POLITECNICO DI MILANO

Scuola di Ingegneria dei Sistemi

Polo Territoriale di Como

Master of Science in

Management, Economics and Industrial Engineering

MANAGEMENT CONTROL SYSTEMS & SOCIAL MEDIA: TOWARDS A NEW FRAMEWORK

Supervisor: Prof. Michela ARNABOLDI

Co-supervisor: Ing. Deborah AGOSTINO

Master Graduation Thesis by:

Jacopo RADAELLI

Student Id. number: 760799

Yulia SIDOROVA

Student Id. number: 762778

Academic year 2011 - 2012

"You know, we don’t grow most of the food we eat. We wear clothes other people make. We speak a language that other people developed. We use a mathematics that other people evolved… I mean, we’re constantly taking things. It’s a wonderful, ecstatic feeling to create something that puts it back in the pool of human experience and knowledge."

Steve Jobs

ABSTRACT

In current fast changing reality managers and heads of companies have to face new challenges and difficulties, for making fast decisions, and improving their competitive positions. Social media is entered this business life becoming an important source of communication and information, making the market even more turbulent and sensitive respect to the past, but also creating new opportunities.

This new reality of the business is forcing to answer these calls, leading enterprises to implement new tools, seeking for new resources of information with the final goal to reach a competitive advantage through Social media. At the academic level this "explosion" of Social media has been studied by marketing and innovation scholars, while it has been almost neglected management control and performance management researches, notwithstanding the great potential of the Big Data stemming from social media. In this context our research aims at exploring the role of social media in Management Control System (MCS) addressing specifically these questions: how the social media could be included in MCS and what are the most efficient ways for such implementation? For which departments social media could provide the most important contribution to the control cycle? What are the current and perspective uses of the social media for management purposes? What are the main benefits and difficulties of using social media in MCS?

Empirically this research project adopted a two-level methodology developing a multi case study and an action research experimentation. This later was chosen as the most appropriate method to develop an innovative approach in the management control systems that should bring improvements into the system itself (confirmed by the use of this approach on the visible and reliable study) and follow ultimate trends in business, benefiting from the progress of the information technologies

Keywords: Social media, Management control systems (MCS), Key performance indicators (KPI), Metric, Social networks, Sentiment analysis, Opinion mining, Text mining, Big data, Web 2.0

Nella realtà attuale, in continuo mutamento, i manager e la componente direttiva delle società devono far fronte, nel prendere decisioni rapide e per aumentare la loro competitività, a nuove sfide e difficoltà.

I social media hanno debuttato all'interno di questa realtà di mercato diventando un importante mezzo di comunicazione e di condivisione di informazioni, rendendo il mercato più sensibile e turbolento rispetto al passato, ma creando anche nuove opportunità.

Questa nuova realtà del mondo dell'impresa porta a dover rispondere a queste nuove richieste, conducendo le aziende a sviluppare nuovi strumenti per cercare nuove risorse di informazioni con l'obiettivo finale di ottenere una posizione di vantaggio attraverso i social media.

A livello accademico questa "esplosione" dei social media è stata studiata da studenti di Marketing e Innovation, mentre è stata trascurata dai ricercatori di Management Control e Performance Management, non intuendo il grande potenziale insito nei Social Media. In questo contesto la nostra ricerca vuole porsi come obiettivo l'identificazione del ruolo dei Social Media nel Management Control System (MCS), rispondendo specificamente a questi quesiti: come i Social Media possono essere integrati in MCS e quale può essere il percorso più semplice per questa integrazione? Per quali rami aziendali i Social Media possono dare il maggior contributo alla catena di controllo? Quali sono gli usi attuali e futuri dei social media nell'ambito del management? Quali i maggiori vantaggi e quali le difficolta nell'utilizzo dei Social Media in MCS?

In concreto il progetto di ricerca ha adottato una metodologia a due livelli, sviluppando un Multi Case Study e una sperimentazione attraverso Action Research.

Questa è poi stata scelta come la metodologia più appropriata per sviluppare un approccio innovativo in MCS che potrà portare miglioramenti nel sistema stesso (ciò è confermato dall'uso di questa metodologia in uno studio visibile e attendibile) e seguendo gli ultimi business trend beneficiare del progresso delle tecniche informatiche.

Parole chiave: Social Media, Sistemi di controllo di gestione, Indicatori di prestazioni, Metrica, Social networks, analisi del sentiment, Opinion mining, Text mining, Big data, Web 2.0

ACKNOWLEDGMENTS

We would like to express our appreciation and thanks to prof. Michela Arnaboldi for giving us inspiration, courage and great opportunity to work on this challenging theme. We are thankful for her wise advices and support in what we were doing, together with the transmission of knowledge and detailed approach to the work. As well as we would like to gratitude ing. Deborah Agostina for important comments and hints in developing our research project and her friendly attitude.

Moreover, we would like sincerely thank Nextbit in the face of Federico Pagani for faith and trust in our team, and especially Fabio and Silvia, who were working with us during all this time. They have shared with us their critical business approach, helping to understand boundary of business necessities interrelated with the research.

Special gratitude we would like to express to Euro-Fer and Marco Mazza for this great possibility to work on real experimentation, approaching the innovative methodologies to meet company's objectives. This cooperation provide as a incredible experience, that we believe was of mutual benefit.

Personally, I would like to thank my family for transferring this startling approach of being curious about everything, moving forward each day of my life and being able to critically analyse things done. Thanks to all my friends in Russia, Italy and all over the word, this work would be impossible without your help, support, encouragement and strong believe in me (Yulia Sidorova).

Vorrei ringraziare tutta la mia famiglia, in particolare mio fratello, per ricordare che potremo sempre contare l'uno sull'altro, mia madre, che mi ha saputo trasmettere valori preziosi e importanti, e mio padre, che mi ha sempre dato supporto durante il mio percorso di studi. Merita un riferimento speciale l'amicizia, quella vera e pura, che molte persone mi hanno saputo dimostrare soprattutto nei momenti difficili, perchè è stata la più importante risorsa che ha permesso di raggiungere questo obiettivo (Jacopo Radaelli).

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY 11
1.1 Introduction and objectives 11
1.2 Methodology 13
1.3 Results 14
1.4 Structure of thesis 16
2 OVERVIEW OF SOCIAL MEDIA 18
2.1 Social Network 18
2.2 Acknowledgement 18
2.3 Web 2.0 19
2.4 Classification Of Social media 21
2.5 Enterprise 2.0 25
3 MCS 32
3.1 Conceptual definition of MCS 32
3.2 Evolution 34
3.3 Open issues and Future perspective 37
3.4 MCS & performance of Social media 37
4 CONCEPTUAL FRAMEWORK 56
4.1 Indicators 57
4.2 Uses 68
4.3 Framework overview 71
5 METHODOLOGY 73
5.1 Literature review methodology 74
5.2 Empirical analysis 79
5.3 Analysis of results 107
6 CASE STUDIES 108
6.1 Case # 1 108
6.2 Case # 2 109
6.3 Case # 3 110
6.4 Case # 4 112
6.5 Case # 5 114
6.6 Case # 6 116
6.7 Case # 7 118
6.8 Case # 8 120
6.9 Synthetic evidence: case studies 122

7 EXPERIMENTATION 126
7.1 Company overview 126
7.2 Euro-Fer requirements 126
7.3 Italian market: setting up 127
7.4 Italian market: data gatheing 128
7.5 Italian market: data analysis 131
7.6 Russian market: setting up 150
7.7 Russian market: data gatheing 151
7.8 Rusian market: data analysis 153
7.9 Company uses 170
8 CONCLUSION 172
REFERENCES AND BIBLIOGRAPHY 182
APPENDIX 193

INDEX OF TABLES

Table 1: Social media usage 15
Table 2: Acknowledgement terms 19
Table 3: Differences between WEB 1.0 and WEB 2.0 19
Table 4: Social media classification [Kaplan and Haenlein, 2009] 21
Table 5: Social software in the company [Haefliger et al, 2011] 29
Table 6: Open issues in Social software research 30
Table 7: Package of control elements 33
Table 8: Classification of indicators measuring Social media [Doerflinger, 2012] 47
Table 9: KPIs from Social media business objectives 52
Table 10: KPI used by Egg. Co 53
Table 11: KPI used by TMW agency 53
Table 12: KPIs used in commercial tools 54
Table 13: Summary of Social media metrics 55
Table 14: Social media KPIs classification 58
Table 15: Social media metrics performance 68
Table 16: Impact of SM KPIs on the control cycle 69
Table 17: Example of literature review analysis of the scientific & research papers 75
Table 18: Example of literature review of practitioners materials 77
Table 19: Summary table of the interviews 84
Table 20: Developed taxonomy of words for data analysis 91
Table 21: Example of the entry points analysis 92
Table 22: Stages of data analysis and instruments 97
Table 23: Case # 7 - KPIs' metrics for social networks 119
Table 24: Synthetic evidence of the interviews 123
Table 25: Taxonomy of Italian words 128
Table 26: Italian fonts selection 129
Table 27: Quantity of comments per font in Italian 131
Table 28: Taxonomy clusters for the keyword "ferro battuto" in IT1 font 132
Table 29: Taxonomy clusters for the keyword "acciaio inox" in IT1 font 134
Table 30: Taxonomy clusters for the keyword "ferro battuto" in IT2 font 137
Table 31: Taxonomy clusters for the keyword "acciaio inox" in IT2 font 139
Table 32: Taxonomy clusters for the keyword "ferro battuto" in IT3 font 141
Table 33: Taxonomy clusters for the keyword "acciaio inox" in IT3 font 144
Table 34: Taxonomy clusters for the keyword "ferro battuto" in IT4 font 146
Table 35: Data synthetic evidence for keyword "ferro battuto" 148
Table 36: Data synthetic evidence for keyword "acciaio inox" 149
Table 37: Taxonomy of Russian words 150
Table 38: Russian fonts selection 151
Table 39: Quantity of comments per font in Russian 152
Table 40: Taxonomy clusters for the keyword "ковка" in RU1 font 154
Table 41: Taxonomy clusters for the keyword "металл" in RU1 font 157

Table 42: Taxonomy clusters for the keyword "ковка" in RU2 font 159
Table 43: Taxonomy clusters for the keyword "металл" in RU2 font..... 161
Table 44: Taxonomy clusters for the keyword "ковка" in RU3 font 164
Table 45: Taxonomy clusters for the keyword "металл" in RU3 font..... 167
Table 46: Data synthetic evidence for keyword "ковка" 168
Table 47: Data synthetic evidence for keyword "металл" 169
Table 48: Italian market dashboard 170
Table 49: Russian market dashboard..... 171
Table 50: Cross-case analysis..... 174

INDEX OF FIGURES

Figure 1: Social media families..... 23
Figure 2: Social media families evolution..... 24
Figure 3: Relationships and links of Social media families 25
Figure 4 : MCS operational system 34
Figure 5: Performance management systems (PMSs)..... 36
Figure 6: Network information access..... 40
Figure 7: Online Social network information access..... 41
Figure 8: Stages of Social media measurement..... 47
Figure 9: Pillars of Social media measurement 50
Figure 10: Measurement methodology..... 60
Figure 11: Impact of SM KPIs on the planning stage 70
Figure 12: MCS control cycle 71
Figure 13: Timeline of project research 73
Figure 14: Indicators usage map..... 125
Figure 15: Example of crawled comments for Italian fonts 130
Figure 16: 10 most frequent terms for the keyword "ferro battuto" in IT1 font..... 131
Figure 17: Word cloud for the keyword "ferro battuto" in IT1 font..... 132
Figure 18: Clusters for the keyword "ferro battuto" in IT1 font 133
Figure 19: 10 most frequent terms for the keyword "acciaio inox' in IT1 font 133
Figure 20: Word cloud for the keyword "acciaio inox" in IT1 font 134
Figure 21: Clusters for the keyword "acciaio inox" in IT1 font..... 135
Figure 22: 10 most frequent term for the keyword "ferro battuto" in IT2 font 136
Figure 23: Word cloud for the keyword "ferro battuto" in IT2 font..... 136
Figure 24: Clusters of the terms for the keyword "ferro battuto" in IT2 font 137
Figure 25: 10 most frequent terms for the keyword "acciaio inox" in IT2 font 138
Figure 26: Word cloud for the keyword "acciaio inox" in IT2 font 138
Figure 27: Clusters for the keyword "acciaio inox" in IT2 font..... 139
Figure 28: 10 most frequent terms for the keyword "ferro battuto" in IT3 font..... 140
Figure 29: Word cloud for the keyword "ferro battuto" in IT3 font..... 141
Figure 30: Clusters for the keyword "ferro battuto" in IT3 font 142
Figure 31: 10 most frequent terms for the keyword "acciaio inox" in IT3 font 142

