methodological process for interpreting qualitative data. The process is an iterative one in which a ‘part’ of the qualitative data is interpreted and reinterpreted in relation to developing sense of the ‘whole’. These iterations are necessary because a holistic understanding develops over time. Furthermore, initial understandings of the text often modified as later readings provide a more developed sense of the text ‘whole’ meaning (Kozinets, 2010). These two different analytic processes analytic coding mentioned in detail above and hermeneutic interpretation overlap in many interesting ways especially the step of checking and refinement (Kozinets, 2010). In practice, I have used some part of hermeneutic analysis in step VI while doing checking and refinement. Hence, my analysis is not purely analytic coding but a hybrid as it has shades of hermeneutic interpretation as well.

At this point, I would like to debate the choice of my method for data interpretation. I am using text-mining software for data collection while interpreting it manually so why not use software to interpret too. The reasons for not using computer aided qualitative data analysis software (CAQDAS) is because the text captured by data collection software for every hit is not long to apply a software for sorting it into categories and codes because it is already arranged according to the search term and number of hits. Secondly, the various CAQDAS programs available vary in functionalities and usefulness for different analytic situations and in my case are of not much use, hence manual interpretation of text-mined data. Lastly, a bounded investigation like mine requires closeness to data is better suited with a manual interpretation process.

2.6 Research Quality

For capturing consumer insights from social media, one of the methods is text mining. It is a way to understand the discourse and interactions of people engaging in computer mediated communication about market-oriented topics (Kozinets, 2002). The potential insight to be
gained from the analysis of social media content, verified or not, enshrined within a workable ethical policy however is at once enticing and incalculable (Ampofo, 2011). One of the benefits of text mining is that the researcher has the opportunity to conduct unobtrusive research without disturbing the research environment, something that is not possible with other research methods (Kozinets, 2010). Johns et al (2004) underscored this point, claiming that a ‘benefit of virtual research is the extent to which it provides one with the ability to conduct research with virtually no “observer effects.” Thus, virtual settings may provide the opportunity for “naturalistic research” in the extreme (Johns, Chen & Hall, 2004).

One of the main advantages of an undisturbed research environment is that the participants are free to express themselves in a more ‘naturalistic’ way; in which features such as one’s age, gender, ethnicity, and aesthetic appearance do not dominate social interaction (see fig 2.6). In such a setting, people are more likely to respond to the content of other’s interaction rather than their appearance or personality’ (Johns, Chen & Hall, 2004, p.6). As with any new methods, of course, there may be as many uncertainties and challenges as opportunities.

Essentially, text mining social media suffers from at least four major weaknesses:

I. Authenticity

Text mining social media faces its most important issue of being unable to verify the accuracy of the information contained in social media content. Cyberspace appears to be a dark hallway filled with fugitive egos seeking to entrap the vulnerable neophyte (Beckmann & Langer, 2005).
II. An underdeveloped analytical toolkit

Despite the proliferation of publicly available research tools, text mining has a less developed system of analytical tools, often relying on manual interpretation of the researcher (Xun, 2010). I have raised this point earlier while describing the reason for manually interpreting data.

III. Potentially poor quality of textual discourse

Contributing to discourse electronically through a discussion forum or by any other mechanism that requires keying in responses automatically limits the communication of written cues. This requires users both to have a certain degree of literacy as well as to pay more attention when composing messages, or interpreting and paraphrasing the remarks of others. In addition, online social media may be difficult to manage in terms of the flow and order of discussion. This requires the researcher to identify the sequence of the discussion messages, and perhaps apply a degree of “cleaning” of raw data into a “proper” logical flow. However, this reconstitution of words and/or discourse needs careful triangulation with other methods to ensure an accurate and objective analysis (Xun, 2010).

IV. Research Ethics

Ethical concerns about text mining turn on two nontrivial, contestable, and interrelated issues: (1) Are online forums a private or a public site? and (2) What constitutes "informed consent" in cyberspace? Heated debate centres on what online information is private as opposed to public. Elgesem (2002) argues that on the Internet, the anonymity derived from such techniques as using nicknames offers an alternative way of protection. It is suggested that informed consent is not required (Newhagan, 1997). King (1996) firmly believes that the boundary of public and private is blurred on the Internet, which renders the consumer “deluded about the quasi-public nature of their ostensibly private communications” thus text mining social media research may pose a real risk to general online members. He argues that consent must be obtained from the participants. This may not be practical, according to Hudson and Bruckman, (2004). However, Frankel and Siang (1999) argue that from a legal point of view it is the informant’s responsibility to determine what information they make public on the Internet. A clear consensus on ethically appropriate procedures for text mining is yet to emerge. Johnson (2001) also blames high software development costs for the slow adoption of digital consumer research. Perhaps, the inertia of ‘ivory tower’ academic mindsets may also be to blame to inhibit the growth and development of this novel set of methods (Xun, 2010).

2.6.1 Validity and Generalization

Validity according to Saunders et al. (2007) is concerned with whether the results that we are getting are really what they appear to be about. There are two types of validity; Internal and external (Merriam, 1998).

I. Internal

Internal validity deals with the concern if research findings actually match reality (Merriam, 1998). In every qualitative research setting there is an interpreter observing the phenomena being studied. In this scenario, I am the interpreter and this fact changes unconsciously the reality slightly according to my understanding. There are however ways to go about this issue. To
strengthen the internal validity in this particular study I have tried to incorporate strategies posed by Merriam’s (1998). An important way to increase the internal validity is to triangulate i.e. using multiple sources of data, or methods to confirm the findings. As shared earlier I am using both analytic and hermeneutic methods of interpretation to enhance validity.

II. External

External validity sometimes referred to as generalization (Saunder, Lewis and Thornhill, 2007). It deals with uncertainty if results of one study are applicable to other situations (Merriam, 1998). The aim of this study is to generalize my findings outside the empirical setting and explore the use of social media for not only consumer insights but also for finding consumer complaints to improve customer service for any organization that has interest in mining social media. Therefore, generalization is relevant in my research in the sense that Saunders et al. (2007) discuss the term. However, there is a view of external validity called, reader or user generalization where the findings of one study can apply to other people in other situations (Merriam, 1998).

2.6.2 Reliability

Reliability according to Saunders et al. (2007) refers to the extent which data collection techniques or analysis procedures would yield consistent findings. Assessment of reliability of research happens by asking three questions:

I. Will the measures yield the same results on other occasions
II. Will similar observations be reached by other observers
III. Is there transparency in how sense was made from data

In the context of text mining social media, as I am using automated data collection process the data collected is on the search terms and if the software runs on another occasion would include similar results while adding on the latest consumer generated data. Hence, consistency in data collection rules out any subject or participant error and bias, two of the threats to reliability. This leads to the question of observer bias and error. The use of automated text mining software ensures consistency in data collection and all the hits for the search terms would be similar if any other researcher were to collect data from text mining of social media using same attributes attached to the search terms. To reiterate any new consumer generated content would be additionally visible if data collection happens on two different dates. This take care of observer error too, the third threat to reliability.

By using a hybrid approach of analytic coding and hermeneutics for manually interpreting data, I have tried to minimize observer bias. As shared earlier in previous section 2.5, this hybrid approach ensures not only transparency in data interpretation but also consistency in observations. Another observer would identify a consumer insight similarly subject to using the same type of coding structure. It is my endeavor to share interpreted data, which is coherent, contradiction free, comprehensible to the reader and supported with relevant examples and ensure minimal observer bias.
3 Theoretical Framework

This chapter defines, explains and reviews the theoretical concepts used in the thesis. The chapter starts with an introduction of social media followed functional blocks of social media. The motives for mining social media based on theoretical frameworks and prior research discussed followed by the model proposed for mining social media.

3.1 What is social media?

There seems to be confusion among managers and academic researchers alike as to what exactly should be included under this term, and how Social Media differ from the seemingly interchangeable related concepts of Web 2.0 and User Generated Content. It therefore makes sense to take a step back and provide insight regarding where Social Media come from and what they include (Kaplan & Haenlin, 2010).

By 1979, Tom Truscott and Jim Ellis from Duke University had created the Usenet, a worldwide discussion system that allowed Internet users to post public messages. Yet, the era of Social Media as we understand it today probably started about 20 years earlier, when Bruce and Susan Abelson founded “Open Diary,” an early social networking site that brought together online diary writers into one community. The term “weblog” was first used at the same time, and truncated as “blog” a year later when one blogger jokingly transformed the noun “weblog” into the sentence “we blog.” The growing availability of high-speed Internet access further added to the popularity of the concept, leading to the creation of social networking sites such as MySpace (in 2003) and Facebook (in 2004) (Kaplan & Haenlin, 2010).

This, in turn, coined the term “Social Media,” and contributed to the prominence it has today. The most recent addition to this glamorous grouping has been so-called “virtual worlds”: computer based simulated environments inhabited by three-dimensional avatars. Perhaps the best-known virtual world is that of Linden Lab’s Second Life (Kaplan & Haenlein, 2009). Although the list of the aforementioned applications may give some idea about Social Media, a formal definition of the term first requires drawing a line to two related concepts that frequently are named in conjunction with it: Web 2.0 and User Generated Content. Web 2.0 is a term that first used in 2004 to describe a new way in which software developers and end-users started to utilize the World Wide Web. Used as a platform whereby content and applications are no longer created and published by individuals, but instead are continuously modified by all users in a participatory and collaborative fashion (Kaplan & Haenlin, 2010).

The concept of “Web 2.0” began with a conference between O’Reilly Media and MediaLive International. Dale Dougherty, a web pioneer and vice president of O’Reilly, noted that “Web 2.0”—the second generation of web companies—had emerged (O’Reilly, 2005). Internet expert O’Reilly (2005) demonstrated the following seven principles of Web 2.0: web as platform, harnessing collective intelligence, data as the next “Intel inside,” the end of software release cycle, lightweight programming models, software above the level of a single device, and rich user experiences. Turban et al. (2009) argue that Web 2.0 is the popular term for advanced Internet technologies and applications including blogs, wikis (documents written collaboratively in a simple Web markup language), RSS, and social bookmarking (Chia-Liang et al., 2011). When Web 2.0 represents the ideological and technological foundation, User Generated Content (UGC) can be seen as the sum of all ways in which people make use of Social Media. The term,
which achieved broad popularity in 2005, is usually applied to describe the various forms of media content that are publicly available and created by end-users (Kaplan & Haenlin, 2010). According to the Organization for Economic Cooperation and Development (OECD, 2007) UGC needs to fulfill three basic requirements in order to be considered as such. First, it needs to be published either on a publicly accessible website or on a social networking site accessible to a selected group of people. Second, it needs to show a certain amount of creative effort; and finally, it needs to have been created outside of professional routines and practices. The first condition excludes content exchanged in e-mails or instant messages; the second, mere replications of already existing content (e.g., posting a copy of an existing newspaper article on a personal blog without any modifications or commenting); and the third, all content that has been created with a commercial market context in mind (Kaplan & Haenlin, 2010). Within this general definition there are various types of Social Media that need to be distinguished further. However, although most people would probably agree that Wikipedia, YouTube, Facebook, and Second Life are all part of this large group, there is no systematic way in which different Social Media applications is categorized. In addition, new sites appear in cyberspace every day, so it is important that any classification scheme takes into account applications which may be forthcoming. To create such a classification scheme, and to do so in a systematic manner, a set of theories in the field of media research (social presence, media richness) and social processes (self-presentation, self-disclosure), the two key elements of Social Media come handy. Regarding the media-related component of Social Media, social presence theory (Short, Williams, & Christie, 1976) states that media differ in the degree of “social presence” defined as the acoustic, visual, and physical contact that can be achieved—they allow to emerge between two communication partners. Social presence is influenced by the intimacy (interpersonal vs. mediated) and immediacy (asynchronous vs. synchronous) of the medium, and can be expected to be lower for mediated (e.g., telephone conversation) than interpersonal (e.g., face-to-face discussion) and for asynchronous (e.g., e-mail) than synchronous (e.g., live chat) communications. The higher the social presence, the larger the social influence that the communication partners have on each other’s behavior. Closely related to the idea of social presence is the concept of media richness. Media richness theory (Daft & Lengel, 1986) based on the assumption that the goal of any communication is the resolution of ambiguity and the reduction of uncertainty. It states that media differ in the degree of richness they possess—that is, the amount of information they allow to be transmitted in a given time interval—and that therefore some media are more effective than others in resolving ambiguity and uncertainty (Kaplan & Haenlin, 2010).