Figure 32: Word cloud for the key word "acciaio inox" in IT3 font 143
Figure 33: Clusters for the keyword "acciaio inox" in IT3 font 144
Figure 34: 10 most frequent terms for the keyword "ferro battuto" in IT4 font..... 145
Figure 35: Word cloud for the keyword "ferro battuto" in IT4 font..... 146
Figure 36: Clusters for the keyword "ferro battuto" in IT4 font 147
Figure 37: Example of crawled comments for Russian fonts..... 152
Figure 38: 10 most frequent terms for the keyword "ковка" in RU1 font 153
Figure 39: Word cloud for the keyword "ковка" in RU1 font 154
Figure 40: Clusters for the keyword "ковка" in RU1 font 155
Figure 41: 10 most frequent terms for the keyword "ковка" in RU1 font 155
Figure 42: Word cloud for the keyword "металл" in RU1 font..... 156
Figure 43: Clusters for the keyword "металл" in RU1 font..... 157
Figure 44: 10 most frequent terms for the keyword "ковка" in RU2 font 158
Figure 45: Word cloud for the keyword "ковка" in RU2 font 159
Figure 46: Clusters for the keyword "ковка" in RU2 font 160
Figure 47: 10 most frequent terms for the keyword "металл" in RU2 font..... 160
Figure 48: Word cloud for the keyword "металл" in RU2 font..... 161
Figure 49: Clusters for the keyword "металл" in RU2 font..... 162
Figure 50: 10 most frequent terms for the keyword "ковка" in RU3 font 163
Figure 51: Word cloud for the keyword "ковка" in RU3 font 164
Figure 52: Clusters for the keyword "ковка" in RU3 font 165
Figure 53: 10 most frequent terms for the keyword "металл" in RU3 font..... 166
Figure 54: Word cloud for the keyword "металл" in RU3 font..... 166
Figure 55: Clusters for the keyword "металл" in RU3 font..... 167
Figure 56: SM usage by cases 178
Figure 57: Euro-Fer change of SM usage..... 179

1 EXECUTIVE SUMMARY

This research project investigates the role of Social media within and for Management control systems. This chapter provides an overview on the whole thesis, illustrating the main elements of the research carried out: in section 1.1 we introduce the topic, its' relevance and main objectives of the work; in section 1.2 an overview of the methodologies used during the research is provided; section 1.3 presents key results for the research, followed by the detailed description of the thesis' chapters contents (section 1.4).

1.1 Introduction and objectives

In the fast changing reality of nowadays it is crucial to have appropriate instruments of management in order to react and face the challenges of the modern business. Actually any company right now cannot imagine its' internal and external communication without the usage of the internet, more generally without the usage of the telecommunication devices. Therefore, the management instruments should be able to fulfil certain requirements in order to take fast and correct decisions, motivating the company's personnel, as well as sustaining ongoing learning process and have positive impact on overall external accountability of the company.

Within this landscape, the social networks are playing now an increasingly role. EMarketer predicts there will be a massive 1.43 billion social network users in 2012, representing a 19.2% increase over 2011 figures. Of course at some point society is bound to reach a social media saturation point, but the current trend is upcoming and it will not arrive to this point in the forthcoming decade. Alongside personal usage, more businesses are taking advantage of the benefits social media sites can offer. A joint study by global management consulting firm Booz & Company and Buddy Media found that 96 percent of companies surveyed would be increasing their investments in social media. Advertising and promotions, public relationship and customer services were listed as the main uses or benefits but other uses such as market research and recruitment were reported. Pratically speaking social software concepts can be useful in the human resource management in order to improve the process for talent management and successfull planning.

The adoption of Social media inside the company is directly correlated to use of the social software in paticular to their contexts and company's environment, becoming for the management an ongoing task that involves user's role and adapts the strategy according to the nature of the work and activities [Leonardi and Barley, 2010].

Another critical issue is represented by the technology, that tackle both the competitive dynamics in industries and the structure of the organization. The costs to interact with users get lower using social software, in addition the competitive landscape in each sector is influenced by comments and recommendation on social network. Thus, social software can be both tool and mediator between company and end-users [Frey, Luthje, and Haag, 2011]. In this context the contribution are correlated to the empowering and restricting of the users, so another balance based on the power relations is to be considered because the design and implementation of social software can alter or disrupt organizational processes [Leonardi, 2007]. As last social software must take in consideration the grow and build of a community, particularly take decision upon leadership and boundaries. Here as well the choice is the outcome of a balance process the makes the community possible [Jarvenpaa and Lang, 2011].

The peculiarity of social application is generation of big amount of data that could be used for the companies benefits i.e. for external and internal activities. Such information could be a good contribution to the management control systems in order to support it's main goals: decision-making, external accountability, motivation and [Azzone, 2006] and learning [Kloot, 1997]. MCS being a "package" of elements that are implicated in the relationship between organizational context and effectiveness [Otley, 1980]. Starting from the history of the MCS and in detail the indicators evolution from the traditional financial to the innovative and non-financial, the usage of the social media information could be the next step in the KPIs development, helping to overcome and complement the current requirements' drawbacks of the system, such as timeliness, precision and long-term orientation [Azzone, 2006].

Others open issues for the researcher are to evaluate the performance of each component of the package and the interrelation between the components, taking in consideration the MCS implemented in the company [Malmi and Brown, 2008].

Summarizing all above, we have seen the great opportunity for the management control systems development together with the social media information through a new framework. We have analysed the role of social media MCS, focusing on decision-making. Therefore, aim of research was to answer the following questions:

- How the social media could be included in the MCS and what are the most efficient ways for such implementation?
- For which departments social media could provide the most important contribution to the control cycle?
- What are the current and perspective uses of the social media for management purposes? What are the main benefits and difficulties of using social media in the MCS?

1.2 Methodology

The empirical analysis is represented by two main methodologies: action research (experimentation study) and case studies approach (interviews).

The action research (AR) methodology was chosen as the most appropriate way in order to develop our thesis by creating an innovative approach in the management control systems that should bring improvements into the system itself and follow the ultimate trends in business, benefiting from the progress of the information technologies. We also thought that the action research is the most suitable methodology, as it is widely used for the social science and organizational development: "in these areas it has a particular niche among professional who want to use research to improve their practices" [Denscombe, 1998, p. 57]. Moreover, after the primary step of the literature review by making deep analysis more than 90 scientific and research papers, together with the overview of the practitioners materials we detected the room for the new framework implementation in the MCS, which by now was only partially developed in the marketing sphere and in the internal business processes. Furthermore, within the AR we have introduced the experimentation in order to gather information directly from the company where the new framework have been applied for the phenomenon observation (from making market analysis, sending proposals to consulting companies, career service of Politecnico di Milano, human resource departments of the big companies, we have chosen the Euro-Fer case, as most suitable one to implement the new framework by developing two projects for Italian and Russian market).

Thus, as a primary step, we have introduced the case studies (interviews) methodology for collecting and listening the people experiencing usage of the social media for internal and external purposes inside their companies, developing synthesized evidence table to identify and classify indicators, their measurement and possible uses. Last, but not the least.

Observation is one of the three elements of the triangulation model of multiple source of evidence. We have used archival data which were mostly represented by the "practitioners" materials review and taken into account with the certain degree of reliability, as such sources in our cases are not scientifically trustful, however reflecting much more clearly true state of the current business practices.

In order to perform all these tasks we were following the methodological principles, as well as we have developed together with Nextbit own methodology for the data analysis for the experimentation, together with the provided analysis of the instruments and tools used for data gathering. All this was done in order to insure the reliable and valid results of our research project.

1.3 Results

According to the 8 case studies (interviews) held with the significant players on the current market of technology, consultancy, manufacturing and etc., we have identified the main areas of the usage of social media information, as well as trend of the KPIs and their possible ways of measurement, that could be seen in table # 1 below:

Case #	Field	Metric Objective	Use Department
1	Manufacturing	Sales forecasting Benchmarking	Marketing
2	Beverage	Product promotion. Public awareness	Marketing
3	Telecommunication	Product co-creation Product promotion Customer service Recruitment Benchmarking	Marketing Human resource Sales Management
4	IT conglomerate	Product co-creation Product promotion Listening platform Customer service Recruitment	Marketing Human resource Sales Management
5	Consulting	Sales forecasting Listening platform Recruitment	Marketing
6	IT corporation	Product co-creation Product promotion Listening platform Customer service Recruitment.	Marketing Human resource Sales Management
7	IT agency	Customer service Sales forecasting Product promotion Listening platform	Consulting Marketing Human resource Sales Management
8	Data intelligence	Marketing Communication	Marketing Management

Table 1: Social media usage

The usage of the Social media is linearly correlated with the companies' level of the implementation of the technology and telecommunications inside, and is used as one of the comparative advantages, instead other companies are outsourcing such solutions or developing them inside, which are related to the significant investments.

As regarding to the experimentation study we have performed two projects with two different objectives: for Italian market the sustainable development and for Russian the evaluation of market and possibilities to enter. In order to give the complete overview of the analysis performed, we have introduced the following metrics to sum up information and achieve the objectives of the projects:

- Relevant sources; this metric show the number of sources in the web that talk in a relevant way about the a certain topic. Usually the topic is related to one or

more keywords employed for the research. This measure should be taken each three or six months.

- Sentiment polarity; this metric is the weighted mean of sentiment of the sources depending on the sentences in each source in a range of [-1,1] where a positive value indicates a positive sentiment and instead a negative one shows a negative pattern. This measure frequency depends on the kind of topic, but it is suggested to run it monthly or weekly.
- Relevant topics; this metric evaluate the most talked topics depending on the taxonomy defined across each source. The frequency of this measure should be taken weekly or monthly.

Overall results of this analysis provided the company with the detailed information derived from the main objectives of two projects and leaded to the development of the indicators and their implementation in the social media sources ("Facebook" and "Youtube" personal pages), as well as usage of other social sources to reach strategic aims and improve performance of the company.

In terms of the research results we have implemented a new framework, based on the case studies outcomes and theory background, giving good results. However, the action research methodology is requiring the reflection stage in order to move the research forward for the next level of the spiral, by means calculating the results of the action taken base on the analysis and its comparison with the similar decision made based on the traditional source of information.

1.4 Structure of thesis

The thesis is organized as follows. Chapter 2 is giving the profound overview of the Social media including the social networks, Web 2.0, Enterprises Web 2.0 and their main features, together with the open issues and future perspectives. Instead Chapter 3 provides the state of art of management control system and in particular the evolution of the system, it's main drawbacks and existing way of interaction with the social media. Chapter 4 is devoted to the development of the new framework, concentrating effort on the indicators and their measurement, as well as on the uses of such indicators in the business. Chapter 5 describes the methodology of action research and information triangulation used for the implementation and verification of the new framework. In Chapter 6 we have included all the case studies - interviews done in order to verify and

understand the current use of the Social media inside the companies providing systematized evidence of the case studies. In Chapter 7 we have described in details the experimentation done for Euro-Fer applying the new framework for the concrete company's objectives, with the colligating conclusions of such experimentation. Chapter 8 provides the overall conclusions of the research work done, outlining the aims achieved and future development, coherence with the theory approach and main findings in terms of the scientific and managerial fields.

2 Overview of Social media

The objective of this paragraph is to analyse the extant literature on social media distinguishing between its definition (social networks are just a declination of social media world), its evolution, its distinctive characteristics and the opportunities and difficulties offered to business companies by these tools.

2.1 Social Network

The common concept of network has been adopted in the web-sites environment providing services for the users. The "social network" allows to the end-users the following main features:

- Making of a public or semi-public profile within a bounded system,
- Generation of a list of user in which create connection (ties),
- Viewing and browsing the list of connections and public contents owned by other users.

The social networking provides several sources of data that represent information about behaviour of people. Often the term "social web" is used as synonymous in a wrong way, in fact in this way we are referring to the specific web technology or to the websites [Keenan & Shiri, 2009]. The theoretical basis of network data analysis and the structure of the group have major implications for the researchers in the social web [Keenan, 2008]. Usually the community created in the social web are discussing around specific topics making niche groups of users, in web context it becomes relevant the identity of the user and the level of influence that he have in a group.

The explosion of the Web 2.0 allows a fast growth of this technology that brought at the current situation, the one in which we are living and communicate.

2.2 Acknowledgement

The term "Social media" is usually employed as synonymous of social network or Web 2.0. It is crucial to use the right terminology in fact all this word belongs to the same world but they are addressed to different object. In the table # 2 it has been reported the main definition on the Social media topic:

Term	Definition
Web 2.0	*"Web 2.0 is a term coined in 1999 to describe web sites that use technology beyond the static pages of earlier web sites [O'Reilly]"*
Social Network	*"A social network is a social structure made up of a set of actors (such as individuals or organizations) and the dyadic ties between these actors. The social network perspective provides a clear way of analyzing the structure of whole social entities"[Wasserman, Stanley; Faust, Katherine]*
Social Network Service	*"A social networking service is an online service, platform, or site that focuses on facilitating the building of social networks or social relations among people who, for example, share interests, activities, backgrounds, or real-life connections. A social network service consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. [Boyd, Danah; Ellison, Nicole]"*
Social media	*"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]."*
Social Software	*"Social software is a tool that supports weak ties [Granovetter] and social production [Benkler, Tapscott] following the idea of egalitarism [Surowiecki] and mutual service provisioning [Vargo, Lusch]".*

Table 2: Acknowledgement terms

The literature in the following pages uses the word social media as general reference to the world including social network, social software and Web 2.0; while the other terms are used to refer to the specific scope. Particularly the dissertation aim to analyse information got from Web 2.0 in different contexts.

2.3 Web 2.0

The social technology has been spread worldwide, it can been said that is the natural development of Internet environment. The mass adoption of this kind of technology has brought to a rapid diversification of purposes and kind of usage among the different Websites.

WEB 1.0	WEB 2.0
DoubleClick	Google AdSense
Ofoto	Flickr
Akamai	Bittorrent
Domain name speculation	Search engine oprimization
Pulishing	Partecipation
Britannica Online	Wikipedia
Content management systems	wikis
Screen scraping	Web services
Directories (taxonomy)	Tagging (folksonomy)
Personal website	blogging
Page views	Cost per click

Table 3: Differences between WEB 1.0 and WEB 2.0

To get a fully understanding of this pattern we must come back to the born of the "WEB 2.0". The first definition comes out from a brainstorm discussion between Tim O'Reilly at MediaLive International in 2001, in which it was highlighted how the increasing of the internet bubble had generated a lot of new interesting applications and the last survived dot-com companies seems to save some common elements that allowed them to escape from the collapse. This situation was called turning point or better starting point of the "WEB 2.0". This revolution was founded on the paradigm "The web as a platform" (defined for the first time in October 2004), this new era is characterized by the user that make the web without any clear boundary, the new environment created is dynamic and allows the end-user to better satisfy his needs through participation and the sharing. The table # 3 above is summing up the main changes in term of services or application in the "new" web.

In this context the social tools allows to socialize through a wide variety of features such as private messages, photos, videos and music sharing. A first division can be done identifies two major group of sites: people focused and activity focused.

People-focused websites are based on the user-driven personal content used to socialize and communicate with other people, basically each person is in a community sharing information through a personal profile page. The most famous and populated network of this kind is Facebook with almost 1 billion of users worldwide, while other well-know websites are Twitter, LinkedIn, mySpace and Google+.

Instead the websites activity-focused are based on the web-site content, which is usually clustered by theme and it's often enriched by users comments and contributions. The specific content is uploaded by an user and goes from film making to the dating, actually these sites can contain also more user focused feature or incorporate in a direct way content from people-focused websites. It's been classified as activity-focused websites Youtube (video), Flickr (photos) and Delicious (bookmarks), but there is a longer list depending on the content.

The main purpose of this framework is the communication and socialization among people, this generate a huge amount of information that can be used in several fields if well selected. Hence, it's important to understand how the key elements of the social

software (users and contents) act, for example in terms of anatomy, influence and work as a system.

2.4 Classification Of Social media

The formal definition of this world is "Internet-based tool for sharing and discussing information among human", so actually the social media is something about people connecting, interacting and sharing. The web 2.0 boosts the websites that are linked with this domain, this created an huge proliferation of this kind of sites based on the same paradigm but with a different purpose.

The very first social media is more than 30 years old, in fact in 1979 Tom Truscott and Jim Ellis had created "Usenet", a worldwide discussion system in which Internet users were allowed to post public messages. Since that idea and with the spreading of internet connection among the people brought to the parent of the blog "Open diary" created by Bruce and Susan Abelson about 20 years ago.

The first element of classification is the degree of "social presence"[Short, Williams and Christie 1976], in terms of kind of contact that can be achieved such as visual, acoustic or physical, which influences the communication patterns. The second element is the "media richness" [Daft and Lengel 1986] that has got as main goal the reduction of the uncertainty and ambiguity of the communication getting the greatest amount of information in a given time period.

A second level of classification is represented by the degree of "self-presentation", in terms of creation of impressing and consistent image of own personal identity [Schau & Gilly, 2003], and the "self-disclosure", which is directly correlated to the consistency of personal information given with the image created.

The table # 4 shows the very first classification of social media based on this criteria [Kaplan and Haenlein, 2009].

		Social Presence/Media Richness		
		Low	Medium	High
Self- presentation/ Self-disclosure	High	Blogs	Social networking sites	Virtual worlds
	Low	Collaborative projects	Content Communities	Virtual game worlds

Table 4: Social media classification [Kaplan and Haenlein, 2009]

- Blogs: basically personal websites, they can be compare to personal diaries that summarize all the relevant information of the user life.
- Collaborative projects: enabling of simultaneous creation of contents by the users, the idea is that the joint effort can bring to better results. From corporate side it should be underlined that this kind of social media are becoming the main source of information.
- Social networking sites: applications that allow connection between users creating personal information profiles and using any kind of content.
- Content communities: sharing of media content of any type between users without the compulsory use of a personal profile page.
- Virtual social words: get the possibility to live a virtual life similar to the real one using avatars in a three dimensional replication of world environment without any specific rules to interact.
- Virtual game words: platforms that simulate three-dimensional environments using personalized avatars and setting specific rules to interact.

Another kind of classification is based on the type of tool that each social media provides, getting 10 different families [Cavazza, 2012] showed in the figure # 1.

Figure 1: Social media families

- Publish: constituted basically by wikis, blogs and personal journalism portals. Here the contents usually are trustfully, made by recognizable authors.
- Share: combining different kind of resources such as video, photo, link and music. It's likely to find also product's reviews and feedback.
- Discuss: formed by forums and all the platforms in which it's possible taking a conversation such as video forums, instant messaging and VoIP.
- Social networks: usually platforms that allow the creation of "linkages" between the users that populate the environment.
- Microblog: composed by smaller contents (micropost) exchanged among users.
- Lifestream: aiming to group a huge amount of people to socialize
- Livecast: constituted by platform for hosting broadcast contents.
- Virtual worlds: platforms that create a parallel universe with main character in it the people.
- Social games: formed by different platforms that allow to play simple game with other people.
- MMO(massively multiplayer online game): these are real online video games that support thousand of players simultaneously.

Recently some social media has become silos for information trustfully and usable personal information about the user, for these reasons some website accounts is become also the key to enter in most of the social media website. The precious accounts are formed usually social network such as Facebook, Twitter, Google+ or LinkedIN [Cavazza, 2012].

Figure 2: Social media families evolution

Hence, the information that a user decides to include in the profile of his account is crucial as well as the number of user that a social network has got. The more a social network is populated the more is important in the social media landscape, so it can become a linker. This power brings also the social network to cover other areas, in order to provide a unique and integrated service and of course get into new sector [Cavazza, 2012]. This move usually brings two main results more people that will join to a particular social network and so more power towards other social actors. The figure # 3 present the map of the services covered by the social media with the historical actors, the social network that succeeded in differentiation and competitors.

Figure 3: Relationships and links of Social media families

2.5 Enterprise 2.0

A company can use this new growing environment in order to improve the performances through costs reduction or revenues increase. Each firm has chosen to use the web 2.0 in different ways both in the processes and in the departments, exploiting the huge amount of information and the users networks generated. The very innovative part is the chance for people to communicate and interact in a very easy way, basically resulting in creation of networks [Niall Cook, 2011].

This opportunity brings a huge amount of information at "low cost" for companies, changing the perspective to get them from top-down to bottom-up [Stewe Boyd, 2006]. The tools adopted to managing the "social issue" is called social software and the main characteristics is to support the interaction inside and outside the company among individuals [Shirky, 2005].

Thus, social software tools has become very trendy topic from practitioners to researchers, in fact this instrument supports the management to improve the level of the output of people both inside and outside the form, and also have strategic implications [Haefliger, Monteiro, Foray and von Krogh, 2011].

2.5.1 Social software main features

Social software has been used in company for years but there is not a formal definition of best practice, but it has been formalize a paradigm that allow to describe the main characteristics [Mcafee, 2006; Dion, 2005; Mandini, 2011]:

- Open belonging, company boundaries are becoming more willing to absorb information and innovation from external parties such as end-users, partners and suppliers using ways of interaction and engagement.
- Sociability, the network of relationships has become a core asset of individuals that organizations are recognizing and support to enhance the positive effects on problem solving, thereby stimulating the creation of social networks expanded, opening and transparency in reporting and sharing of information and knowledge.
- Co-creation, promote the creation and spread of content, knowledge and information from the people, encouraging creativity and innovation regardless of hierarchical structure
- Emergent collaboration, enable emerging cooperation approaches to create cooperative settings in a fast, flexible way, even outside formal organizational patterns
- Flexibility, in order to respond to changes in the business, organizations are increasingly seeking to keep dynamics in organizational roles and processes in order to ensure greater speed in the change
- Virtuality, ability to access tools, information through rich and fast interactions, regardless of physical location and office hours, in order to gain productivity and enable forms of work organization more adaptable to changing conditions of the people.

The mix of the elements of this paradigm allows identifying a framework in which social software can influence the orientation of the company [Mandini, 2011]:

- Open Enterprise, focused on the expansion and opening of organizational boundaries
- Adaptive Enterprise, enable the possibility to make flexibility and reconfigurable process
- Social Enterprise, characterized by the new collaboration, knowledge sharing and organizational schemes.

In order to realize such new kind of enterprise the social software has to be employed in three different part of the firm: the organization, the information service or the processes. In this sense right strategy, depending on right analysis of the current and

future situation, is important in order to choose the orientation that best fits to firm needs.

2.5.2 Make reason out of social software

The dimensions previously defined can be linked to more general domains: strategy, technology and community. They can be analyzed form inside and outside the company [Haeflinger, 2011].

2.5.2.1 Strategy

Actually it is been proposed that the availability of critical information could bring to similar firms output in terms of offer and products [Sampler,1998].The social software allows to get a very similar simulation in fact the information are on internet and free for all people able to get them. The information are related to new product development, preferencies and substitutes products in community (social networks) that often are gathered pursuing of similar objectives.

Basically the creation of value is empowered because it facilitate interaction with end users and customer, making shared identity and learning for internal technology use [Lakhani and von Hippel, 2003; Kuk, 2006, Hertel et al., 2003]; the tools for this objective are evolving getting sophisticated, the email lists represent the old parents.

The second point is the level of integration of the external users in the organization of the company regarding the essential business process. The outcome is the creation of a user-centered business model [Hienerth, Keinz and Lettl, 2006] that manage both the community based platform and the firm based platform in order to get good results in the co-creation process [Javernpaa and Leng, 2011].

2.5.2.2 Technology

The user sees the social software as a tool that allows the creative expression and identity building in the online context [Schau and Gilly, 2003; Muniz and O'Guinn, 2001]. The personal comments of customers about experiences with products or company is boosted by the visibility and peer recognition, actually this can brings to create sub-cultures with specific vocabularies, creative expressions, and behavior [Kozinets, 2002]. The social software, the architecture of digital artifacts, and the rules used to collaborate influence users [Baldwin and Clark, 2006] and make links between the ICT system and the user behavior.

Tackling the issue of the technology from a strategic perspective, it is important paying attention to the motivation of user inside and outside the company. Social software can represent a mediator between the two parties performing broadcast search on the platform, this can brings substantial contributions, if it succeeds in attracting motived individuals with different knowledge. [Haefliger, Monteiro, Foray and von Krogh, 2011].

2.5.2.3 Community

Social software are crucial to the process of formation of online community, it is really easy to access and get a group of voluntary and qualified contributors who can be involved in a process of a firm.

To facilitate this practice can be used the leadership, even if it takes years of commitment and investments to be gained [Spaeth et al., 2010]; in addition the involvement of the firm can change the motivation of the members [Shah, 2006; Stewart et al., 2006]. It is needed a deep analysis to get differences and risks in the adoption of this practice especially in a corporate context [Markus, 2007; O'Mahony and Ferraro, 2007].It's been developed a model called of "emergent leadership" [Sutanto, Tan, Battistini, and Phang, 2011] in a framework that allows peoplw to interact and creates network ties. The model can predicts the percived leadership from the patterns of interaction and it can provide to company levers to communicate to user communities in a right way.

The second topic about community is the boundaries form to all the users that interact about shared purpose or interest [Ren et al., 2007]. The main risk is the "Knowledge leakage" [Hustad and Teigland, 2008], so it is need some kind of mechanism to protect and check the intelectual property. Unfortunatly the setting up of this practise is interpreted as concerning in loosing control over their work [O'Mahony, 2003], so it is important motivate in the right way the users. This objective take full priority respect to the leadership issue, which will be treated in a second moment. It's beeen taken an holistic perspective to manage the boundaries in platform-based online communities [Javernpaa and Leng, 2011]. To achieve community goals and create and innovate contents is needed integrated management of power, identity, competence and efficiency boundaries between the platform owners and the users who form the community.

In the table # 5 it is been summed up all the considerations introduced previously.