<table>
<thead>
<tr>
<th>Low</th>
<th>Self-presentation/ Self-disclosure</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Blogs</td>
<td>High</td>
</tr>
<tr>
<td>Low</td>
<td>Collaborative projects (e.g., Wikipedia)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Content communities (e.g., YouTube)</td>
<td></td>
</tr>
</tbody>
</table>

**Fig 3.1 Classification of Social Media**
Combining both dimensions leads to a classification of Social Media, which is above in Fig 3.1. With respect to social presence and media richness, applications such as collaborative projects (e.g., Wikipedia) and blogs score lowest, as they are often text-based and hence only allow for a relatively simple exchange. On the next level are content communities (e.g., YouTube) and social networking sites (e.g., Facebook) which, in addition to text-based communication, enable the sharing of pictures, videos, and other forms of media. On the highest level are virtual game and social worlds (e.g., World of Warcraft, Second Life), which try to replicate all dimensions of face-to-face interactions in a virtual environment. Regarding self-presentation and self-disclosure, blogs usually score higher than collaborative projects, as the latter tend to be focused on specific content domains. In a similar spirit, social networking sites allow for more self-disclosure than content communities do. Finally, virtual social worlds require a higher level of self-disclosure than virtual game worlds, as the latter are ruled by strict guidelines that force users to behave in a certain way (e.g., as warriors in an imaginary fantasy land) (Kaplan & Haenlin, 2010).

Going further, I now define the six different types of social media:

I. Collaborative projects

Collaborative projects enable the joint and simultaneous creation of content by many end-users and are, in this sense, probably the most democratic manifestation of UGC. Within collaborative projects, one differentiates between wikis—that is, websites which allow users to add, remove, and change text-based content—and social bookmarking applications—which enable the group-based collection and rating of Internet links or media content. Exemplary applications within this category include the online encyclopedia Wikipedia, a wiki currently available in more than 230 different languages, and the social bookmarking web service Delicious, which allows the storage and sharing of web bookmarks. The main idea underlying collaborative projects is that the joint effort of many actors’ leads to a better outcome than any actor could achieve individually; this is similar to the efficient-market hypothesis in behavioral finance (Fama, 1970). From an organizational perspective, firms must be aware that collaborative projects are trending toward becoming the main source of information for many consumers. As such, although not everything written on Wikipedia may actually be true, it is believed to be true by more and more Internet users (Kaplan & Haenlin, 2010).

II. Blogs

Blogs, which represent the earliest form of Social Media, are special types of websites that usually display date-stamped entries in reverse chronological order (OECD, 2007). They are the Social Media equivalent of personal web pages and can come in a multitude of different variations, from personal diaries describing the author’s life to summaries of all relevant information in one specific content area. Blogs are usually managed by one person only, but provide the possibility of interaction with others through the addition of comments. Due to their historical roots, text-based blogs are still by far the most common. Nevertheless, blogs have also begun to take different media formats. For example, San Francisco-based Justin.tv allows users to create personalized television channels via which they can broadcast images from their webcam in real time to other users. Yet, as is the case with collaborative projects, blogs do not come without risks. These generally present in two fashions. First, customers who for one reason or another turn out to be dissatisfied with or disappointed by the company’s offerings may decide
to engage in virtual complaints in the form of protest websites or blogs (Ward & Ostrom, 2006), which results in the availability of potentially damaging information in online space. Second, once firms encourage employees to be active on blogs, they may need to live with the consequences of staff members writing negatively about the firm. Microsoft’s former “technical evangelist” Robert Scoble, for example, had a tendency to fiercely criticize the products of his employer before he decided to leave the Redmond-based software company in 2006 (Kaplan & Haenlin, 2010).

III. Content communities

The main objective of content communities is the sharing of media content between users. Content communities exist for a wide range of different media types, including text (e.g., BookCrossing, via which 750,000+ people from over 130 countries share books), photos (e.g., Flickr), videos (e.g., YouTube), and PowerPoint presentations (e.g., Slideshare). Users on content communities are not required to create a personal profile page; if they do, these pages usually only contain basic information, such as the date they joined the community and the number of videos shared. From a corporate viewpoint, content communities carry the risk of being used as platforms for the sharing of copyright-protected materials (Kaplan & Haenlin, 2010).

IV. Social networking sites

Social networking sites are applications that enable users to connect by creating personal information profiles, inviting friends and colleagues to have access to those profiles, and sending e-mails and instant messages between each other. These personal profiles can include any type of information, including photos, video, audio files, and blogs (Kaplan & Haenlin, 2010).

V. Virtual game worlds

Virtual worlds are platforms that replicate a three-dimensional environment in which users can appear in the form of personalized avatars and interact with each other as they would in real life. In this sense, virtual worlds are probably the ultimate manifestation of Social Media, as they provide the highest level of social presence and media richness of all applications discussed thus far. Besides their use for in-game advertising (similar in idea to product placement in blockbuster movies), the high popularity of virtual game worlds can also be leveraged in more traditional communication campaigns. Japanese automotive giant Toyota, for example, used pictures and mechanics from the World of Warcraft application in its latest Tundra commercial to reach the 2.5 million players in the U.S. alone (Kaplan & Haenlin, 2010).

VI. Virtual social worlds

The second group of virtual worlds, often-referred to as virtual social worlds, allows inhabitants to choose their behavior more freely and essentially live a virtual life similar to their real life. As in virtual game worlds, virtual social world users appear in the form of avatars and interact in a three-dimensional virtual environment; however, in this realm, there are no rules restricting the range of possible interactions, except for basic physical laws such as gravity. This allows for an unlimited range of self presentation strategies, and it has been shown that with increasing usage intensity and consumption experience, users of virtual social worlds—“residents,” as they prefer to be called—show behavior that more and more closely mirrors the one observed in real
life settings (Haenlein & Kaplan, 2009; Kaplan & Haenlein, 2009a, 2009b). Arguably, the most prominent example of virtual social worlds is the Second Life application, founded and managed by the San Francisco-based company Linden Research Inc. Besides doing everything that is possible in real life (e.g., speaking to other avatars, taking a walk, enjoying the virtual sunshine), Second Life also allows users to create content (e.g., to design virtual clothing or furniture items). In addition, it allows users to sell this content to others in exchange for Linden Dollars, a virtual currency traded against the U.S. Dollar on the Second Life Exchange. Some residents are so successful in this task that the virtual money earned that way complements their real life income. Virtual social worlds offer a multitude of opportunities for companies in marketing (advertising/communication, virtual product sales/v-Commerce, marketing research), and human resource and internal process management (Kaplan & Haenlin, 2010).

*Observation* The section above defines various types of social media. From a research perspective, the primary question is *which type of social media is to be text mined for maximum consumer insights?*

### 3.2 The seven functional blocks of social media

![Fig 3.2 Honeycomb of Social Media](image)

Although it is clear that for better or for worse social media is very powerful, many executives are reluctant or unable to develop strategies and allocate resources to engage effectively with social media. Consequently, firms regularly ignore or mismanage the opportunities and threats presented by creative consumers (Berthon, Pitt, McCarthy, & Kates, 2007). One reason behind this ineptitude is a lack of understanding regarding what social media are, and the various forms
they can take (Kaplan & Haenlein, 2010). To help address this gap in knowledge Fig 3.2 above illustrates a honeycomb framework of seven social media building blocks. Utilized individually and together, these blocks can help managers make sense of the social media ecology, and to understand their audience and their engagement needs. In true social media fashion, the origins of this framework can be attributed to a number of bloggers: principally, Gene Smith (2007) of the Atomiq.org, who developed and combined ideas discussed by Matt Webb (2004) of interconnect.org; Stewart Butterfield (2003) of sylloge.com and Peter Morville (2004) of semanticstudios.com (Keitzmann et al., 2011).

The framework in Figure 3.2 is a honeycomb of seven functional building blocks: identity, conversations, sharing, presence, relationships, reputation, and groups. Each block allows the reader to unpack and examine (1) a specific facet of social media user experience, and (2) its implications for firms. These building blocks are neither mutually exclusive, nor do they all have to be present in a social media activity. They are constructs that allows making sense of how different levels of social media functionality is configured (Keitzmann et al., 2011).

I. Identity

The identity functional block represents the extent to which users reveal their identities in a social media setting. This can include disclosing information such as name, age, gender, profession, location, and also information that portrays users in certain ways. For instance, Kaplan and Haenlein (2010) explain that the presentation of a user’s identity can often happen through the conscious or unconscious ‘self-disclosure’ of subjective information such as thoughts, feelings, likes, and dislikes. Consequently, users and social media sites have different discourse preferences and aims. Many individuals who participate in online activities use their real names (e.g., Guy Kawasaki, a leading blogger and managing director of Garage Technology Ventures), while other influential social media mavens are known by their nicknames, or ‘handles’ (e.g., hummingbird604 is Raul Pacheco, a blogger and educator on environmental issues) (Keitzmann et al., 2011).

II. Conversations

The conversations block of the framework represents the extent to which users communicate with other users in a social media setting. Many social media sites are designed primarily to facilitate conversations among individuals and groups. These conversations happen for all sorts of reasons. People tweet, blog, et cetera to meet new like-minded people, to find true love, to build their self-esteem, or to be on the cutting edge of new ideas or trending topics. Yet others see social media as a way of making their message heard and positively impacting humanitarian causes, environmental problems, economic issues, or political debates (Beirut, 2009) (Keitzmann et al., 2011).