		View from inside the firm	View from outside the firm
Strategy	Value creation	Inviting and empowering customers to contribute to product development [Fuchs & Schreler, 2010]	Strategic interaction with other users and learning benefits [Kuk,2006; Lakhani & von Hippel, 2003]
	Value appropiation	Firm's differentiated involvement in communities, dual licensing, selecting revealing better innovation performance [Dahlander, 2007; Henkel, 2006; Stam 2009]	Availability and dissemination of assets under Open Source and Creative Commons licenses or appropriation by user entrpreneurship [Lerner & Tirole, 2005; Haefliger, 2010]
Technology	SS as a tool	Gaining access to creative users, utilizing their judgement and their know-how [Fuller, 2010]	Use of blogs and community participation for self-sexpression and identity building [Schau & Gilly, 2003; Muniz & O'Guinn, 2001; Kozinets, 2002]
	SS as a mediator	Platform- induced biases, gropus and user generated content and behaviour as "runtime effect" [Dellarocas and Wood, 2008; Shiriky, 2005; Cooke & Bukley, 2008]	Techonology architecture signal value and suggest assignment of tasks and specialization [Baldwin & Clark, 2006; Yamauchi, 2000; von Krogh, 2003]
Community	Leadership	Trade-offs between community founding and sponsorship, community leadership costly and complex [West & O'Mohony, 2005; Shah, 2006; Spaeth, 2010]	Central role of most achieved member of the community, social skills matter beyond technical savvy [Moon & Sproull, 2001; Raymond, 1999; O'Mahony & Ferrero, 2007]
	Boundary	Cultural differences as challenge, risk of knowledge leakage [Pauleen & Yoong, 2001; Hustad & Teigland, 2008]	Firm involvement makes a difference in terms of contribution and motivation yet firm recognition matters [Shah, 2006; Stewart, 2006; Jappesen, 2006]

Table 5: Social software in the company [Haefliger et al, 2011]

2.5.3 Open issues and future perspective

Pratically speaking social software concepts can be useful in the human resource management in order to improve the process for talent management and succesion planning; actually it can brings in health management systemsnew ways of storing, accessing, and locating patient data and of course can support the new produt development initiatives. The adoption of this tool is really depending on the contexts and the environment systems, so the management of social software becomes an ongoing task that involves user's role and adapts the strategy according to the nature of the work and activities [Leonardi and Barley, 2010]. Anyway regarding to each dimension there are open issues that don't still have a straight answer represented in the table # 6:

		Open Issue for strategy research
Strategy	Value creation	What are the successful conditions for successful value co-creation? What are the conditions for mutual buy-in of internal and external stakeholders? How can long-term relationships between firms and users emerge and sustain? [Burger Helmechen & Cohendet, 2011]
	Value appropiation	Are the generic strateies for firms to appropriate value from co-created assets? What are successful business models for entrepreneurial firms engaging in co-creation of value with users? How could information flows, altered by social software, open new opportunities for value appropriation? [Hienerth, Keinz and Letti, 2011]
Technology	SS as a tool	Are the optimal use patterns of social software in certain industries or for specific business processes? How can social software be used to catalyze organizational change? When is sourcing and when in-house development and adaptation optimal? [Denyer, Parry and Flowers, 2011]
	SS as a mediator	How can power realationships during implementation be made transparent? How can the organizational impact during the implementation phases be negotiated to proceed fairly? How can management best assimilate, respect and act on ethical concerns of users? [Frey, Luthje and Haag, 2011]
Community	Leadership	Can governance structures of user communities be emulated by firms? Can economic arguments buy influence and what roles do other values play? When is firm leadership optimal and when to defer to community leadership? [Sutanto, Tan, Battistini and Phang, 2011]
	Boundary	What are effective policies and strategy for access to communities and social networks? Which areas of business should best be involved with managing the changing community boundaries? Can different community platform ownership structures be designed and network positions be build and, if yes, which ones are desirable to whom? [Jarvenpaa an Lang, 2011]

Table 6: Open issues in Social software research

Regarding the strategy issue we can say that the value increase through the interaction with customers and users inside and outside the firm. It is important both protect knowledge assets and sharing knowledge with consumers and users, obviously these two actions are in trade off so it is crucial getting the right balance level [Jarvenpaa and Majchrzak, 2010]. The co-creating strategy may be focused on the creation of very active and loyal customers and users outside the company, allowing strategists to understand and internalized the two perspective (inside and outside the firm).

Second issue is represented by the technology, that tackle both the competitive dynamics in industries and the structure of the organization. The costs to interact with users get lower using social software, in addition the competitive landscape in each sector is influenced by comments and recommendation on social network. Thus, social software can be both tool and mediator between company and end-users [Frey, Luthje, and Haag, 2011]. In this context the contribution are correlated to the empowering and restricting of the users, so another balance based on the power relations is to be

considered because the design and implementation of social software can alter or disrupt organizational processes [Leonardi, 2007].

As last social software must take in consideration the grow and build of a community, particularly take decision upon leadership and boundaries. Here as well the choice is the otucome of a balance process the makes the community possible [Jarvenpaa and Lang, 2011]

3 MCS

The objective of this paragraph is to introduce the control system elements, particularly some of them can be affected by the recent history of external environment (it's important to try to use reliable variable to describe a system), in this context can be adopted social networks metrics to get more reliable performance and also try to solve some historical problems that is affecting MCS. Moreover it is described and listed the main ways to evaluate the performance in Social media. For each performance it can be assign a metric linked to each objective of the company.

3.1 Conceptual definition of MCS

The management control system (MCS) can be defined as some sort of navigator that enables the coherence between objectives and available resources in a company, compares different alternatives and measures if the actions taken are aligned with the goals [Azzone, 2006].

Nonetheless there isn't a unique definition for MCS, in fact in the years the concept is become broader and broader respecting to its role in the company. The following statements are the more relevant in literature to explain the nature of MCS:

- MCS as the systematic use of management accounting practices and other forms of controls to achieve certain purposes, a perspective that we follow [Chenhall, 2003].
- MCS as a "package" of elements that were implicated in the relationship between organizational context and effectiveness [Otley, 1980].
- Management control as "the process by which managers assure that resources are obtained and used effectively and efficiently in the accomplishment of the organization's objectives" [Anthony, 1965].
- MCS as the means used by senior managers to successfully implement their intended strategies [Simons, 1995].

The main two dimensions emerged from MCS design are the characteristics of information produced and the management techniques used. To get the point, this distinction is comparable in financial accounting between form (how the information is disclosed) and content (what information is disclosed) [Mak, 1989].

Actually the MCS is a package indeed, in fact it comprises tools that allow different level of detail relevance and orientation depending on how they are "packaged" [Amigoni, 1978]. The coordination and intention designed among the elements determine if the system is an MCS, instead in case of designing for different organizational units it will be considered as a package [Malmi & Brown, 2008].

A recent framework [Malmi and Brown, 2008] consists of five control elements, representing the package of systems: planning, cybernetic, reward and compensation, administrative, and cultural controls.

<table>
<tr><td colspan="7">Cultural Controls</td></tr>
<tr><td colspan="2">Clans</td><td colspan="4">Values</td><td>Symbols</td></tr>
<tr><td colspan="2">Planning</td><td colspan="4">Cybernatic Controls</td><td rowspan="2">Reward and Compensation</td></tr>
<tr><td>Long range planning</td><td>Action planning</td><td>Budgets</td><td>Financial Measurement Systems</td><td>Non Financial Measurement Systems</td><td>Hybrid Measurement Systems</td></tr>
<tr><td colspan="7">Administrative Controls</td></tr>
<tr><td colspan="2">Governance Structure</td><td colspan="4">Organisation Structure</td><td>Policies and Procedures</td></tr>
</table>

Table 7: Package of control elements

Firms employ MCS packages to meet information needs, they contribute to value creation producing material for decision-making, resource allocation, performance evaluation, control, and cost management [Cheanhall, 2003].

Though the classical MCS operational system is composed by three subsystems [Azzone, 2006]:

- Budgeting, it defines the ways through objectives and resources are linked to the responsible of each part of the company (BU, process, product line).
- Performance measurement, it collects information needed to measure the performances of different business units and it evaluates the value in each time period.
- Reporting, It compares the results with the objectives and provides information about corrective actions.

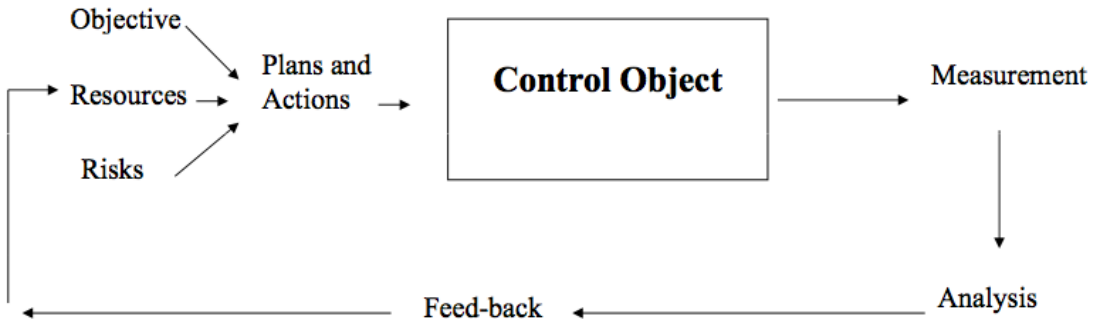

Figure 4 : MCS operational system

Simons [1995; 2000] stressed that both forms of MCS used are important and complementary; in fact one will perform the traditional role for monitoring and controlling [Otley & Berry, 1980] and the second one acts as a more elaborate form of monitoring performance through double loop learning [Argyris & Schön, 1978].

For both systems is crucial the concept of 'broad scope' information (BSI) [Chenhall and Morris, 1986]. This is referring to the characteristics of the information that is external, future-oriented and non-financial, taking in consideration as well financial, internal and historical information [Gordon & Narayanan, 1984]. BSI contrasts with traditional scope information, which is based conventional financial accounts, events and data generated inside the company [Mia & Chenhall, 1994].

The firms depending on information needs choose the techniques. Each choice will define the features of the package of control and also it will characterize the MCS package chosen [Gordon & Miller, 1976].

To conclude, management control system (MCS) help organizations to increase the probability that employees make decisions and take actions that are in the organizations' best interest [Chow, Shields, & Wu, 1999; Chenhall, 2003; Ferreira & Otley, 2005; Merchant & Van der Stede, 2007]

3.2 Evolution

The main components of MCS keep changing over time; in the past can be recognised two main frameworks that bring to current situation [Azzone, 2006]:

- First framework: here basically the goals of the MCS were the decision-making support and the internal accountability. The tools employed took as target the

measure of cost of production, monitoring the resources and evaluating the variance between standard and budgeting costs.

- Traditional framework: emerged in 80s, here it was introduced the MCS system that includes planning, measurement, variance analysis and feedback actions. For the first time financial indicators were introduce in the measurement of the performances, each business unit was in charge to report for their own activities and it was used both standard and historical references for budgeting.
- Modern framework: the one used nowadays, it is depicted with the used of Balance Scorecard, innovative financial and non financial KPIs for the measurement; reporting system extended also over the boundaries of the company, benchmarking and risk management techniques and an high integration with the internal information systems of the company.

Actually the figure that manages the situation keep changing during the years, in fact while the controller was essentially a engineer, with the traditional framework enlarged his competencies getting accountant background, while nowadays is called "business analyst" with competencies in process analysis, information system, risk management, internal control and business strategy. It sounds he should be a "know-it-all", instead the real important task is to understand the different languages of the organizational units in order to manage the operative management of the company [Azzone, 2006].

So both the definition and the actors of MCS has evolved over the years from the focus on providing formal and financially quantifiable information to one that includes a much broader set of information. It is composed by external information related to markets, customers, non-financial info related to production processes, predictive information, a broad array of decision support mechanisms, informal personal and social control [Chennal, 2003].

The modern techniques propose to get the "relevance lost" by traditional MCS techniques [Johnson and Kaplan, 1987] and broadening the scope and value-creation role of management accounting. Anyway the traditional techniques keep on being useful to organization, but giving other kind of information.

The Performance Measurement system (PMS) is viewed as the evolution formal and informal of mechanism, processes and systems used by firms to support the strategic

process and also the MCS package. Hence the PMS try to extend the model proposed by the MCS, suggesting ways to design and operative for control systems [Otley, 1998], avoiding issues raised by the adoption of the ideal model.