III. Sharing

Sharing represents the extent to which users exchange, distribute, and receive content. The term ‘social’ often implies that exchanges between people are crucial. In many cases, however, sociality is about the objects that mediate these ties between people (Engestrom, 2005); the reasons why they meet online and associate with each other (Keitzmann et al., 2011).
IV. Presence

The framework building block presence represents the extent to which users can know if other users are accessible. It includes knowing where others are, in the virtual world and/or in the real world, and whether they are available. In the virtual world, this happens through status lines like ‘available’ or ‘hidden.’ Given the increasing connectivity of people on the move, this presence bridges the real and the virtual. Another direct implication of presence is that it is linked to the traits of other functional blocks in the honeycomb framework, including conversations and relationships. For instance, drawing upon ideas by Kaplan and Haenlein (2010), firms should recognize that social media presence is influenced by the intimacy and immediacy of the relationship medium, and that higher levels of social presence are likely to make conversations more influential (Keitzmann et al., 2011).

V. Relationships

The relationships block represents the extent to which users can be related to other users. By ‘relate,’ we mean that two or more users have some form of association that leads them to converse, share objects of sociality, meet up, or simply just list each other as a friend or fan. Consequently, how users of a social media platform are connected often determines the what-and-how of information exchange. The general rule is that social media communities, which do not value identity highly, also do not value relationships highly. Because the implications of the relationship block are numerous, two properties structure and flow from social network theory (Borgatti & Foster, 2003; Granovetter, 1973) help to explain the importance of different relationship traits. The structural property of a user’s relationships refers to how many connections they have and their position in their network of relationships. Research shows that the denser and larger a user’s portfolio of relationships is, and the more central his or her position in the portfolio, the more likely that user is to be an influential member (‘influencer’) in their network. The flow property of user relationships refers to the types of resources involved in individual relationships and how these resources were used, exchanged, or transformed. It describes the strength of a relationship: strong relationships are ‘long-lasting, and affect-laden’ (Krackhardt, 1992), while weak ones are ‘infrequent and distant’ (Hansen, 1999). It also refers to the ‘multiplexity’ of relationships; that is, when users are connected by more than one type of relationship (e.g., they are work colleagues and friends). (Keitzmann et al., 2011).

VI. Reputation

Reputation is the extent to which users can identify the standing of others, including themselves, in a social media setting. Reputation can have different meanings on social media platforms. In most cases, reputation is a matter of trust, but since information technologies are not yet good at determining such highly qualitative criteria, social media sites rely on ‘mechanical Turks’: tools that automatically aggregate user-generated information to determine trustworthiness (Keitzmann et al., 2011).

VII. Groups

The group’s functional block represents the extent to which users can form communities and sub communities. The more ‘social’ a network becomes, the bigger the group of friends, followers, and contacts. A widely discussed relationship-group metric is Dunbar’s Number, proposed by
anthropologist Robin Dunbar (1992), who theorized that people have a cognitive limit, which restricts the number of stable social relationships they can have with other people to about 150. Social media platforms have recognized that many communities grow well beyond this number, and offer tools that allow users to manage membership. Two major types of groups exist. First, individuals can sort through their contacts and place their buddies, friends, followers, or fans into different self-created groups (e.g., Twitter has lists). Second, groups online can be analogous to clubs in the offline world: open to anyone, closed (approval required), or secret (by invitation only). Facebook and Flickr have groups, for instance, with administrators who manage the group, approve applicants, and invite others to join. The direct implication of groups is fairly straightforward. It is assumed that a social media community would enjoy a way to group its users, even when the number of likely contacts is low for each member initially (Keitzmann et al., 2011).

*Observation* The section above defines the seven functional building blocks of social media. The relevance of the building blocks is high as they can help in identifying the right media for mining based upon the purpose of mining. From a research perspective, *how can building blocks help in identifying the appropriate social media for text mining of consumer insights?*

### 3.3 Motives for text mining Social Media: Deliberate or Emergent Strategy

How do strategies form in organizations? Research into the question is necessarily shaped by the underlying conception of the term. Since strategy has almost inevitably been conceived in terms of what the leaders of an organization ‘plan’ to do in the future, strategy formation has, not surprisingly, tended to be treated as an analytic process for establishing long-range goals and action plans for an organization; that is, as one of formulation followed by implementation. As important as this emphasis may be, Minzberg argues that it is seriously limited, that the process needs to be viewed from a wider perspective so that the variety of ways in which strategies actually take shape can be considered the process of strategy formation based on the definition of strategy as ‘a pattern in a stream of decisions’ (Mintzberg, 1972, 1978; Mintzberg and Waters, 1982, 1984; Mintzberg et al., 1986, Mintzberg and McHugh, 1985; Brunet, Mintzberg and Waters, 1986).

This definition was developed to ‘operationalize’ the concept of strategy, namely to provide a tangible basis on which to conduct research into how it forms in organizations. Streams of behavior could be isolated and strategies identified as patterns or consistencies in such streams. The origins of these strategies may be investigated, with particular attention paid to exploring the relationship between leadership plans and intentions and what the organizations actually did. Using the label strategy for both of these phenomena one called intended, the other realized—encouraged that exploration. Comparing intended strategy with realized strategy, as shown in Figure 3.3, allows us to distinguish deliberate strategies-realized as intended-from emergent strategies patterns or consistencies realized despite, or in the absence of, intentions. For a strategy to be perfectly deliberate—that is, for the realized strategy (pattern in actions) to form exactly as intended—at least three conditions would seem to have to be satisfied. First, there must have existed precise intentions in the organization, articulated in a relatively concrete level of detail, so that there can be no doubt about what was desired before any actions were taken. Secondly, because organization means collective action, to dispel any possible doubt about whether or not the intentions were organizational, they must have been common to virtually all the actors: either shared as their own or else accepted from leaders, probably in response to some
sort of controls. Thirdly, these collective intentions must have been realized exactly as intended, which means that no external force (market, technological, political, etc.) could have interfered with them.

Fig 3.3 Types of Strategies

The environment, in other words, must have been either perfectly predictable, totally benign, or else under the full control of the organization. These three conditions constitute a tall order, so that we are unlikely to find any perfectly deliberate strategies in organizations. Nevertheless, some strategies do come rather close, in some dimensions if not all. For a strategy to be purely emergent there must be order-consistency in action over time-in the absence of intention about it. (No consistency means any strategy or at least unrealized strategy-intentions not met.) It is difficult to imagine action in the total absence of intention-in some pocket of the organization if not from the leadership itself-such that we would expect the purely emergent strategy to be as rare as the purely deliberate one (Mintzberg & Waters, 1985).

Below is a list of a variety of types of strategies that fall along this continuum, beginning with those closest to the deliberate pole and ending with those most reflective of the characteristics of emergent strategy:

I. Planned Strategy

In this first type, called planned strategy, leaders at the centre of authority formulate their intentions as precisely as possible and then strive for their implementation-their translation into collective action-with a minimum of distortion, ‘surprise-free’. To ensure this, the leaders must first articulate their intentions in the form of a plan, to minimize confusion, and then elaborate this plan in as much detail as possible, in the form of budgets, schedules and so on, to pre-empt discretion that might impede its realization. Those outside the planning process may act, but to the extent possible, they cannot decide. Programmes that guide their behavior built into the plan, and formal controls instituted to ensure pursuit of the plan and the programmes. But the plan is of no use if it cannot be applied as formulated in the environment surrounding the organization so the planned strategy is found in an environment that is, if not benign or controllable, then at least rather predictable (Mintzberg & Waters, 1985).

II. Entrepreneurial Strategy

In this second type of strategy, the condition of precise, articulated intentions is relaxed. Here, one individual in personal control of an organization is able to impose his or her vision of direction on it. Because such strategies are rather common in entrepreneurial firms, tightly controlled by their owners, they can be called entrepreneurial strategies. In this case, the force for pattern or consistency in action is individual vision, the central actor’s concept of his or her
organization’s place in its world. This is coupled with an ability to impose that vision on the organization through his or her personal control of its actions (e.g. through giving direct orders to its operating personnel). Of course, the environment must again be co-operative. However, entrepreneurial strategies most commonly appear in young and/or small organizations (where personal control is feasible), which are able to find relatively safe niches in their environments. Indeed, the selection of such niches is an integral part of the vision. These strategies can, however, sometimes be found in larger organizations as well, particularly under conditions of crisis where all the actors are willing to follow the direction of a single leader who has vision and will. In two important respects, however, entrepreneurial strategy can have emergent characteristics as well. First, vision provides only a general sense of direction. Within it, there is room for adaptation: the details of the vision can emerge en route. Secondly, because the leader’s vision is personal, it can also be changed completely. To put this another way, since here the formulator is the implementer, step by step, that person can react quickly to feedback on past actions or to new opportunities or threats in the environment and can thus reformulate vision. This adaptability distinguishes the entrepreneurial strategy from the planned one (Mintzberg & Waters, 1985).

III. Ideological Strategies

Vision can be collective as well as individual. When the members of an organization share a vision and identify so strongly with it that they pursue it as an ideology, then they are bound to exhibit patterns in their behavior, so that clear realized strategies can be identified. These may be called ideological strategies. Can an ideological strategy be considered deliberate? Since the ideology is likely to be somewhat overt (e.g. in programmes of indoctrination), and perhaps even articulated (in rough, inspirational form, such as a credo), intentions can usually be identified. The question thus revolves around whether these intentions can be considered organizational and whether they are likely to be realized as intended. In an important sense, these intentions would seem to be most clearly organizational. Whereas the intentions of the planned and entrepreneurial strategies emanate from one centre and are accepted passively by everyone else, those of the ideological strategy are positively embraced by the members of the organization (Mintzberg & Waters, 1985).

IV. Umbrella Strategy

Consider the scenario where the condition of tight control (whether bureaucratic, personal or ideological) over the mass of actors in the organization and, in some cases, the condition of tight control over the environment as well is relaxed. Leaders who have only partial control over other actors in an organization may design what can be called umbrella strategies. They set general guidelines for behavior-define the boundaries-and then let other actors maneuver within them. In effect, these leaders establish kinds of umbrellas under which organizational actions are expected to fall-for example that all products should be designed for the high-priced end of the market (no matter what those products might be). When an environment is complex, and perhaps somewhat uncontrollable and unpredictable as well, a variety of actors in the organization must be able to respond to it. In other words, the patterns in organizational actions cannot be set deliberately in one central place, although the boundaries may be established there to constrain them. From the perspective of the leadership (if not, perhaps, the individual actors), therefore, strategies are allowed to emerge, at least within these boundaries. In fact, we can label the umbrella strategy not only deliberate and emergent (intended at the centre in its broad outlines but not in its
specific details), but also ‘deliberately emergent’ (in the sense that the central leadership intentionally creates the conditions under which strategies can emerge) (Mintzberg & Waters, 1985).

V. Process Strategy

The leadership functions in an organization in which other actors must have considerable discretion to determine outcomes, because of an environment that is complex and perhaps also unpredictable and uncontrollable. Instead of trying to control strategy content at a general level, through boundaries or targets, the leadership instead needs to exercise influence indirectly. Specifically, it controls the process of strategy making while leaving the content of strategy to other actors. Again, the resulting behavior would be deliberate in one respect and emergent in others: the central leadership designs the system that allows others the flexibility to evolve patterns within it (Mintzberg & Waters, 1985).

VI. Unconnected Strategy

The unconnected strategy is perhaps the most straightforward one of all. One part of the organization with considerable discretion—a subunit, sometimes even a single individual because it is loosely coupled to the rest, is able to realize its own pattern in its stream of actions. How deliberate or emergent are these unconnected strategies? Since they come neither from a central leadership nor from intentions in the organization at large, they would seem to be relatively emergent from the perspective of the entire organization. However, from the perspective of the unit or individual involved, clearly they can be deliberate or emergent, depending on the prior existence of intentions (Mintzberg & Waters, 1985).

VII. Consensus Strategy

In no strategy so far discussed above is the condition of prior intention left out and his type is rather more clearly emergent. Here many different actors naturally converge on the same theme, or pattern, so that it becomes pervasive in the organization, without the need for any central direction or control and called the consensus strategy. Unlike the ideological strategy, in which a consensus forms around a system of beliefs (thus reflecting intentions widely accepted in the organization) the consensus strategy grows out of the mutual adjustment among different actors, as they learn from each other and from their various responses to the environment. Thereby find a common, and probably unexpected, pattern that works for them. In other words, the convergence is not driven by any intentions of a central management, nor even by prior intentions widely shared among the other actors. It just evolves through the results of a host of individual actions. Of course, certain actors may actively promote the consensus; perhaps even negotiate with their colleagues to attain it (as in the congressional form of government). However, the point is that it derives more from collective action than from collective intention (Mintzberg & Waters, 1985).

VIII. Imposed Strategy

All the strategies so far discussed have derived in part at least from the will (if not the intentions) of actors within the organization. The environment has been considered, if not benign, then at least acquiescent. But strategies can be imposed from outside as well; that is, the environment can directly force the organization into a pattern in its stream of actions, regardless of the
presence of central controls. The clearest case of this occurs when an external individual or group with a great deal of influence over the organization imposes a strategy on it (Mintzberg & Waters, 1985).

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Major features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planned</td>
<td>Strategies originate in formal plans: precise intentions exist, formulated and articulated by central leadership, backed up by formal controls to ensure surprise-free implementation in benign, controllable or predictable environment; strategies most deliberate</td>
</tr>
<tr>
<td>Entrepreneurial</td>
<td>Strategies originate in central vision: intentions exist as personal, unarticulated vision of single leader, and so adaptable to new opportunities; organization under personal control of leader and located in protected niche in environment; strategies relatively deliberate but can emerge</td>
</tr>
<tr>
<td>Ideological</td>
<td>Strategies originate in shared beliefs: intentions exist as collective vision of all actors, in inspirational form and relatively immutable, controlled normatively through indoctrination and/or socialization; organization often proactive vis-à-vis environment; strategies rather deliberate</td>
</tr>
<tr>
<td>Umbrella</td>
<td>Strategies originate in constraints: leadership, in partial control of organizational actions, defines strategic boundaries or targets within which other actors respond to own forces or to complex, perhaps also unpredictable environment; strategies partly deliberate, partly emergent and deliberately emergent</td>
</tr>
<tr>
<td>Process</td>
<td>Strategies originate in process: leadership controls process aspects of strategy (hiring, structure, etc.), leaving content aspects to other actors; strategies partly deliberate, partly emergent (and, again, deliberately emergent)</td>
</tr>
<tr>
<td>Unconnected</td>
<td>Strategies originate in enclaves: actor(s) loosely coupled to rest of organization produce(s) patterns in own actions in absence of, or in direct contradiction to, central or common intentions; strategies organizationally emergent whether or not deliberate for actor(s)</td>
</tr>
<tr>
<td>Consensus</td>
<td>Strategies originate in consensus: through mutual adjustment, actors converge on patterns that become pervasive in absence of central or common intentions; strategies rather emergent</td>
</tr>
<tr>
<td>Imposed</td>
<td>Strategies originate in environment: environment dictates patterns in actions either through direct imposition or through implicitly pre-empting or bounding organizational choice; strategies most emergent, although may be internalized by organization and made deliberate</td>
</tr>
</tbody>
</table>

Fig 3.4 Summary description of strategy types

Strategy formation walks on two feet, one deliberate, and the other emergent. As noted earlier, managing requires a light deft touch-to direct in order to realize intentions while at the same time responding to an unfolding pattern of action. The relative emphasis may shift from time to time but not the requirement to attend to both sides of this phenomenon. A summary is shared in Fig 3.4 above (Mintzberg & Waters, 1985).

Observation In order to understand and appreciate the reason behind an organization's decision to text mine social media it is imperative to know if the strategy is deliberate or emergent. From a research perspective, the question is, Electrolux is interested in exploring the concept of text mining social media a novel and new method, to supplement consumer insights. What type of Organizational strategy does it have and is this strategy deliberate or emergent?

In the next section, I discuss the new product development process and level of collaboration.
3.4 Motives for text mining Social Media: New product development and Collaboration

In literature and in practice, product innovation is generally conceptualized as a five-stage New Product Development (NPD) process—ideation, concept development, product design, product testing, and product introduction (e.g., Ulrich & Eppinger, 2003; Urban & Hauser, 1993). Firms use varied techniques to solicit customer input in order to create better new products faster. In the front-end stages of the NPD process (ideation and concept development), firms use market research techniques like focus groups, consumer surveys and quantitative techniques like conjoint analysis to create, test, and refine new product concepts. At later stages in the NPD process, firms use quality function deployment, prototyping, product testing, and test marketing to design and improve products and marketing strategies for new product introduction (Urban & Hauser, 1993). While firms have always sought to hear the “voice of the customer,” consumers have traditionally tended to play a passive role as recipients of the firm’s innovation activities. Firms seek to improve fit between their offerings and consumer needs by surveying consumers and importing knowledge from leading-edge consumers into the firm (von Hippel, 1988). Drivers of the firm’s innovation success include the firm’s market sensing ability (Day, 1994), effective R&D and manufacturing routines (Hayes, Wheelwright, & Clark, 1988) and the right balance of organizational competences (Verona, 1999). The traditional perspective on consumer engagement implicitly views value creation and innovation as a firm-centric activity, with most information flowing in one direction from the customer to the firm (Prahalad & Ramaswamy, 2004). When consumers, viewed as passive recipients of innovation, the firm has a limited understanding of customer knowledge developed within their specific contexts of experience, and there is little emphasis on iterative dialogue to refine and enhance ideas. Further, if one excludes costly tools like participant observation (Leonard & Rayport,1997), there is little opportunity to engage communities of consumers to tap into the social aspects of knowledge. Finally, the firm tends to be biased towards listening to its current customers, and even among these, to its most important customers (Sawhney et al., 2005).

Involvement of consumers to support the development of products is also a subject of debate. Some critics state that consumers do not know what they want in the future (Ulwick, 2005) and cannot formulate those needs (Ciccantelli and Magidson, 1993). Consumers are notoriously lacking foresight, since, according to Hamel and Prahalad (1994), they cannot imagine something that does not exist. Firms can lose their position of industry leadership, if they listen too carefully to their customers (Christensen and Bower, 1996). We can think of situations in which consumer involvement will be of little use or even detrimental, we will follow the lead of those who think that it will be useful in most cases, because “continuing involvement of consumers with developers in an integrated fashion sustains the melding of consumer needs with technical capabilities”(Saguy and Moskowitz, 1999: 70). The involvement of consumers (by need inputs, concept reviews and product tests) contributes to the superiority of a product (e.g., Cooper and Kleinschmidt, 1993; Chandy and Tellis, 2000; Van Kleef et al., 2005).

The concept of consumer-led new product development is a market-oriented innovation strategy developed specifically for the manufacturers of consumer goods, as it focuses on the share of market intelligence pertaining to the end-users. It is also an integrated concept concerning the application of consumers’ current and future needs, and its determinants, in the development of innovative products with true added value (Grunert et al.,1996; Lord, 2000; Urban & Hauser, 1993; van Trijp & Steenkamp, 1998; Costa, et al., 2006).
Its main pillars are:

I. Consumer needs should be the starting point of New Product Development processes (Costa, et al., 2006).

II. New Product Development should aim at the fulfillment of consumer needs and the realization of consumer value rather than at the development of products or enabling technologies per se (Costa, et al., 2006).

III. Given that increased sales and satisfactory returns on investment can be achieved if consumer needs are effectively identified and satisfied, the measure of success of a New Product Development process should be the degree of fit between the new product and the needs of the targeted consumers (Costa, et al., 2006).

The key stages in the formulation of the consumer-led NPD concept follow closely a market-oriented approach: need identification, idea development to address the need, product development to substantiate the idea and the product’s market introduction to communicate the fulfillment of a need (Urban & Hauser, 1993) as shown in Fig. 3.5 (Costa, et al., 2006).

The Internet is an open, cost-effective and ubiquitous network (Afuha, 2003). These attributes make it a global medium with unprecedented reach, contributing to reduce constraints of geography and distance (Cairncross, 1997). Further, the Internet potentially allows firms to overcome the trade-off between richness and reach because it is interactive in nature (Evans & Wurster, 1999). In the physical world, communicating (and absorbing) rich information requires physical proximity or personal interactions with customers. These constraints limit the number of customers that the firm can have dialogue with. On the other hand, the firm can interact with a large number of customers through customer surveys, but this type of interaction does not allow for a rich dialogue. However, Internet-based virtual environments allow the firm to engage a much larger number of customers without significant compromises on the richness of the interaction (Sawhney et al., 2005).
Internet-based collaboration mechanisms can be mapped to the NPD process based on two important dimensions—the nature of customer involvement that is needed, and the stage of the NPD process at which the customer involvement is desired. In terms of the nature of customer involvement, Internet-based collaboration mechanisms are classified into mechanisms that emphasize reach versus mechanisms that emphasize richness of the interaction. While the reach-versus-richness trade-off is not as severe on the Internet as it is in the physical world, it still is a decision that the firm needs to make. The firm may want to emphasize richness over reach if it is interested in generating ideas and insights, while it may value reach over richness if it is interested in validating hypotheses with a representative sample of customers. Internet-based collaboration mechanisms may also be classified in terms of their usefulness at different stages of the NPD process: some mechanisms are more relevant at the front-end stages of the process (idea generation and concept development stages), while others are better applied to enhance the back-end stages of the process (product design and testing). Figure 3.6 shows a variety of Internet-based mechanisms classified on these two dimensions.

![Fig 3.6 Collaboration Mechanisms Based on Nature of Collaboration and Stage of NPD](image)

Observation Firms in their product development cycles may have different degrees of collaboration with online consumers at different stages of product development process. The question from research perspective becomes, *At Electrolux what is the collaboration mechanism while text mining social media for consumer insights?*
3.5 Research Framework

Social media and text mining are two relatively new fields and not much research is available on the concept of text mining social media for consumer insights. Hence, the search for theoretical frameworks, models and prior research has been difficult to say the least. On this note, I would agree that information is available in the scientific community about the importance of social media and how organizations should have a social media strategy but very limited literature is available on text mining social media for consumer insights. Researcher’s current focus is on citing the benefits of social media strategy but very little is available about harnessing the strengths of social media. This unavailability of research gets a supporting hand by the lack of availability of analytical tool kits for analyzing text mined social data. In view of the above, I had very little information available to create a theoretical framework, which I explain in the next section.