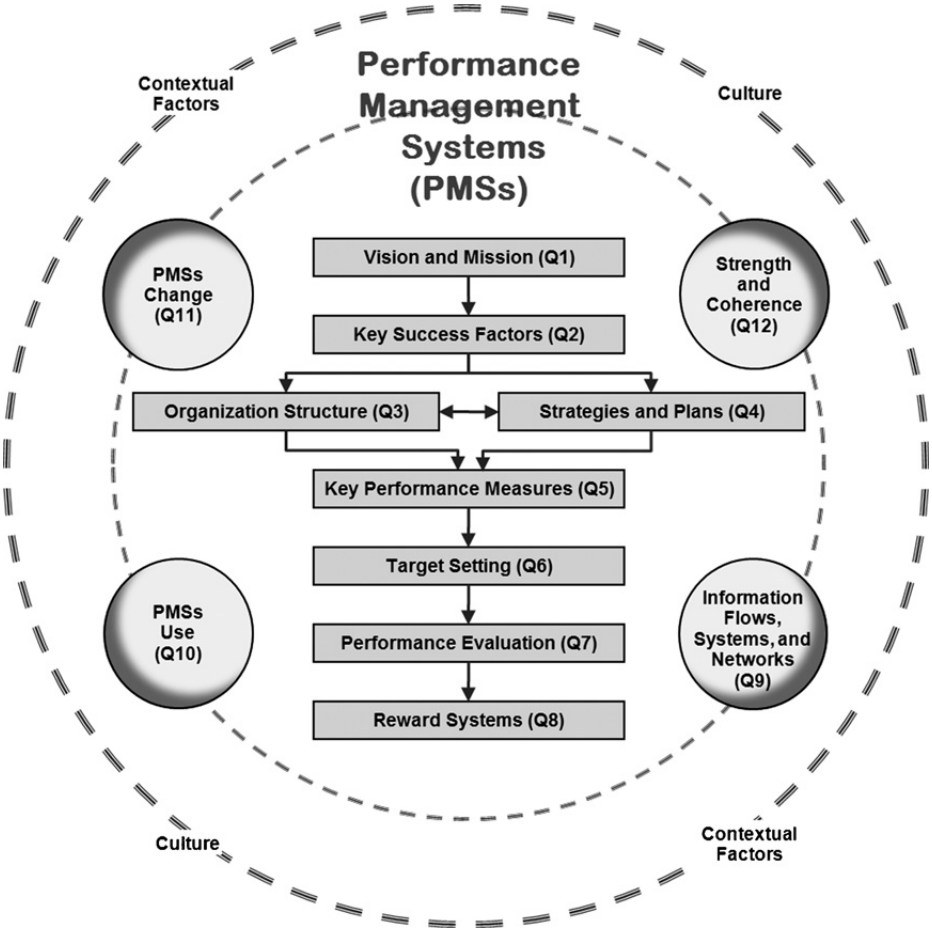

Figure 5: Performance management systems (PMSs)

Hence we use the term performance management system to encapsulate these more general processes, so the definition of a PMS includes not only the formal mechanisms, processes, systems, and networks used by organizations, but also the more subtle, yet important, informal controls that are used (Chenhall, 2003; Malmi and Brown, 2008).

Concluding, the PMSs act like supporting role for a broader range of managerial activities, including strategic formulation and implementation and on going management [Ferreira and Otley, 2009]. Actually with the learning and change facilitation role, PMS allows emergent strategy [Mintzberg, 1978].

3.3 Open issues and Future perspective

Many techniques developed are proposed to solve the limitation of information derived from traditional methods, such as being out dated, out-of-touch, too aggregated, internally focused, not fostering value creation, and not supporting effectively strategy implementation [Eccles, 1991; Johnson & Kaplan, 1987; Kaplan & Norton, 1996a; Stewart, 1991].

During the recent years scholars reported the negative "side-effects" deriving from MCS use. Individuals can be pushed to engage dysfunctional activities when they have the perception that it can bring to a personal gain or it can serve to their own interest [Birnberg et al., 1983; Simons, 1995]. The dysfunctional behaviours include smoothing, biasing, focusing, gaming, filtering, and illegal acts and were argued to be more prevalent in situations of low measurability, low verifiability, and low analysability of data [Birnberg et al., 1983]. They emerge because MCSs are used for controlling, evaluating, and rewarding performance [Simons, 1995] and because they may disrupt the extant balance of power [Tuomela, 2005]. These problems happen when there is a misalignment with goals between managers and organizations [Simons, 1995], in addition they could affect the performance of well-designed MCS as well in a negative way. The main issue in this case remains the distortion of information.

Others open issues for the researcher are to evaluate the performance of each component of the package and the interrelation between the components, taking in consideration the MCS implemented in the company [Malmi and Brown, 2008].

3.4 MCS & performance of Social media

The modern approach to performance measurement includes some non-financial indicators that are particularly useful to solve the problem of financial one: mainly the timeliness. There are two main kind of metrics related to the social media:

- Metrics describing use of social media (Network metrics)
- Metrics coming from social media (Social Metrics)

Both metrics can be used as indicators that can predict in advance some pattern or behaviours, so in some way complete performance measurement system of a company.

3.4.1 Network metrics

This kind of metrics measures the characteristics of the network, so this paragraph introduces which kind of features can be taken as object of control and then proposes the metrics to evaluate them.

3.4.1.1 Network data characteristics

Technically speaking "a network data consists of at least one structural variable measured on a set of actors"[Wasserman and Faust, 1994]. So it's very commonly used in the analysis of certain variables within an environment. This variable can be of two types: structural they are evaluated on a pair of actors, while the composition that analyze the attributes of an actor, such as the social and behavioral factors, that typify an individual.

The social networks are usually depicted as graph composed by nodes, that represent the member, and by ties, that show the relations among the member. The relations can be symmetrical; in case that both people have a mutual relation while it's asymmetric when just a person is relating to the other. Classical example of asymmetric relation can be the reading of a blog or an article, but it becomes symmetric if the reader will leave some comments or hints.

The network approach is not new at all, for all the past century this theme has been object of researches that defined some elementary properties, in particular the "Transactional content", the "Nature of the links" and the "Structural Characteristics".

The Transactional content is referring to kind of thing that usually two actors exchanges when are linked, basically it exists 4 types of content:

- Exchange of affect (liking, friendship)
- Exchange of influence or power
- Exchange of information
- Exchange of good

Usually the social network can take just a content or make them overlapping and the individual's position can vary.

The "Nature of the links" can be defined by 4 characteristics:

- Intensity, which represents the strength of the relation. Mitchell in the 1969 define this as "the degree to which individuals honor obligations or forego personal costs to carryout obligations", for an example if we want to define an appropriate indicator to measure that we can use indicator - "The number of contacts in a unit of time".
- Reciprocity, this is simply the number of the same intensities measurement that there is in a certain area. Clarity of expectation, in this case it's evaluated the degree to which people agree with the behavior in the relation with another person.
- Multiplexity, once identified the role of each individuals, it can be measure the degree to which a pair is referred to a multiple role, in a better way we can claim that a link is stronger the more role requirements are satisfied.

The "Structural characteristics" define a network divided in 4 level:

- External network: in here it is describe the relations with the external domain.
- Total internal network: it's composed by the relationship among the internal actors.
- Cluster within the network: this is a particular area in which similar actors are grouped considering them more close respect to others. The clusters can be of different types such as work group, emergent coalition (people that are grouped temporally for a limited purpose) and cliques (people grouped permanently for a broader range of purposes but informally).
- Individual as special nodes within the network: it can been identified special individuals that are more important than others. These people can link the focal unit to other areas within and outside the organization.

The bottom line of this analysis brings to say that a social network is a social unit with relative stable patterns of the relationship over time. This context allows the exchange of information through the ties among the actors, spreading them within the network. The economist sociologist Mark Granovetter in his paper called "The strength of weak ties" (1973), found out surprisingly that people prefer to get jobs that they notice through individual with not so frequent interactions respect to close people.

To explain the phenomena, we can claim that close people form a cluster (strong ties) in which they are all well-connected, this is composed usually by two people that a person use to talk, so as total there are three people. In this cluster the information are spread quickly to all the members, the strong of weak ties consists to transfer the information among cluster in which the interactions are not so frequent, so creating a sort of bridge. The speed of the connections represent an important charactiristic that influence their performance.

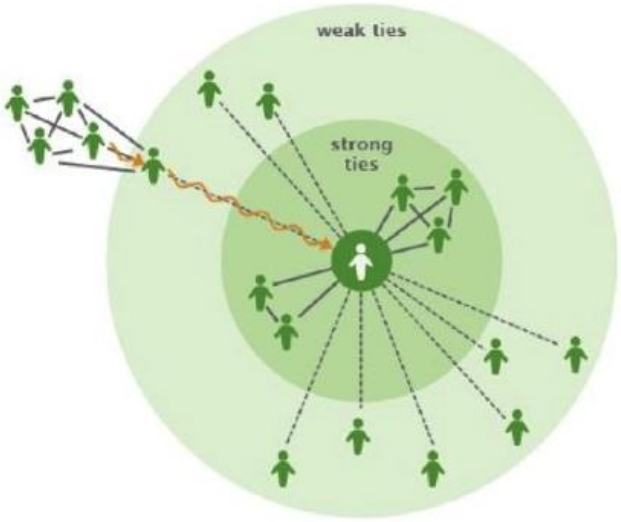

Figure 6: Network information access

This scheme can be adapted to the Internet environment where the information such as news is widely available. In this case there is a crucial factor in the spreading of the information that is represented by the commonalities among individuals, basically the people are in contact if they have something in common (homophiles). This concept bring to the conclusion that similar individuals seek the same kind of information, so also in this case the weak ties are crucial to spread the information all over the network.

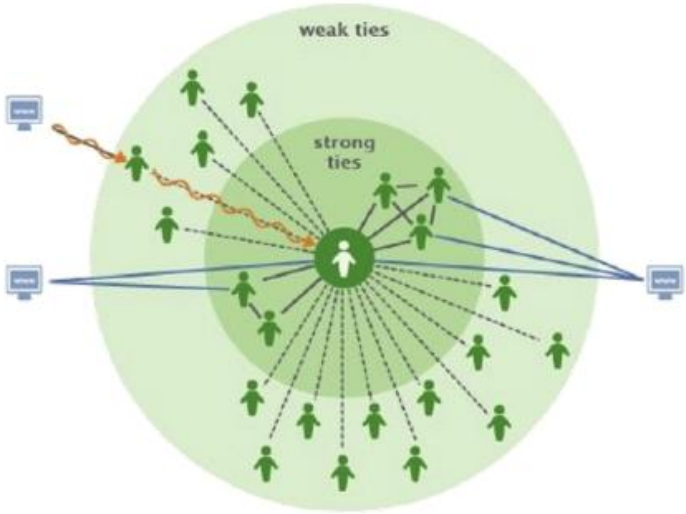

Figure 7: Online Social network information access

The social network is just a declination of broader world defined social media that includes characteristics of networks theories. The next chapter tackles the theme of Social media providing a clear overview of uses, classification and features.

3.4.2 Characteristics measurement

Once understand the dynamics of social media world, It can be recognized some patterns that can be always measured [Wasserman & Faust, 1994].

The first is the density that is the ratio between the existent ties and the all theoretical ones, so a high density network is a network where the member are very connected one with which others, hence probably there are few weak ties. On the opposite side a low-density network is characterized by several weak ties that allow the flowing of information, in fact with very few strong ties the weak ones become crucial both for the new knowledge and the common one.

The network centrality has different measures in the literature, actually it represents the importance and the power of a certain node in the graph, in a better way it could be defined as the power of influence other nodes [Wasserman & Faust, 1994].. This characteristic can be evaluated through:

- Degree of centrality: this is simply define as the number of ties of a member. The degree can be distinguished in "in-degree" and "out-degree". The first type expresses the number of the people related to a member, while the second one

shows the number of ties a person has to other people getting more chances for supporting and for reducing dependency.

- Closeness centrality: this considers all the shortest paths("geodesics") to all other members, so it doesn't take in account just the strong ties. This indicators is inversely proportional to the sum of all shortest paths.
- Betweenness centrality: this gets all the shortest paths to go through a specific node in relation to the other paths. This kind of indicator highlights the dependency of many on this member to connect the others.

There are different methods to study networks depending on the characteristics that we have mentioned before, so basically by the nature of the actors and the properties of the ties. The most common type of network is the "One-mode" that study a single set of actors, which can have of course the usual declinations in term of group and cluster, and the relations among them representing essentially the nature of the connection; in addition there are extra-information related to the actors called attributes. The "Two-modes" network usually is composed by two set of actors or a set of events; these are defined as the time joint affiliation, for example the clubs membership or attendance at social events make this characteristic critical in the information management [Wasserman & Faust, 1994].

Once defined so many characteristics that a network can have, it is important to monitor the structure, the usage of the network and the information flow activities.

The information flow studies aim to get what components are critical in the flow of the information, in order to monitor and trying to optimize them. As we mention before it's important to get what are the critical nodes so the members in the graph, trying to quantifying them. The centrality of the node is the core information that we want to get; in a better way we want to understand "the capability of the vertex of connecting other measure members" [Wasserman & Faust, 1994]. Several metrics exist for this purpose:

- Degree of centrality, measured as the number of direct ties that involve a given node [Dalay & Haahr, 2009]. The metric to measure the degree of centrality of a node p_i where $a(p_i, p_k)=1$ is computed as:

$$C_D(p_i) = \sum_{n=1}^{N} a(p_i, p_k)$$

- Closeness centrality, measured as the reciprocal of the mean geodesic distance $d(p_i, p_k)$ [Dalay & Haahr, 2009]. For a given node, where *N* is the number of reachable nodes, it is computed as follow:

$$C_C(p_i) = \frac{\mathrm{N} - 1}{\sum_{n=1}^{N} d(p_i, p_k)}$$

- Betweenness centrality, measured as the control over information flowing of a node between others [Dalay & Haahr, 2009]. If g_{jk} is the total number of geodesic paths linking p_j and p_k, $g_{jk}(p_i)$ is the number of those paths that include pi calculated as:

$$C_B(p_i) = \sum_{j=1}^{N} \sum_{k=1}^{j-1} \left(\frac{g_{jk}(p_i)}{g_{jk}}\right)$$

Usually the first two types of metrics are used for message transfer (package delivery) [Borgatti, 2011]. The disadvantage of this metrics is that are suited for bounded network, so the employment with large population is very difficult due to complexity of the network. To compute this kind of indicators is required the topology map of the net. To overcome this problem it has been introduce the "ego-network" that is a network consisting on a single actor, in this way we can evaluate the centrality of the node without a complete knowledge of the topology [Wasserman & Faust, 1994].