In order to provide a theoretical framework for this research I went back to basics using the three W’s researchers tend to try to answer, Why, What and When.

I. Literature review reveals benefits of using social media and in previous sections I have established the benefits of doing text mining but as the research is done for benefit of Electrolux, it is important to ask, Why did Electrolux decide to text mine social media? Is this a part of the bigger picture, intentional or the organization went ahead without any previous intentions of using social media for consumer insights? The question helped me bring in the concept of deliberate and emergent strategies based on the work of Harry Mintzberg.
II. The next logical step after asking why text mine social media is *What* type of social media to text mine? The building blocks of social media help in identifying the appropriate media and this forms the second question of my research framework.

III. The consumer insights from text mining social media are collated and ready but *When* does an organization use that information in its product development process? What is the level of collaboration and at what stage of new product development are these insights useful?

In Fig. 3.7 above, I have shared the visualization of the research framework. In my limited knowledge Emergent versus Deliberate Strategy, Types of Social media to text mine and the use of these insights in the New Product Development stage completes the research framework.

To conclude, the basic purpose of this research is to text mine social media and extract consumer insights and by understanding the type of social media along with the Organizational strategy towards text mining and its new product development process, a deeper insight gained from a business and an academic purpose.

In the next chapter, I share research findings.
4 Empirical Findings

The chapter reports the empirical findings by first, giving a short description of Electrolux and second presenting the findings about the consumer focus and new product development process. The second part contains the findings of the three text mining studies conducted for consumer insights.

Electrolux is a global leader in household appliances and appliances for professional use (Fig 4.1), selling more than 40 million products to customers in more than 150 markets every year. The company focuses on innovations that are thoughtfully designed, based on extensive consumer insight, to meet the real needs of consumers and professionals. Electrolux products include refrigerators, dishwashers, washing machines, vacuum cleaners, cookers and air-conditioners sold under esteemed brands such as Electrolux, AEG, Eureka and Frigidaire. In 2010, Electrolux had sales of SEK 106 Billion and 52,000 employees (Annual Report, 2010).

<table>
<thead>
<tr>
<th>Category</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CONSUMER</strong></td>
<td><strong>KITCHEN</strong> For household kitchens throughout the world, Electrolux sells cookers, ovens, refrigerators, freezers, dishwashers, hoods and small appliances. The increasing role of the kitchen as a meeting place for family and friends gives Electrolux a unique display area.</td>
</tr>
<tr>
<td><strong>DURABLES</strong></td>
<td><strong>LAUNDRY</strong> Washing machines and tumble-dryers are the core of the Electrolux product offering for cleaning and care of textiles. Innovations and a growing preference for higher capacity, user-friendliness as well as lower consumption of water and energy are driving demand for Electrolux products.</td>
</tr>
<tr>
<td><strong>FLOOR-CARE</strong></td>
<td><strong>Electrolux</strong> vacuum cleaners and accessories are sold to consumers worldwide. A strong, global distribution network and an attractive product offering are important competitive advantages. All production is located in low-cost areas.</td>
</tr>
<tr>
<td><strong>PROFESSIONAL</strong></td>
<td><strong>PRODUCTS</strong> Electrolux sells a range of products for professional kitchens and laundries. High productivity, maximum utilization of resources and an extensive service network are key factors for purchases by professionals. Electrolux has a global presence, and is largest in Europe.</td>
</tr>
</tbody>
</table>

Fig 4.1 Electrolux product categories
4.1 Organizational Strategy

Based on the information available from the annual report Electrolux has a defined vision statement behind the slogan. “Thinking of you” expresses the Electrolux offering: To maintain continuous focus on the consumer, whether it is product development, design, production, marketing, logistics or service. Electrolux achieves profitable growth by offering products and services that are preferred by consumers, that benefit people as well as the environment, and for which customers are prepared to pay higher prices. Innovative products, lower costs and a strong Electrolux brand create a foundation for improving Group profitability (Annual Report, 2010).

The reason I have italicized some words in the statement is to bring to attention that Electrolux has clear and articulated intentions, to ensure their pursuit, in an environment that is acquiescent. Electrolux has not only articulated their intentions in the form of a plan, to minimize confusion, but also elaborated this plan detail in the form of budgets, schedules and so on, to pre-empt discretion that might impede its realization. In addition, the organization strategy is consumer driven with focus on consumer engagement across multiple activities.

Evident from the information available this deliberate change in strategy is as recent as the past decade. According to the annual report, “Electrolux has completed its transformation from a manufacturing company into an innovative, consumer-driven company with an organization built on a strong understanding of evolving consumer needs. The combination of innovative products and a strong brand in the premium segment with the ability to utilize the global strength and reach of the Group have equipped Electrolux with the best prerequisites ever for profitable growth.” To support my argument I share the statement made by the former CEO Hans Straberg, “We have returned to a consumer focus – meaning that rather than selling what we produce, we produce what sells” (Bari, 2010).

Fig 4.2 Organizational Strategy
4.2 Consumer Focus

In order to identify the depth of consumer focus of Electrolux I requested information from Electrolux and based on the presentation made by Elisabetta Bari, Project Director, Consumer innovation program, I would like to share my findings.

Fig 4.3 Focus of Consumer Innovation Program

In 2004, Electrolux launched the consumer innovation program where product development would focus on consumer insights where opportunity areas identified from consumer prospective. The consumer was the center of the process (Fig 4.3) and by performing conventional marketing research, getting consumer insights was the core. As shared earlier at Electrolux a consumer insight is a focused understanding of unfulfilled needs, problems, wants or desires (Bari, 2010).

Based on definition of consumer insight, this innovation program developed the Product Management Flow, the product development process at Electrolux. The annual report categorically states all new Electrolux products are born out of the Group’s process for consumer-driven product development, which is a holistic process for managing products – from the cradle to the grave (Annual Report, 2010). This product development approach, focuses on the customer, utilizing data on customer preferences, needs and requirements but, in addition, includes display of different solutions/concepts for the customers, so the customers can react to different proposed design solutions. In order to maintain its consumer driven focus Electrolux uses tools such as but not limited to market surveys, interviews and home visits of its consumers to get consumer insights (Bari, 2010).
4.3 Product Management Flow

At Electrolux, the product development process is Product Management Flow and in order to understand the process I would like to share the example of Electrolux RealLife® dishwasher. The RealLife® dishwasher developed for the “real life” in a modern household, where consumers want to be able to fit everything into the dishwasher and get everything properly cleaned, no matter how the machine is loaded. RealLife® dishwasher was launched in the markets successfully in March 2010 (Annual Report, 2010). Due to space constraints, I have put the process in two parts but it is one continuous process.

**Fig 4.4 Product Management Flow Part 1**

It took several consumer market surveys to understand the most important factors when buying a dishwasher. An important conclusion was that nobody loads his or her dishwasher in the same way. The consumer insights from the market surveys were translated into three main consumer needs:

I. more loadable space; consumers want the machine to fit all types of dishware, such as big pots, plates and delicate glasses
II. perfect cleaning, no matter how you load
III. baskets to fit items of any size and shape

The selected concepts translated into features and functions such as FlexiSpray™ Arm and RealLife® Baskets to fit and clean items of any size and shape (Annual Report, 2010).

**Fig 4.5 Product Management Flow Part 2**

The launch of the RealLife® dishwasher was prepared in a consistent and integrated way through several consumer touch points, including point-of-sale support, web campaigns, advertisements and TV. (Annual Report, 2010)
4.4 Study 1: Text mining social media for consumer insights, search terms type 1

Key Search words  

**Text Mining Duration** 07/05/2011– 07/08/2011

**Dinner party** (Additional attributes must include one of the following: disappointed, frustrated, difficult, disaster, stressed)

**Dinner guests** (Additional attributes must include one of the following: disappointed, frustrated, difficult, disaster, stressed)

**Cooking** (Additional attributes must include one of the following: disappointed, frustrated, difficult, disaster, stressed)

**Silverbakk project name** Cooking 1

**OVERVIEW**

![Fig 4.6 Frequency of hits for the duration in social media](image)

*Fig 4.6 Frequency of hits for the duration in social media*

The text mining of social media for the search words of study 1 commenced on May 7 and went on until August 7, a period of three months. The daily frequency of hits of the search terms in social media is in Fig 4.6.

**POSTS PER MEDIA**

![Fig 4.7 Distribution of posts per media](image)

*Fig 4.7 Distribution of posts per media*
The search terms of Study 1 featured in different social media and the distribution per media is in Fig 4.7. On adding up all the number of posts, the data set size was **7983** that required interpretation using analytic coding and hermeneutics for extracting consumer insights if any.

**TAG CLOUD**

```
a ah all applebees articles baking bbq because bob books burger
business cafe ceasers cheese chef chicken choi completed cook
cooking diet disaster drama drive easy eating english
entertainment episode episodes eun evans family fly food for
fried fries gas geomsa grill gun han hard hari health healthy
heavy hee high ho home hoo how hungry hut hyun in jung
kentucky kfc kim king kitchen korean lee little losmcdonalds
meal metal movies new news park perkins pizza princess recipe
recipes restaurant reviews rock romance ronald s
sandwich season secret song step the to tv video weight
wendys with you
```

**Fig 4.8 Tag Cloud Study 1**

For Study 1 the tag cloud generated by the system is above in Fig 4.8 and the terms, which appear the most with search terms, placed accordingly. It is imperative to understand that a tag cloud is a very simple analysis. It might suggest words that be inspected further but is not really an analysis. In these three studies, the tag cloud sharing is for knowing the output types from Silverbakk, the text mining software.

**4.5 Study 2 Text mining social media for consumer insights, search terms type 2**

**Key Search words**

**Text Mining Duration** 07/05/2011– 07/08/2011

**Dinner party** (Additional attributes must include one of the following: I wish I had, if there was a way to, why don’t they have)

**Dinner guests** (Additional attributes must include one of the following: I wish I had, if there was a way to, why don’t they have)

**Cooking** (Additional attributes must include one of the following: I wish I had, if there was a way to, why don’t they have)
The text mining of social media for the search words of study 2 commenced on May 7 and went on until August 7, a period of three months. The daily frequency of hits of the search terms in social media is in Fig 4.9.

**OVERVIEW**

![Fig 4.9 Frequency of hits for the duration in social media](image)

The search terms of Study 2 featured in different social media and the distribution per media is in Fig 4.10. On adding up all the number of posts, the data set size was 1355 that required interpretation using analytic coding and hermeneutics for extracting consumer insights if any.

**POSTS PER MEDIA**

![Fig 4.10 Distribution of posts per media](image)

Imperative to note here is the difference between the number of posts in Study 1 and Study 2 as they do have the same primary search terms but the additional attributes have made all the difference.
For Study 2 the tag cloud generated by the system is above in Fig 4.11 and the terms, which appear the most with search terms, placed accordingly. It is imperative to understand that a tag cloud is a very simple analysis. It might suggest words that be inspected further but is not really an analysis. In these three studies, the tag cloud sharing is for knowing the output types from Silverbakk, the text mining software.