Other important measures to evaluate the information flows are the presence of strong ties in the network [Dalay & Haahr, 2009]. The main metrics adopted are defined as follows:

- Frequency, measured as the number of times in which a node *n* encounter the node *m* [Dalay & Haahr, 2009]. In the metrics *f(m)* is the number of times node *n* encountered node *m* and *F(n)* is the total number of encounters node *n* has observed:

$$FI_n(m) = \frac{f(m)}{F(n) - f(m)}$$

- Intimacy/closeness: measured as the amount of time the node has spent connected to a give node [Dalay & Haahr, 2009] where *d(m)* is the duration of the connection between node *n* and *m*, while *D(n)* is the total time that a node has been connected across all the nodes:

$$ICI_n(m) = \frac{d(m)}{D(n) - d(m)}$$

- Recency, measured as how recently a node has encountered another node [Dalay & Haahr, 2009] where *rec(m)* is how recently *n* encountered *m* and *L(n)* is the total time that *n* has been a part of the network:

$$RecI_n(m) = \frac{rec(m)}{L(n) - rec(m)}$$

- Longevity: defined as the time commitment into a social contact for a social tie [Dalay & Haahr, 2009].
- Reciprocity: evaluate if the valuable contact is recognized by members of the relationship [Dalay & Haahr, 2009]. This indicator can assume value of true or false.
- Multiple social contexts: the breadth of topics discussed is used to show the intimacy of the contact [Dalay & Haahr, 2009].
- Trust: this is used to evaluate the trust level of the information [Dalay & Haahr, 2009].

Usually the tie strength is identified by the composition of this set of social indicators, so giving them a different weight depending on the characteristic of the network.

It exists also some indicators that evaluate the utility of the nodes in the network; in doing so it is possible to rank the node in order to understand the better one to spread a certain kind of information. The component of the formula (Sim=similarity, Bet=betweenness and TS=Tie strength) are weighted and combined to compute the "Utility value".

$$SimUtil_n(d) = \frac{Sim_n(d)}{Sim_n(d) + Sim_m(d)} \; ; \; BetUtil_n(d) = \frac{Bet_n}{Bet_n + Bet_m}$$

$$TSUtil_n(d) = \frac{TieStrength_n(d)}{TieStrength_n(d) + TieStrength_m(d)}$$

Besides the indicators, it has been defined predictors which the objective to predict future connections using historical data. The new relations are sustained by the transitivity property, in which two people that have a lot of acquaintances in common is likely that get in touch [Liben-Nowell and Kleiberg, 2003]. It has been developed

this theory in order to create a model to assign a score to possible collaboration between two actors:

$$score(x, y) = | N(x) \cap N(y) |$$

where N is referring to the number of neighbors.

It has been defined other two metric to evaluate the score, the first is called "Jaccard's coefficient" and it's defined as follow:

$$score(x, y) = \frac{| N(x) \cap N(y) |}{| N(x) \cup N(y) |}$$

An other metric to evaluate the score is called “Adamic and Adar metric”:

$$P(x, y) = \sum_{z \in N(x) \cap N(y)} \frac{1}{\log |N(z)|}$$

These indicators are useful to predict the evolution of the graph and the new connections between nodes that will get in touch in the future.

On another side the directional relations are very important because evaluate the centrality on a different structural properties. The prestige is referring to a node that is object a extensive nodes, in other words the node that receives information [Wasserman & Faust, 1994]. This feature is not always measurable due to the different ways in which increase or decrease this factor, actually it is affected by several variables. Anyway it has been developed metrics that aim to evaluate this feature:

- Degree Prestige, measured as the number of the nominations or choices that a node gets. The idea is that actors who are prestigious tend to receive many nominations and choices [Alexander, 1963].

  $$P'_d(n_i) = \frac{x + i}{g - 1}$$

  The terms in the formulas this ratio gives us the portion of people that choose this node, so the higher is the ratio the higher is the prestige.
- Proximity prestige, for this measure it has been defined different metrics, which have the same objective, the identification of the influence domain of an actor [Wasserman & Faust, 1994].

- The average distance: $\sum d(n_j, n_i)/I_i$

  where *j* are the actors influenced by *i* , this method doesn't consider all the nodes not influenced. This evaluation depends on the size group.

- *Lin* definition: $P_p(n_i) = \frac{I_i/(g-1)}{\sum d(n_j,n_i)/I_i}$

  where the numerator represent the actor reached and the denominator is average distance between the actor. This indicator is a ratio that evaluates the prestige in terms of closeness and proximity.

  - *Hararay* definition: this measure takes in consideration also the "contrastatus" among the actors (total distance to *ni* of all the actors), so not only the prestige. This metric gives indication on the *net status* of an actor.
  - *Zeleny's sociation index*: that is measured as the difference of the average of the overall intensity of ties of the group and the number of choices made by the actor *i*.

All these metrics can be synthesize by the creation of a actor's rank who represent the list of the actors that will choose a certain node *i*. The function that represents the linear combination to create the sociomatrix is the following:

$$P_x(n_i) = x_{1i}P_R(n_1) + x_{2i}P_R(n_2) + \cdots + x_{gi}P_R(n_g)$$

The main problem in creating this matrix is that can have no finite solutions, so we should force constraints to make the system solvable [Tam et al, 1989].

3.4.3 Social metrics

The social metrics are quite a novelty; they come from the web analytics and CRM. It grown quickly from the needs to make sense of the huge amount of information and data activity on the Web generated by the social media. The measurement of this new world is challenging, sometimes even impossible; the effort required is very high this is way most of the company are limited in the use of such metrics [Chris Murdough, 2009].

The more common use is for the evaluation of the brand and the marketing campaign, anyway innovative companies is going to employ these KPIs not only in marketing, but also in the business processes management (social BPM) and HRM especially in the recruitment and benchmarking of the knowledge [Doerflinger, 2012].

Basically the social media measurement consists in three main steps:

- Monitoring
- Analysis
- Social Intelligence

The figure # 8 shows the maturity of the measurement, the higher level it will be reached the best will be the benefits gained.

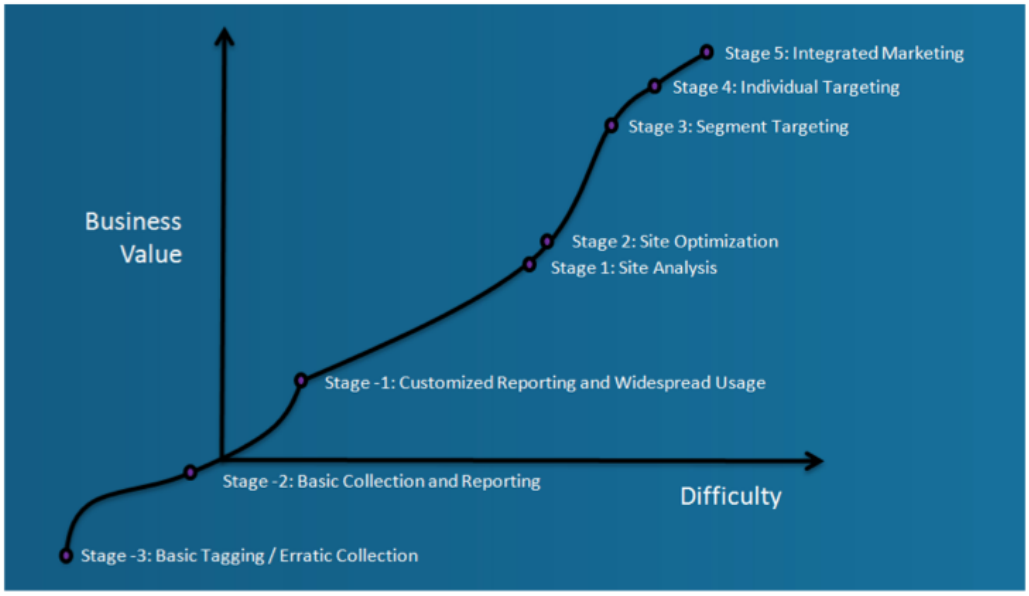

Figure 8: Stages of Social media measurement

Metrics form social are basically related to the Return on Investment (ROI). This metrics tells to company if their effort is giving good or bad results in social media and consequently the benefits gained. The measurement is divided in a classical way using financial and non-financial indicators [Doerflinger,2012] represented in the table # 8:

	Object of measurement	Driver
Financial Metrics	ROI	Employees retention Sales Cost reduction
Non Financial Metrics	Reach Discussion Outcomes	Foster dialog Promote Advocacy Facilitate Support Spur Innovation

Table 8: Classification of indicators measuring Social media [Doerflinger, 2012]

3.4.3.1 Financial metrics to measure social media

The measuring of ROI of social media is a quite complicated issue, just because there is not a unique metric but depends on the cases. The ROI is the concrete answer to what a

company has invested in before, so comparable to the effort done. The main issue is the fact that this KPI is not comparable between companies, due to such many different ways employed to evaluate it [Doerflinger, 2012].

We can provide some example of KPIs related to ROI of social media [Doerflinger, 2012]:

- Using as driver "Employee retention"
    - Investment costs in social business: Summing up all investments for social business including staff trainings, internal wiki and social media setup.
    - Saved costs through social business: Taking the revenue of social business like saved costs and increased revenue resulting from more engaged employees. Of course, the latter is nearly impossible to write down in numbers but indicators of increased employee satisfaction are working too, since there is a proven correlation between the satisfaction of employees and customers.
    - Saved costs through employee retention: It is cheaper to keep trained staff than to hire new members. These different costs are salaries, training and social business costs on the one side versus hire costs like job descriptions and interviews, trainings and lower productivity due to introduction phases etc. on the other side. Calculating the difference identifies the possible benefits.

  The final metric takes in consideration all contributions generated:
  ROI Employee Retention =
  (Revenue – Investment) + Employee Retention / Investment * 100

- Using as driver "Sales"
    - Revenue of online sales: Take the sales based on the respective campaign or social media project. It is important to figure out how exactly the origins can be tracked and consider the given aberration.
    - Investment of generating online sales: Sum up all related costs for social media efforts in terms of acquiring customers.

- Targeted Engagement: This is the summed up value of all new clients. Again, it is very unlikely to get hard numbers. In fact, this metric consists of a couple other metrics like Conversation Reach, Advocate Influence or Advocacy Impact

The final metric takes in consideration all contributions generated:
ROI Online Sales =
(Revenue – Investment) + Targeted Engagement / Investment * 100

- Using as driver "cost reduction"
  - Investment: Total costs of customer service for a given period of time like a month. This includes salaries for staff members related to customer service and all significant phone center costs.
  - Resolution Time: Average time span required solving inquiries. It is assumed that the duration of handling a certain inquiry will decrease due to social media and social business.
  - Satisfaction Score: The presence of a working customer service will result unsatisfied customers. This metric can also include sentiment, churn rate, audience engagement and active advocates (customers help each other).
  - Resolution Rate: Number of successfully solved inquiries.

The final metrics can change depending by the context of analysis. It is crucial to break down each contribution to the bottom line in order to estimate the benefits as precise as possible.

3.4.3.2 Non-financial metrics to measure social media

Non-financial metrics are not that consolidated, basically the literature suggests some methodologies to obtain KPIs starting from the analysis of the Social media linking it the goal of the company.

It is important to measure in some way the impact of social media engagement and not just the ROI since it takes time to be computed. So it is crucial and wise listen to social

conversations before engaging, before trying to drive conversion events. There are 3 main pillars to define metrics from social media [Chris Murdough, 2009]:

- Reach: monitor the quantity of mentions and quality of authors. It is important to evaluate the comments depending by the one who write. The criteria used to evaluate an author could be for example visit and comment frequency, so they can be based on the scale and activity of the fan base. Another important factor is where the comments are, in a better way which kind of social media they are in. This is important to understand which kind of messages and information can be got.
- Discussion: monitor of topic and sentiment of the comments. So first of all understanding which issues the comments are centred on and then evaluate if the sentiment is positive or negative. Of course the company can influence its own perception participating in the discussions and cultivating relationship.
- Outcomes: monitor of behaviour of people (e.g. leads, purchases). Here are measured the most relevant aspects concerning the website even if they are not the primary goals. In any case it is important to link this indicator the real value creation directly or indirectly resulting from social media effort. To provide a context to this social performance, it is crucial to employ the same lens of other primary referral sources (e.g. SEO, SEM, direct/bookmarked, e-mail).