4.6 Study 3 Text mining social media for consumer insights, search terms type 3

<table>
<thead>
<tr>
<th>Key Search words</th>
<th>Text Mining Duration 19/07/2011–07/08/2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microwave</td>
<td></td>
</tr>
<tr>
<td>Bake</td>
<td></td>
</tr>
<tr>
<td>Grill</td>
<td></td>
</tr>
<tr>
<td>Defrost</td>
<td></td>
</tr>
<tr>
<td>Barbeque</td>
<td></td>
</tr>
<tr>
<td>Silverbakk project name</td>
<td>Cooking 3</td>
</tr>
</tbody>
</table>

OVERVIEW

The text mining of social media for the search words of study 3 commenced on July 19 and went on until August 7, a period of three weeks. The daily frequency of hits of the search terms in social media is in Fig 4.12.
The search terms of Study 3 featured in different social media and the distribution per media is in Fig 4.13. On adding up all the number of posts, the data set size was **1026517** that required interpretation using analytic coding and hermeneutics for extracting consumer insights if any. The reason for such high number of hits in social media is due to an intentional use of generic words to target a large sample size for extracting consumer insights.

**Fig 4.14 Tag Cloud Study 3**

For Study 3 the tag cloud generated by the system is above in Fig 4.14 and the terms, which appear the most with search terms, placed accordingly. It is imperative to understand that a tag cloud is a very simple analysis. It might suggest words that be inspected further but is not really an analysis. In these three studies, the tag cloud sharing is for knowing the output types from Silverbakk, the text mining software.
5 Analysis

*This chapter aims at interpreting the empirical findings by comparing them to the key concepts, models and theories presented in the Chapter 3 with the goal to answer the research questions*

Reference is invited to the research framework shared in Chapter 3, which is based on the three key questions:

I. What is the organizational strategy behind the decision of Electrolux exploring the idea of test mining social media for consumer insights? Is it deliberate or emergent?

II. What type of social media is appropriate for text mining to extract consumer insights for Electrolux?

III. When in its product development process does Electrolux incorporate the consumer insights extracted from social media?

My approach towards analysis is to *compare the theoretical construct with the research findings while keeping the analysis crisp and short* for ease of understanding. I start with the identification of strategy, followed by the identification of social media type and lastly use of consumer insights in the product development process.

5.1 Strategy identification

Thinking of you” expresses the Electrolux offering: To maintain *continuous focus* on the consumer, whether it is *product development, design, production, marketing, logistics or service*. Electrolux achieves profitable growth by offering products and services that are preferred by consumers, that benefit people as well as the environment, and for which customers are prepared to pay higher prices. Innovative products, lower costs and a strong Electrolux brand create a foundation for improving Group profitability (Annual Report, 2010). The reason I have italicized some words in the statement is to bring to attention that Electrolux has clear and articulated intentions, to ensure their pursuit, in an environment that is acquiescent. Electrolux has not only articulated their intentions in the form of a plan, to minimize confusion, but also elaborated this plan detail in the form of budgets, schedules and so on, to pre-empt discretion that might impede its realization. The first reaction on reading the strategy statement of Electrolux is that this is a planned strategy and some planning is behind the decision to explore text mining of social media. Imperative to note here the leaders do not exercise tight control (whether bureaucratic, personal or ideological) over the mass of actors in the organization. They have set general guidelines for behavior-define the boundaries-and then let other actors maneuver within them. In effect, leaders establish kinds of umbrellas under which organizational actions are expected to fall. Hence, Electrolux has an umbrella strategy, based on the classification by Mintzberg. I would like to state that the strategy may not have an initial impact on the decision to explore text mining but once the results of the exploration are out, the reaction is strategy dependent. If actors are found to stray outside the boundaries (whether inadvertently or intentionally), the central leadership has three choices: to stop them, ignore them (perhaps for a time, to see what will happen), or adjust to them. In fact, we can label the umbrella strategy not only deliberate and emergent (intended at the centre in its broad outlines but not in its specific details), but also ‘deliberately emergent’ (in the sense that the central leadership intentionally creates the
conditions under which strategies can emerge). In my limited understanding, the strategy to explore text mining is *deliberately emergent*, as leadership has given actors the bandwidth to explore areas where strategic learning can take place.

### 5.2 Selecting appropriate social media type

It is difficult to stay abreast of the choices people have for social media platforms. It seems that new sites and services emerge every day, vying for the attention of individuals and communities online. Using tools like social media classification based on media richness and self-disclosure in conjunction with the honeycomb framework is important to understand the social media landscape and identify appropriate social media for text mining of consumer insights.

*Theoretical Framework* Kaplan & Haenlein (2010) in their paper “Users of the world, unite! The challenges and opportunities of Social Media” have shared an interesting classification of social media based on two parameters, media richness and self-disclosure. Media richness has its basis on the assumption that the goal of any communication is the resolution of ambiguity and the reduction of uncertainty. Self-disclosure is the conscious or unconscious revelation of personal information (e.g., thoughts, feelings, likes, dislikes) that is consistent with the image one would like to give. *From the perspective of social media text mining, higher a social media is in self-discourse, the better is the information quality shared by the user.* This framework of social media along with its importance for text mining is in Fig 5.1

<table>
<thead>
<tr>
<th>Social Media Type</th>
<th>Media Richness</th>
<th>Self Disclosure</th>
<th>Examples</th>
<th>Text Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative Projects</td>
<td>Low</td>
<td>Low</td>
<td>Wikipedia</td>
<td></td>
</tr>
<tr>
<td>Blogs</td>
<td>Low</td>
<td>High</td>
<td>Blogs</td>
<td>Important</td>
</tr>
<tr>
<td>Content Communities</td>
<td>Medium</td>
<td>Low</td>
<td>Youtube, Flickr</td>
<td></td>
</tr>
<tr>
<td>Social Networking Sites</td>
<td>Medium</td>
<td>High</td>
<td>Facebook, Twitter</td>
<td>Important</td>
</tr>
<tr>
<td>Virtual Game worlds</td>
<td>High</td>
<td>Low</td>
<td>World of Warcraft</td>
<td></td>
</tr>
<tr>
<td>Virtual Social worlds</td>
<td>High</td>
<td>High</td>
<td>Second Life</td>
<td>Important</td>
</tr>
</tbody>
</table>

**Fig 5.1 Social Media Types**

*Research Findings* In order to either accept or reject the theoretical framework I would like to share the data Fig 5.2, from the three text-mining studies conducted, which shows that maximum consumer information mined from social media type with high self-discourse.

<table>
<thead>
<tr>
<th>Social Media</th>
<th>Study 1</th>
<th>Study 2</th>
<th>Study 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogs</td>
<td>1956</td>
<td>18</td>
<td>18031</td>
<td>20005</td>
</tr>
<tr>
<td>Facebook</td>
<td>1727</td>
<td>293</td>
<td>252860</td>
<td>254880</td>
</tr>
<tr>
<td>Micro Blogs</td>
<td>4014</td>
<td>1040</td>
<td>741544</td>
<td>746598</td>
</tr>
<tr>
<td>Images</td>
<td>12</td>
<td>0</td>
<td>4335</td>
<td>4347</td>
</tr>
<tr>
<td>Videos</td>
<td>274</td>
<td>4</td>
<td>9747</td>
<td>10025</td>
</tr>
</tbody>
</table>

**Fig 5.2 Social Media posts of search terms**

The search tool used for these studies, Silverbakk identifies social networking sites like Twitter as micro-blogs. An operational definition of micro-blog is that it is a broadcast medium in the
form of blogging. A micro-blog differs from a traditional blog in that its content is typically smaller size. Micro-blogs allow users to exchange small elements of content such as short sentences, individual images, or video links. These are mostly short status updates of:

I. What users are doing  
II. Where they are  
III. How they are feeling (the core reason for mining)  
IV. Links to other sites

I would now like to discuss the second classification parameter of social media that is media richness. From the perspective of classification media richness is the amount of information a media allows to transfer and thereby reducing any ambiguity in communication. Though an important parameter but from the perspective of selecting social media for text mining it is relatively less important. The reason behind this statement is that mining social media for consumer insights is about finding what the consumers are sharing online, the content and not the richness of the social media used. In addition, an observation from the theoretical perspective and the research findings I would like to share is the absence of virtual social worlds from the process of text mining. Although rich in self-disclosure, the reason behind the absence of this social media type is technological constraints. First, as this type of social media exhibits very strong and robust user control and privacy policies the technology used in text mining is under developed to access the virtual forums. Second, I would also like to share again the absence of developed analytical tool kits to interpret the data from virtual social worlds. In addition, this reignites the debate about private and public content in online media and brings in the question of ethical text mining. If user controls deny access for content to be captured then any act of trying to capture the said content would not only be unethical but against the spirit of research.

Social media selection criteria 1 Based on the theoretical framework shared above and supported by the research findings it is safe to conclude that any organization interested in text mining social media for consumer insights should focus on social media with high degree of self-disclosure.

Theoretical framework 2 Keitzman et al., in their paper (2011)“Social media? Get serious! Understanding the functional building blocks of social media Business Horizons” shared a honeycomb framework of seven functional building blocks: identity, conversations, sharing, presence, relationships, reputation, and groups. Based on this tool, the two key building blocks of any social media for text mining are conversations and sharing as that is where maximum users exchange or share information.

<table>
<thead>
<tr>
<th>Social Media</th>
<th>Building block</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikipedia</td>
<td>Sharing</td>
</tr>
<tr>
<td>Blogs</td>
<td>Conversation, Sharing</td>
</tr>
<tr>
<td>Youtube, Flickr</td>
<td>Sharing</td>
</tr>
<tr>
<td>Facebook, Twitter</td>
<td>Relationship, Conversation</td>
</tr>
<tr>
<td>World of Warcraft</td>
<td>Group, Presence</td>
</tr>
<tr>
<td>Second Life</td>
<td>Relationship, Identity</td>
</tr>
</tbody>
</table>

Fig 5.3 Social Media types and Building blocks
In Fig 5.3, I have shared the six types of social media and their key building blocks based on the honeycomb tool. In order to identify the appropriate social media for text mining using the honeycomb tool it is important to understand that these building blocks are neither mutually exclusive, nor do they all have to be present in a social media activity.

Research Findings In order to either accept or reject the theoretical framework I would like to share the data Fig 5.4, from the three text-mining studies conducted, which shows that maximum consumer information mined from social media building block type conversation:

<table>
<thead>
<tr>
<th>Social Media</th>
<th>Building block</th>
<th>Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikipedia</td>
<td>Sharing</td>
<td>0</td>
</tr>
<tr>
<td>Blogs</td>
<td>Conversation</td>
<td>20005</td>
</tr>
<tr>
<td>Youtube, Flickr</td>
<td>Sharing</td>
<td>14372</td>
</tr>
<tr>
<td>Facebook, Twitter</td>
<td>Relationships, Conversation</td>
<td>1001478</td>
</tr>
<tr>
<td>World of Warcraft</td>
<td>Group, Presence</td>
<td>0</td>
</tr>
<tr>
<td>Second Life</td>
<td>Relationship, Identity</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig 5.4 Social Media types, Building blocks and posts

In the three studies conducted for extracting consumer insights by mining social media the maximum posts are from the social media type conversation followed by building block sharing. The findings are in agreement with the theoretical framework 2.

Fig 5.5 Building blocks in single social media

When examining the social media ecology, it quickly becomes clear that many sites have struck a careful balance among the different blocks of the honeycomb. Some focus more on identity, some more on sharing, et cetera. In Fig 5.5 I have shared, the building blocks of Facebook and
YouTube to share the existence of multiple building blocks. The darker the color of a block, the greater this social media functionality is within the site (Keitzmann et al., 2011).

**Social media selection criteria 2** Based on the theoretical framework shared above and supported by the research findings it is safe to conclude that *any organization interested in text mining social media for consumer insights should focus on social media comprising of building blocks conversation and sharing.*

Going further, I have compared the two selection criteria’s in Fig 5.6 with my research findings to ensure the findings support the use of both criteria’s. In addition, the number of posts from media types helps in confirming the two selection criteria’s and identifying in the appropriate media type.

<table>
<thead>
<tr>
<th>Social Media Type</th>
<th>Self Disclosure</th>
<th>Building block</th>
<th>Examples</th>
<th>Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative Projects</td>
<td>Low</td>
<td>Sharing</td>
<td>Wikipedia</td>
<td>0</td>
</tr>
<tr>
<td>Blogs</td>
<td>High</td>
<td>Conversation, Sharing</td>
<td>Blogs</td>
<td>20005</td>
</tr>
<tr>
<td>Content Communities</td>
<td>Low</td>
<td>Sharing</td>
<td>Youtube, Flickr</td>
<td>14372</td>
</tr>
<tr>
<td>Social Networking Sites</td>
<td>High</td>
<td>Relationship, Conversation</td>
<td>Facebook, Twitter</td>
<td>1001478</td>
</tr>
<tr>
<td>Virtual Game worlds</td>
<td>Low</td>
<td>Group, Presence</td>
<td>World of Warcraft</td>
<td>0</td>
</tr>
<tr>
<td>Virtual Social worlds</td>
<td>High</td>
<td>Relationship, Identity</td>
<td>Second Life</td>
<td>0</td>
</tr>
</tbody>
</table>

**Fig 5.6 Selection criteria and media type**

The research findings share the idea of selecting social media for text mining which are *high on self-disclosure and comprise of building blocks conversation and sharing.* This is evident from the table above as highlighted in green the social media type Blogs and Social networking sites both are high on self-disclosure, comprise of building blocks conversation and sharing and in the three studies had the highest number of posts for the search terms.

To conclude, any organization interested in extracting consumer insights by text mining social media should consider blogs and social networking sites as their primary targets.

**5.3 Consumer Insights**

Before I share the use of consumer insights by an organization, I would like to share the result of the three search studies for consumer insights. In order to find consumer insights from social media the following is the size of the data set collected over the period of search.

<table>
<thead>
<tr>
<th>Social Media</th>
<th>Study 1</th>
<th>Study 2</th>
<th>Study 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogs</td>
<td>1956</td>
<td>18</td>
<td>18031</td>
<td>20005</td>
</tr>
<tr>
<td>Facebook</td>
<td>1727</td>
<td>293</td>
<td>252860</td>
<td>254880</td>
</tr>
<tr>
<td>Micro Blogs</td>
<td>4014</td>
<td>1040</td>
<td>741544</td>
<td>746598</td>
</tr>
<tr>
<td>Images</td>
<td>12</td>
<td>0</td>
<td>4335</td>
<td>4347</td>
</tr>
<tr>
<td>Videos</td>
<td>274</td>
<td>4</td>
<td>9747</td>
<td>10025</td>
</tr>
<tr>
<td><strong>Total Posts</strong></td>
<td><strong>1035855</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The data set interpretation steps are below:

I. Coding

For each study, the daily report evaluated for relevant information in the posts while using the definition of consumer insight as a reference. The daily reports had many irrelevant posts simply because the search terms were present in the post. Such posts, which did not have any information about needs, wants, desires or problems, were rejected. All the posts in which the online users mentioned any of the four characteristics captured using the following labels key phrase, emotion, wish list/pain point and finally suggested functionality.

II. Noting

Posts, which included a generic problem but could not be labeled under the categories mentioned above were made a note of but not reported in the list of consumer insights but as observations.

III. Abstracting and Comparing

The selected posts from the three studies were consolidated and compared for any recurring consumer insight.

IV. Checking and Refinement

The collected data checked once again for any oversights and the list of consumer insights consolidated.

V. Generalizing

The consumer insights that were common to the three studies consolidated and a generic statement prepared.

5.3.1 Gender based findings

I. Majority of users feels their mother is the best cook
II. Female users would like men to cook or cook for them
III. Male users ‘perceive’ cooking as time consuming and prefer women to cook

5.3.2 Generic Insights

I. Cooking is a time consuming “skill” need appliances to improve “skill”
II. Household users “wish” cooking would be easier, less energy consuming
III. Interactive appliances to make cooking “easier” by showing steps or sharing recipes
IV. Single consumers “wish” cooking was “fun”
V. Consumers “wish” appliances could help making “exotic” dishes easier
VI. Appliances to help “beginners” from avoiding “incidents”
VII. Appliances which are “child friendly”
By text mining social media, an organization can find consumer insights are evident in Fig 5.8. It is a new and a novel method and has a lot of scope of development. One of the key problem areas faced during the research is the amount of “noise” in the data collected. The number of posts with the search terms in them but of no value outnumbered the relevant posts by many times. The analysis done manually is time consuming and if automated analytical tool sets are, available text mining has the potential to supplement the existing, conventional marketing research methods. Below I share consumer insights from the text mining project 1:

**Consumer Insight 1  Cooking & Shopping**

*What to do when you get home and the fridge is empty? A trip through rush hour traffic and a scoot round the supermarket or an expensive takeaway seems inevitable.*

Online consumers have shared the problem of impulse buying as they are unaware of what they have in their kitchens, sometimes ending up with food products they already have and missing out on what they actually require.

I. **Need**  An awareness of which food items are required

II. **Want**  A reminder for which food items to buy

III. **Problem**  Lack of time and additional costs involved due to duplication of purchase

IV. **Desire**  Intelligent appliance that shares reminders about pre-defined food items
Consumer Insight 2 Cooking & Time

*There are days when I come home on an evening completely exhausted and these are the days when good intentions go by the wayside. I am too exhausted to cook.*

Online consumers have shared their concern about cooking being a time consuming activity. A few consumers have shared the idea of using slow cookers to ensure food is prepared over duration of time without any manual intervention once ingredients have been added.

I. Need  Less time consuming cooking process without much manual intervention  
II. Want   Food to cook quickly while consumers attend office or domestic chores  
III. Problem Lack of time coupled with physical and mental fatigue  
IV. Desire  Cooking appliance that takes longer than usual yet no manual intervention  

Consumer Insight 3 Cooking & Time-2

*How To Make Cooking Quicker: I hate cooking, I think it is a waste of my life to spend so long cooking something I will just eat in 5 minutes anyway, and I think it is difficult and complicated to cook decent meals.*

The second largest number of posts were complaints from online consumers about the time taken to cook even a basic meal as the time taken to consume that meal is much lesser. Though no solution was shared in the text mined data, it is important to understand these consumers are not in favor of pre-cooked meals.

I. Need Satisfy hunger  
II. Want To make cooking quicker  
III. Problem Time taken to cook meals by consumers who are not using frozen meals  
IV. Desire Cost effective, efficient and intelligent appliances  

Consumer Insight 4 Cooking & Health

*I want to lower my cholesterol, lose weight, loose fat and get in shape.*

Online consumers have displayed a keen awareness of cooking habits and health concerns. Phrases like energy boosting food and healthy food figure repeatedly in the social media text search. One of the most common statements was only if the home cooking appliance could share recipes for healthy eating.

I. Need High nutrition food, low fat meal plans without compromising taste  
II. Want Suggestions about alternates for food types to break monotony  
III. Problem Eating habits, unhealthy food, monotony  
IV. Desire Cooking appliance that suggests specific food
Consumer Insight 5 Cooking & Food Validity

According to a survey 15% of us serve “floor pie”, that’s food that’s been scooped off the floor, to our dinner party guests. Have you ever done that? Have you served up food you know is out of date?

Online consumers though aware their identities cannot be revealed have confessed to serving food that was past its validity. They lament the fact that no such device or process exists which can help them in monitoring the validity of the food products they buy.

I. Need Consumption of fresh food
II. Want Reminders about the expiry date of food products
III. Problem Food wastage, use of food products beyond the best before date
IV. Desire Monitoring of the best before date of the food items in their appliances

Consumer Insight 6 Cooking & Recipes

Lack of cooking skills recipe for disaster while delicious need not be difficult. What did you cook at your first dinner party?

Online consumers have repeatedly shared the lack of ease of finding recipes for cooking and imperative to note the search for recipes is partially driven by the appliance owned. Consumers search for food recipes which are appliance specific shares the lack of awareness about the possible uses of the appliance owned.

I. Need Variety in food consumption
II. Want Information to incorporate variety in consumption habits
III. Problem Lack of valid and reliable information about various types of dishes
IV. Desire Single source of valid, reliable and easy to access information on recipes

Consumer Insight 7 Cooking & Recipes-2

If I am cooking meals for only two people, I have had difficulty finding recipes that don’t make tons of food—or require tons of time in the kitchen.

Online consumers have repeatedly shared the lack of ease of finding recipes for cooking and as shared in the previous slide the search for recipes is also driven by the number of individuals. Consumers are specific in their search for recipes for cooking for single individuals, two and larger groups.

I. Need Variety in food consumption
II. Want Information to incorporate variety in consumption habits
III. Problem Lack of valid and reliable information about various types of dishes
IV. Desire Single source of valid, reliable and easy to access information on recipes

Consumer Insight 8 Cooking & Children

As a parent I am concerned about how to get my kids cooking and making healthy food choices. Healthy cooking is often difficult as most of us do not want to spend time planning and preparing meals that our families refuse to eat.

According to online consumers cooking pertaining to toddlers as well as growing kids presents a handful of unique challenges such as but not limited to time, taste and nutrition. Although no definite information is shared but the problem of cooking for kids has been repeatedly raised in text collected after text mining social media. Secondly, many parents share in social media their desire for cooking appliances that are child friendly.

I. Need Safety

II. Want Kitchen appliances with high level of child friendliness

III. Problem Accidents in kitchen involving children

IV. Desire Appliances that are child friendly

Consumer Insight 9 Cooking & Children- Science education

Teaching children to cook creates great opportunity to work to teach the school is a plus there is an added bonus. Children and parents bond and build memories.

One of the rare insights from consumers is about how to teach children cooking and help them understand about the scientific phenomenon involved?

I. Need Independence

II. Want Structured process that explains every step involved in cooking

III. Problem Unavailability of guidance during cooking stages

IV. Desire Intelligent and inter-active appliances

Consumer Insight 10 Cooking & Accessories

What is the optimum cookware form for cooking on an electric stove? It can be time for a significant upgrade from my scratched low-cost aluminum cookware from Wally Earth.

Online consumers have shared the lack of information available about compatible accessories for the cooking appliances they own. The time and energy spent on looking for information regarding accessories for cooking appliances is evident by the number of questions posted online.