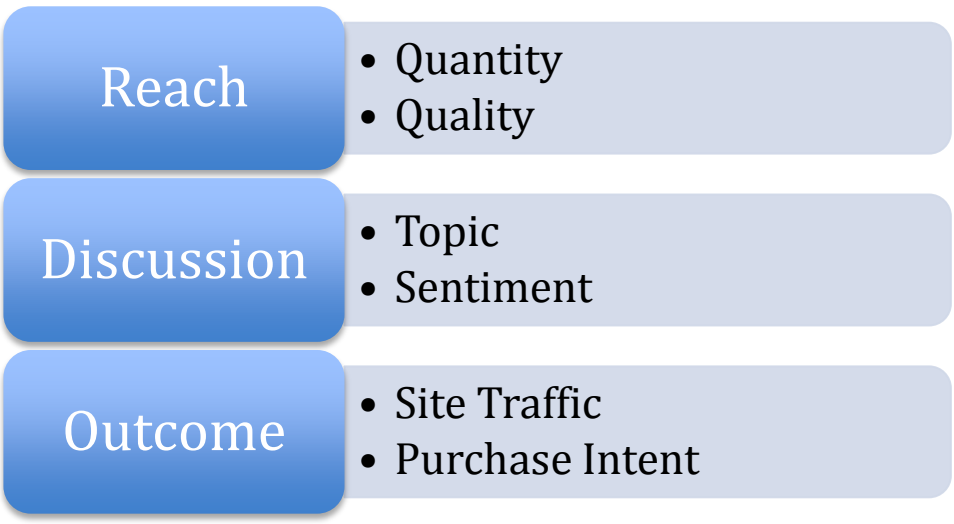

Figure 9: Pillars of Social media measurement

Once decided which indicators are suited for our goals, it is important to highlight the relationships of non-financial impact on financial ROI in social media [Olivier Blanchard, 2012]. The following approach can be adopted to tackle the previous issue:

- Establishing a baseline: Companies take their current sales as starting point and set time frames for easier control (monthly, yearly).
- Creating activity time lines: In addition to sales-based timelines, companies should define activity timelines as well. These have to be set in smaller intervals, preferably on a weekly base. This way, the impact of single actions can be measured.
- Looking at sales revenue: At the same time, it is necessary to keep an eye on the evolution of sales and on the number of transactions, also how many of those purchases originate online and how much revenue they generated compared to offline sales. This requires companies to already being measuring for a couple of months.
- It is good if sales are rising – but even better would be to know the concrete reasons for the rise. Tracking down purchase processes as well as possible helps drawing direct conversion lines. This can be done by using metrics like positive or negative sentiment, online store traffic (and its origin), website visitors, advocacies etc.
- By overlaying all these different patterns of the same given, time frame companies can start to investigate purchase paths and prove touching points. This could be interpreted as financial ROI, as the main formula is "Gain from Investment minus Cost from Investment divided by Cost from Investment".
- Finally, companies should start to take further advantage of their gathered data. As there is always room for improvement, it is recommended to experiment with data sets, combine them with other metrics and draw conclusions of potentially meaningful insights.

The process takes long time to be completed; anyway the non-financial metrics play an important role in the forecast of the ROI.

Besides the previous practices, there are other KPIs that are more innovative and they are theoretical proposal. In fact these metrics are suggestions which companies can

adapted aligning them with the overall changes [Robert W. Group, 2011]. The metrics written are linked to four business objectives:

- Foster Dialog, establishing talks and engagement individuals to bring back information to organization.
- Promote Advocacy, it is about getting closer to customers who have a certain influence and are better connected across networks than usual customers, in order to create advocate client (people who are connected to the company for certain kind of products).
- Facilitate Support, customer support has downright been revolutionised by social media in terms of speed, number of channels and empowered staff.
- Spur Innovation, learn about product needs, service requirements, innovation opportunities and even emerging market gaps.

Business Objective	**Key Performance Indicator**
Foster Dialog	• Share of voice • Audience engagement • Conversation reach
Promote Advocacy	• Active advocates • Advocate influence • Advocacy impact
Facilitate Support	• Resolution rate • Resolution time • Satisfaction score
Spur Innovation	• Topic trends • Sentiment ratio • Idea Impact

Table 9: KPIs from Social media business objectives

3.4.4 Practitioners contribution

Practitioners provide a great contribution in the definition of metrics to monitor social media activities generating several reports. The largest part of them is focused on marketing metrics.

Egg. Co divide the social media KPIs depending on the their nature, they can measure a qualitative or a quantitative goal. An important dimension is the scope of analysis that is directly linked to the kind of measure.

Kind of measure	Scope	Goal	Metric
Non-financial	Loyalty and Trust	**N** amount of positive conversations about your company or product per week	- Amount of positive comments sent to customers per week within a given time ceiling - Amount of conversations that started from the comment.
	Satisfaction	**N** amount of suggestions collected per month	- Amount of good suggestions that your company hadn't thought of.
	Authority	**N** Pagerank by a certain date.	- Amount of influential blogs linking to you.
Financial	Offline sales	**$N** monthly sales	- monthly sales
	Online sales	**$N** monthly sales	- monthly sales attributable directly to Twitter

Table 10: KPI used by Egg. Co

Instead Richard Pentin of TMW agency defines three main parameters to set KPIs:

- Type of Social media platform (Facebook, twitter, blogs, forum, etc…)
- 4 As (awareness, appreciation, action, advocacy) representing the Intent
- Soft metrics/hard financial metrics

In the table # 11 there is a sum up of all proposed metrics by intent, moreover the remaining criteria are applied depending on the needs of the company:

Intent	Metric
Cost per Impression (CPI)	Cost per unique visitor
Cost per Engagement (CPE)	Cost per subscriber
Cost per Lead (CPL)	Average purchase value
Awareness	# of tweets
Appreciation	# of comments
Action	Website referrals
Advocacy	# of tweets/retweets

Table 11: KPI used by TMW agency

Different tools employed for data analysis provide similar set of KPIs represented in the table # 12 below:

Company	Scope	Metric
Polvora	Awareness	Page Views
	Influence	Ratings / Rankings
	Engagement	Comments and Trackbacks
IAB	Social media Sites	Unique Visitors
	Blogs	Content Freshness and Relevance
	Widget and applications	Active Users
Razorfish - Fluent	Net Sentiment for the Brand	(Positive + Neutral Conversations – Negative Conversations) / Total Conversations for the Brand
	Net Sentiment for the Industry	(Positive + Neutral Conversations – Negative Conversations) / Total Conversations for the Industry
	SIM Score	Percentage of positive conversations for the brand in relation to the total conversations for the industry.
Radian 6	Activity and Engagement	Number of Members
	Revenue and Biz Dev	Speed of sales cycle
	Cost Savings	Issue Resolution Time
	Awareness and Value	Brand loyalty/affinity

Table 12: KPIs used in commercial tools

3.4.5 State of the art

The contributions described suggest metrics both from academics and practitioners. Usually the academic studies provide more general solution focusing on classical classification, while the practitioners are more focused on the scope and the objective of such metric. Anyway practitioner metrics can be classified depending on the financial or non-financial nature of the KPIs.

Nature	Academics drivers	Practitioners drivers		
		Egg. Co	TMW agency	Others [Polvora, IAB, Razorfish-Fluent, Radian 6]
• Financial	• Employees retention • Sales • Cost reduction	• Offline Sales • Online Sales	• Cost per impression • Cost per engagement • Cost per lead	• Cost savings
• Non-Financial	• Foster dialog • Promote Advocacy • Facilitate Support • Spur Innovation	• Loyalty and Trust • Satisfaction • Authority	• Awareness • Appreciation • Action • Advocacy	• Awareness, Influence, Engagement • Social media sites, Blogs, Widget and applications • Net Sentiment for brand, industry and SIM Score • Activity and Engagement, Revenue and Biz Dev, Awareness and Value)

Table 13: Summary of Social media metrics

4 CONCEPTUAL FRAMEWORK

The objective of this paragraph is to propose a new framework of MCS with the integration of information coming from the social media world. According to the literature review the management control system is used to provide information through numbers to pursue main goals: support decision-making, motivation, external accountability [Azzone, 2006] and learning [Kloot, 1997]. Therefore, we believe that the huge amount of information, coming from social media, could support MCS by creation of new metrics to measure company performances depending on the goal of analysis:

- Decision-making, in each phase of the control cycle, information from Social media can be useful to improve the standard procedure: from planning to feedback. It is important to translate into numbers information correlated to an object of control.
- Motivation, metrics coming from social media can be used internally to motivate people. In fact they give a new spectrum of performance which enables to drive human behaviour increasing flexibility inside the company.
- External accountability, new KPIs derived from the social media could be a good support of direct communication to the stakeholders the most important overviews on the social, environmental, financial, human resources and other internal company's activities, creating larger and immediate awareness, as well as possibility to benchmark with direct competitors.
- Learning, both KPIs and content analysis performed on the Social media can be a source of new concepts coming from the crowd sourcing and clients' complains. Moreover, the social business process management already introduced in some companies for more efficient interaction, together with new platforms for internal cooperation in multinational companies, such as "Jive" also gives additional possibilities to interact and learn from company's customers and suppliers in order to co-create new products and services.

The framework proposed aims to expand the decision-making purpose focusing on indicators and uses. In order to do this it is crucial to define a methodology to create indicators, standardizing the procedure of obtaining relevant data and analysing impact

of such information on the control cycle. Thus, the framework of analysis is composed by two main parts: the first is related to indicators, their classification and methodology allowing to get innovative metrics of measurements. While the second part could be entitled "Uses" and it is analysing the implications of usage of social media indicators in different stage of control cycle.

4.1 Indicators

The nature of these indicators is non-financial that is why they could be useful to support decision-making and overall the control cycle. This paragraph outlines the main types of indicators that could be used and their measurement methodology.

4.1.1 Types

As for all indicators also social metrics refer to a control object, which can be:

- A product/service
- Company
- Issue (e.g. design elements made of metal with Swarovski).

This latter has entered in the measurement "world" since the explosion of social media; companies have glanced the possibility to explore what people think of specific issues, which are related to their business area. For instance energy companies can monitor web thinking over nuclear energy in specific geographical areas, or automotive companies can explore which is the understanding of new (competing) services such car sharing.

According to the literature review findings and gap identification, we can classify the indicators following social media data characteristics:

- Nature of contents;
- Source of information;
- Network characteristics.

Nature of contents [Borgatti, ALGIM 2011] classifies indicators by the origin of the analysis and divide them into punctuated and text-drives:

  - Punctuated indicators are based on quantitative analysis of the number of users, sources or even opinions (n. like in Facebook fan page), but based on unambiguous information already available in the web.

- Text-derives are instead indicators based on the analysis of the text; this can be done through more traditional content analysis or sentiment analysis.

Sources of information [Hanna, Rohm, Crittenden 2011] that classifies data according to the accessibility of the information by the company itself dividing it into paid, owned and earned:

- Paid, to access to this sources is required a payment of a fee (Banners, SEM/paid search, in-store media).
- Owned, this data comes from company's own Internet pages (website of company, profiles in social networks, mobile applications).
- Earned, data generated by people talking on directly about your company (word of mouth, Facebook, blog, forum and Twitter).

Structural characteristics [Granovetter, 1973; Wasserman and Faust, 1994] classifying indicators according to two main features of social network: nodes and ties:

- Node: which include indicators based on the how many and what question within sources, but without looking at relationships;
- Ties: which instead analysis the quantity and type of ties.

		Node	**Tie**
Punctuated	**Paid sources**	Number of articles	Views
	Owned sources	Number of subscribers	Traffic generated (number of active actions)
	Earned sources	Number of users	Number of comments
Derived	**Paid sources**	Rank in search engines/Popularity	Frequency of keywords terms
	Owned sources	Rank of features	Frequency of most used terms
	Earned sources	Number of new ideas	Sentiment of discussions

Table 14: Social media KPIs classification

4.1.2 Measurement methodology

Usually the process of metric definition is correlated to the making of a scorecard in which data employed comes from internal process or annual report of the company

[Azzone, 2006]. The metrics are linked through critical success factor to the goals of the company that drives the creation of values.

The challenge in Social media metrics is to validate sources, obtaining good quality data making sense out of the loads of information present in the World Wide Web. So while the in the common process of KPIs definition it is used reliable sources such as financial statement; now the goals of the company should be linked, not only with the final metrics, but also with the source of data.

The process is composed by four main stages:

- Setting up, the more conceptual stage, needed to set the parameters for the analysis.
- Data gathering, this is the longest part in which are collected and validated the data for the following analysis.
- Data analysis, here is performed the three out of four analysis useful to get relevant and structured information
- Measurement, here the information is synthesized in indicators that can be employed to monitor and measure different types of performances.
- Monitoring and reporting, this is the standard part of the process in which are formalized reports and monitored the process.