I. Need Information

II. Want Compatible cooking accessories for appliances to increase efficiency
III. Problem  Lack of information from manufacturers about compatible products
IV. Desire  Single repository from manufacturers about suggested accessories

Consumer Insight 11 Cooking & Natural disaster preparedness

_In August 2005, over 15 million people lost power due to Katrina. Those who knew how turned to old-fashioned solutions, such as cooking over an open fire, to ensure basic needs for their families._

Online consumers highly likely from disaster zones have shared their problems with power outage and discuss cooking appliances which may run on alternate fuel or at least have the option to.

I. Need  Cooking solutions
II. Want  Cooking appliances with multiple energy sources
III. Problem  Dependence on conventional cooking appliances
IV. Desire  Hybrid cooking appliances

Consumer Insight 12 Cooking & Exotic Cuisines

_Cooking Chinese, More difficult than it seems! Americans Doing Italian Cooking!_

Online consumers have shown a high level of interest in cooking dishes from across the world but seem unsure if the current cooking appliances and accessories can help in cooking them. Queries about using the current functionalities of appliances to cook ‘exotic’ dishes are high in numbers.

I. Need  Variety in food consumption
II. Want  To cook cuisines from different regions and cultures
III. Problem  Lack of information about functionality and appropriateness of appliances
IV. Desire  Intelligent cooking appliances with functionalities to cook different cuisines

Consumer Insight 13 Cooking & Familial roles

_My Quest to be a better Mommy!_

Online consumers have shared an innate desire to excel in their societal roles. Cooking healthy food which is not only tasty but also less time consuming is the common feeling shared by consumers with children. Cooking thereby assumes significance in daily activities and hence the need to improve cooking skills, and appliances which enhance efficiency.

I. Need  A psychogenic need for social acceptance
II. Want  To provide food that is appreciated by all family members and friends
III. Problem  Changing social roles with cooking becoming gender independent

IV. Desire  Cost effective, efficient and intelligent appliances

Consumer Insight 14 Cooking equals Disaster

Cooking = Disaster, Cooking = Fail

On text mining social media the largest number of posts were by online consumers equating cooking to a disaster waiting to happen. The attitude towards cooking is that of a skill which is not only hard to acquire by learning recipes but also beyond the reach of most consumers, irrespective of gender. The fear of failure experienced by consumers is an opportunity for intelligent appliances, which guide the consumer through the cooking process or reposition products that mitigate the risk of failure.

I. Need  Success

II. Want  To be able to cook dishes that are edible

III. Problem  Lack of information about cooking process and unavailability of recipes

IV. Desire  An intelligent, inter-active appliance that guides the user

I share the use of these insights in the product development phase of Electrolux.

5.4 Consumer Insights and Product Development Process

Organizations in their product development process may have different methods to extract insights and depending upon the method used the insights may be applicable to either the ideation and concept stages or to product design and testing stage. In this section, I try to share the best possible scenario of using consumer insights from social media.

Theoretical Framework  Sawhney et al., (2005) in their paper collaborating to create: The Internet as a platform for customer engagement in product innovation have shared the nature of collaboration whether broad (high reach) or deep (high richness) may be incorporated in either the front end (idea) stage or back end (testing) stage of the new product development cycle.

After sharing, the consumer insights collected from social media in the previous section I attempt to map the text mining process along with the Electrolux’s product management flow to the theoretical framework (Fig 3.7) I shared earlier in section 3.4.

Research Findings  In order to either accept or reject the theoretical framework I would like to start with characteristics of Text mining. Fig 1.6 in Chapter 1 helps in understanding social media analysis as a high reach and unaided, unprompted research method. Text mining (sometimes-called knowledge discovery in text) refers to the process of extracting useful, meaningful, nontrivial information from unstructured text (Dörre, Gerstl, and Seiffert 1999; Feldman et al., 1998; Feldman and Sanger 2006). Using the above-mentioned characteristics of text mining this translates into a broad (high reach) method of collaborating with consumers. The reason behind not considering text mining as deep (high richness) is because the consumer insights are unsolicited and captured using an unobtrusive method. According to the framework,
a broad (high reach) method is most effective in the front end (idea) stage of product development.

Secondly, product management flow, the name of product development process in Electrolux has consumer insights as its starting point (Fig 4.4). According to the annual report 2010, all new Electrolux products are born out of the Group’s process for consumer-driven product development, which is a holistic process for managing products – from the cradle to the grave. This means that consumer insights are incorporated at the front-end (idea, concept) stage of the new product development at Electrolux. The research findings are in congruence with the framework as the real life practice at Electrolux is similar to the model shared by Sawhney et al.,

At this point, I would like to share another perspective for use of consumer insights which is outside the purview of the research framework. In section 1.8 I have briefly hinted at the relationship between marketing research and R&D function. Consumer insights extracted from social media may be shared with R&D so that a consumer needs perspective is available to a team that is primarily driven by a technological perspective.
6 Conclusion

In this paper, I propose a method by which organizations can listen to consumers’ ongoing discussions over the Web with the goal of converting the online discussion to consumer insights. Text mining of social media can act as sonar for listening to the electronic word of mouth. I use text mining to overcome the difficulties involved in extracting the wealth of online data that consumers generate.

I now discuss some of the academic and business implications, which arise because of this research. From an academic perspective, by text mining social media for consumer insights and interpreting my findings through the three lenses of strategy, social media type and use of consumer insights in the new product development process, has improved the chances of this novel market research method finding business application. Methodologically, too, my research explores and therefore demonstrates that text mining is a useful practice by sharing the extracted consumer insights to address research objectives. Leveraging a mixed method approach to analyze, this particular research has delivered a clearer view of the identifying appropriate social media and then text mining it to discover the electronic word of mouth that is reverberating in social media.

There are two main business implications arising from this research. First, information asymmetry persists online despite the advent of computer mediated communication and information exploration: in other words, there is still a gap between consumers’ needs and wants, and market offerings. As a result, consumers spend a fair amount of time in searching for information on products and features. Secondly, this research shows a window of opportunity for not only manufacturers but also advertisers to improve their marketing practice based on new insights into consumers’ needs and wants extracted from social media.

I would like to conclude my paper with a quote from Ray Poynter which aptly describes the purpose of my research “Why ask some of us when you can listen to all of us?”
7 Future of Text Mining

The immediate prediction is that text mining is here to stay. If there is a method of listening what customers are saying to each other about brands and services they have to be listened to, otherwise opportunities may be missed and potential problems not avoided. The tools for text mining are likely to continue to improve with one of the drivers being the research conducted in academic sector. The trend now is for more and more information on the internet to made easily available for example the recent changes to Facebook’s defaults and the open structure of twitter. This trend will make more data available and with improved tools will increase resolution of mining. However, it is unlikely that in the near future text mining will become cheap. The two key drivers are cost of dealing with very large amounts of data and the need for human analysis interpreting it. The application of ethical guidelines to text mining that are consistently employed in traditional research methods is problematic. In addition, the difficulty in analyzing elements such as sentiment and gaining informed consent from participants within the research environment. To effectively and accurately assess the impact of consumer conversations, it is proposed that a new set of internet research guidelines be created that fuse elements from both traditional and non-traditional research methods.
References


Barbier, Geoffrey and Huan Liu (2011) Data Mining in Social Media Chapter 12 Social Network Data Analytics


Bartl Michael, Steffen Hück and Stephan Ruppert (2009) Netnography research Community insights in the cosmetic industry


Cooper, RG and EL Kleinschmidt (1993). Screening new products for potential winners. Long Range Planning, 26(6), 74–81


Engestrom, J. (2005, April 13). Why some social network services work and others don’t Or: the case for object-centered sociality


Gilani, Hamza and Indira Sehic (2010) IT Outsourcing: A Strategic Partnership between buyer and vendor: Outsourcing of IT infrastructure within the deregulated Swedish pharmacy retail market


Glance, Natalie, Matthew Hurst, Kamal Nigam, Matthew Siegler, Robert Stockton, and Takashi Tomokiyo (2005), "Deriving Market Intelligence from Online Discussion," Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, 419–428


Kaplan Andreas M. and Michael Haenlein (2010) Users of the world, unite! The challenges and opportunities of Social Media Kelley School of Business, Indiana University


Kozinets, Robert V (2010) Netnography: Doing Ethnographic research online Sage


Mangold, W., Glynn and David J. Faulds (2009) Social media: The new hybrid element of the promotion mix Kelley School of Business, Indiana University

Merriam, S., 1998, Qualitative Research and Case Study Applications in Education. San Francisco: Jossey-Bass Publishers


Saguy, IS and HR Moskowitz (1999). Integrating the consumer into new product development.Food Technology, 53(8), 68–73


Xian-zhong, Xu & Roland Kaye, G Building market intelligence systems for environment scanning

Xun, Jiyao and Reynolds, Jonathan Applying netnography to market research: The case of the online forum Journal of Targeting, Measurement and Analysis for Marketing (2010) 18, 17–31

List of Figures

Figure 1.1: Levels of consumer insights
Figure 1.2: Innovative problem solutions in social media
Figure 1.3: Impact of social media
Figure 1.4: An Analogy of Consumer Input
Figure 1.5: The four buckets of data
Figure 1.6: Comparison with conventional methods
Figure 1.7: Opportunity in Social media mining
Figure 1.8: Spectrum of types of netnography
Fig 1.9 Project management approach to research
Fig 1.10 Overview of the thesis
Fig 2.1 Typology of mixed research
Fig 2.2 Basic types of designs for case studies
Fig 2.3 Thesis structure
Fig 2.4 Data Collection process in Silverbakk
Fig 2.5 Data Structure and Analysis
Fig 2.6 Online user
Fig 3.1 Classification of Social Media
Fig 3.2 Honeycomb of Social Media
Fig 3.3 Types of Strategies
Fig 3.4 Summary description of strategy types
Fig 3.5 Consumer led new product development concept
Fig 3.6 Collaboration Mechanisms Based on Nature of Collaboration and Stage of NPD
Fig 3.7 Framework of text mining Social Media
Fig 4.1 Electrolux product categories
Fig 4.2 Organizational Strategy
Fig 4.3 Focus of Consumer Innovation Program
Fig 4.4 Product Management Flow Part 1
Fig 4.5 Product Management Flow Part 2
Fig 4.6 Frequency of hits for the duration in social media
Fig 4.7 Distribution of posts per media
Fig 4.8 Tag Cloud Study 1
Fig 4.9 Frequency of hits for the duration in social media study 2
Fig 4.10 Distribution of posts per media study 2
Fig 4.11 Tag Cloud Study 2
Fig 4.12 Frequency of hits for the duration in social media study 3
Fig 4.13 Distribution of posts per media study 3
Fig 4.14 Tag Cloud Study 3
Fig 5.1 Social Media Types
Fig 5.2 Social Media posts of search terms
Fig 5.3 Social Media types and Building blocks
Fig 5.4 Social Media types, Building blocks and posts
Fig 5.5 Building blocks in single social media
Fig 5.6 Selection criteria and media type
Fig 5.7 Data Set
Fig 5.8 Consumer Insights