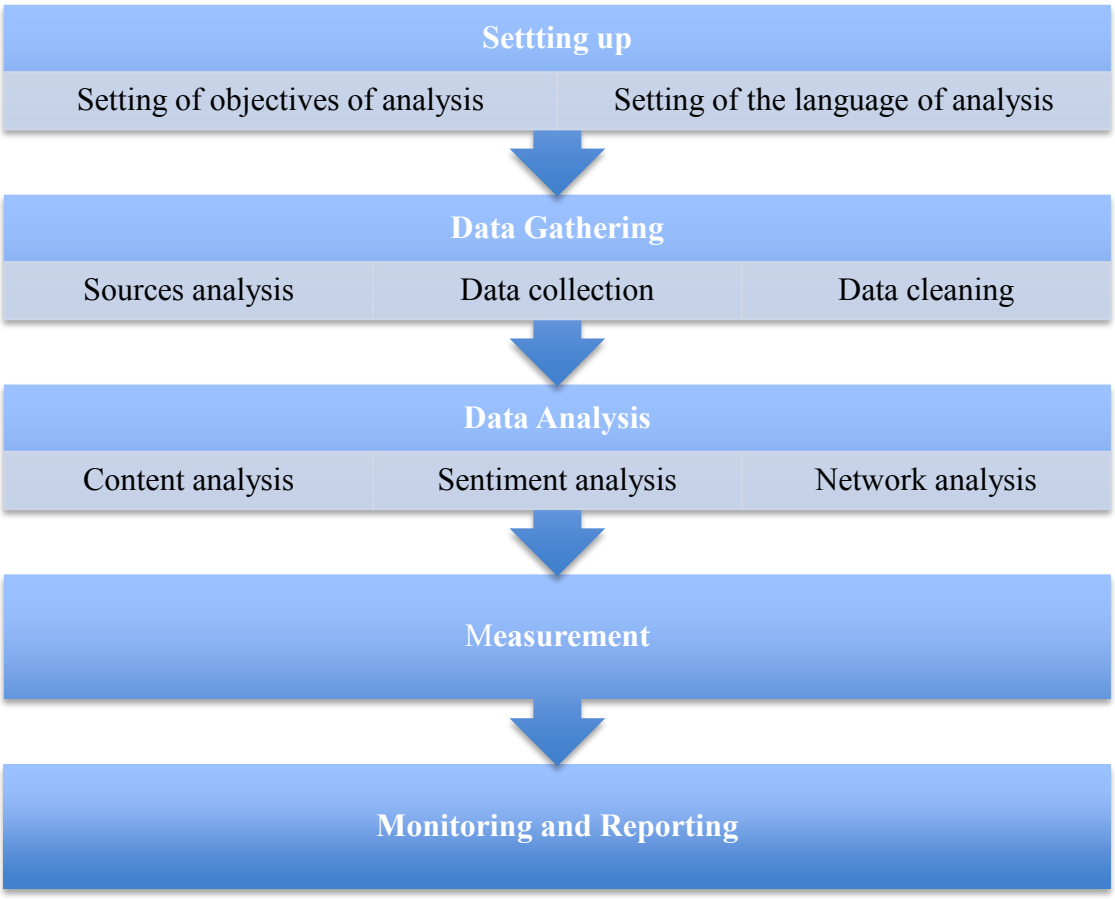

Figure 10: Measurement methodology

4.1.2.1 Setting up

4.1.2.1.1 Setting of objectives of analysis

This step is the very inception of the process, in fact here is defined what are the dimension of analysis, so which kind of information are useful to evaluate the object of analysis. The object of analysis has to be linked with a specific goal of the company in order to make sense of the available knowledge focusing on the issues that are more relevant. Of course can be for instance a product, a person, a service, a place so essentially whatever that for some reasons I want to control and monitor. The more important points related to the analysis to take in consideration are the following ones:

- "Where I can find the data"
- "What data I can find"
- "Who is dealing with the data"
- "How is the data"

Each point correspond to a specific type of analysis carried on in the following steps:

- Sources Analysis
- Content Analysis
- Network Analysis
- Sentiment Analysis

It is useful to support the following analysis through the creation of a taxonomy aimed to classify the concepts of the object of analysis using sentences, concepts and synonyms. In the taxonomy theoretical topics can be included in order to get the features of the object of analysis.

4.1.2.1.2 Setting of the language of analysis

This step is correlated to the choice of the language; in fact the different analysis is dealing especially with text data. This means that the person in charge to carry on the process has a good knowledge of the language chosen in order to manage with semantic rules, ways of saying and new words. In Social media world it is common face unstructured texts, which is supposed to be structured, and for each country and region there are specific semantic rules for the understanding of the meaning.

Usually the decision of the language is directly correlated with the object of analysis, for instance if I want to evaluate the data regarding the object "Financial Stock" in Germany, it is likely to choose German as language.

4.1.2.2 Data gathering

Data gathering is any process of preparing and collecting data. The purpose of data collection is to obtain information to keep on record, to make decisions about important issues, or to pass information on to others. Data are primarily collected to provide information regarding a specific topic [Australian Bureau of Statics]

4.1.2.2.1 Sources analysis

This is the very first analysis that corresponds to a report. The goal is to explore a direct practical application of the social media information in order to select the websites that can be useful for our object of analysis.

To get a starting list of the sources it should be taken in consideration the nature of website in terms of reliability, for instance there could be some data containing comments that are driven to influence people.

The second criteria is to report the quantity of information about our object of analysis if they are many or few ones, in fact it can happen that a website, focused on specialized topics, does not treat our object of analysis at all.

The results of this analysis should be a list of sources in which there is a first evaluation of "what is said" on the object of control made by the human being.

4.1.2.2.2 Data collection

The list of sources is the input for this step that consists in the download of data from websites, which can be performed manually or automatically (web crawling). Usually the web crawling is performed when there is a lot of information and the source code is well structured in terms of patterns.

Each source corresponds to a document containing the following data:

- "Text corpus", represent the core part of the collection and it is commonly named "comment".
- Title of "Text corpus" (usually name of the thread).
- Author of "Text corpus".
- Date of "Text corpus".
- Source of "Text corpus".

The output of this stage is a document of metadata for each source. Depending on the quality of the output some sources can be excluded.

4.1.2.2.3 Data cleaning

Before starting the next analysis, it is useful to perform some transformations aimed to improve the quality of the data extracted from the different sources, through the correction of inconsistencies, inaccuracies and missing values [Vercellis, 2009]. Some of the major shortcomings that are removed during the data cleansing stage are:

- Inconsistencies between values recorded in different attributes having the same meaning;

- Data duplication;
- Missing data;
- Existence of inadmissible values.

Once done the standard methods of cleaning, data needs some other transformations more correlated to the nature of the source. In our case it is likely to manage the following actions:

- Cleaning out the HTML language from the metadata
- Handling special characters and encoding format.
- Filtering out data that are not correlated to our research.

The bottom line of this stage is possibility to analyse each piece of data in the right way through a "Corpus" of documents.

4.1.2.3 Data analysis

The cleaned data allows performing different kind of analysis independent one from another one. So the following analysis can be carried on in parallel and depending on the needs can be performed all of them or just one.

4.1.2.3.1 Content analysis

Content analysis or textual analysis is a methodology that studies recorded human communications, such as books, websites, paintings and laws [Earl Babbie]. More specifically Content analysis is a summarizing, quantitative analysis of messages that relies on the scientific method (including attention to objectivity, intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and hypothesis testing) and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented. [Kimberly A. Neuendorf]

The goal of this analysis is to obtain the topic of the discussions considering the specific words, without taking into consideration the context. Information such as the frequency of words or repeated concepts comes out from this analysis changing its nature in the process from text data to numbers.

There are two main approaches used to perform this analysis: Top-down and Bottom-up. They are basically complementary, in fact can be employed both approaches getting

good quality solutions. Usually it is employed at the beginning the bottom-up that includes the following phases:

- Removing of words without a substantial meaning such as verbs and articles. Usually this steps aim to take just nouns and sometimes the adjectives.
- Removing low frequency themes and bias/errors of typing. The exceptions could be high frequency errors that can be considered in the analysis, for instance it's likely that the bank "Intesa-Sanpaolo" is discussed with the name of "Sanpaolo", "San Paolo" or "Intesa".
- Lemming to reduce the declination of the words in order to get the absolute frequency.
- Clustering of the words trying to get topic of discussion.

As it is been anticipated before, the main use of the top-down approach is to improve the first method, so the starting point is the outcome of the bottom-up analysis. Basically the top-down approach consists in matching of the taxonomy previously defined with the current results in order to get relevant results regarding the object of analysis.

The bottom line of this analysis is a report focused on the contents of the discussion, frequency of the words, words association and clusters of the discussions. The content analysis is carried out for each source to provide a good level of detail in the report, nonetheless it is good to have the overall results aggregated as well.

4.1.2.3.2 Sentiment analysis

While in the content analysis text is categorized by topic, which may involve dealing with whole taxonomies of topics, on the other hand sentiment classification usually deals with two classes (positive vs. negative), a range of polarity (e.g. star ratings for movies), or even a range in strength of opinion [Pang and Lee, 2008].

The essential issue in sentiment analysis is to identify how sentiments are expressed in texts and whether the expressions indicate positive (favourable) or negative (unfavourable) opinions toward the subject [Nasukawa and Yi, 2003]. Thus, sentiment analysis involves identification of:

- Sentiment expressions,
- Polarity and strength of the expressions
- Relationship to the subject.

These elements are interrelated. For example, in the sentence, "*XXX beats YYY*", the expression "*beats*" denotes a positive sentiment toward *XXX* and a negative sentiment toward *YYY*.

However, most of the related work on sentiment analysis to date has focused on identification of sentiment expressions and their polarities. Specifically, the focus items include the following:

- Features of expressions to be used for sentiment analysis such as collocations and adjectives
- Acquisition of sentiment expressions and their polarities from supervised corpora, in which favourability in each document is explicitly assigned manually, such as five stars in reviews, and unsupervised corpora, such as the WWW, in which no clue on sentiment polarity is available except for the textual content.

Besides adjectives, other content words such as nouns, adverbs, and verbs are also used to express sentiments. In principle, a sentiment expression using an adjective, say "*good*", denotes the sentiment towards its modified noun such as in "*good product*," and the whole noun phrase ("*good product*") itself becomes a sentiment expression with the same polarity as the sentiment adjective (positive for "*good*" in this case). Likewise, a sentiment expression using an adverb, say "*beautifully*," denotes the sentiment towards its modified verb such as in "*play beautifully*," and the polarity of the sentiment is inherited by the modified verb. Thus, sentiment expressions using adjectives, adverbs, and nouns can be simply defined as either positive or negative in terms of polarity. In contrast, as in the examples in the previous section such as "*XXX beats YYY*," the polarity of sentiments denoted by the sentiment expressions in verbs may depend on the relationships with their arguments. In this case, positive sentiment is directed towards its subject and negative sentiment is directed towards its object [Nasukawa and Yi, 2003].

Therefore, it has been manually defined sentiment expressions in a sentiment lexicon by using a simple notation that consists of the following information:

- Polarity positive (good), negative (bad), or neutral
- Part of speech (POS) Currently, adjective, adverb, noun and verb are registered in our lexicon
- Sentiment term in canonical form
- Arguments such as subject and object that receive sentiment from a sentiment verb.

In order to improve the accuracy of the sentiment analysis, it is important to properly identify the semantic relationships between the sentiment expressions and the subject.

To perform this analysis there are three main approaches:

- Rule based Approach, rules based on taxonomy (positive and negative words).
- Semantic Tree, seeks to link words based on semantic relations
- Entity Recognition, seeks to locate and classify atomic elements in text into predefined categories such as the name of persons, organizations, locations, expressions

The outcome of this analysis is the polarity of the documents for each source and then in an aggregate way.

4.1.2.3.3 Network analysis

The defining feature of network analysis is its focus on the structure of relationships, ranging from casual acquaintance to close bonds. Network analysis assumes that relationships are important. It maps and measures formal and informal relationships to understand what facilitates or impedes the knowledge flows that bind interacting units, viz., who knows whom, and who shares what information and knowledge with whom by what communication media (e.g., data and information, voice, or video communications. Network analysis is a method with increasing application in the social sciences and has been applied in areas as diverse as psychology, health, business organization, and electronic communications. More recently, interest has grown in analysis of leadership networks to sustain and strengthen their relationships within and across groups, organizations, and related systems [Serrat,2010].

While previously the text is the object of analysis, now the "author" (node) of a comments and the relationship created (tie) in a discussion become the main characteristics to study. This analysis aim, when it is possible, to create a profile of person that is linked to object of analysis and understand which are the more influence over group of people. It is hard to perform this analysis when the sources are blogs or forums where people use to register with nicknames, for instance it is likely find the same person in two different forum with different username.

4.1.2.4 Measurement

Depending on the object of analysis metrics can be developed in order to monitor the trend over time. It is crucial to choose KPIs that are more interesting for our purpose and aligned to the objectives previously defined. So depending on the goal and needs of the company, it is created a dashboard of indicators coming from the analysis and report performed.

This new metrics (social media metrics) can be evaluated using the classical requirements criteria of a MCS [Azzone, 2006] comparing them with the existent ones:

- Completeness, generally this feature depends on the set chosen because the social metrics take in consideration just some factors not all of them. This means that a set of indicators can compensate the drawbacks of the single KPI.
- Measurability, it is low because of the lack of standard methods and metrics, in fact the structure of the metrics is not standard that means that the final value can differ giving information extremely contrasting.
- Long-term orientation, they can be long term oriented depending on the object of analysis, if this is strictly correlated and coherent the main objectives of the company.
- Precision, usually they can predict with good accuracy the component of value creation, for instance the sentiment on a product can foresee the future sales with a good approximation.
- Specific responsibility, depending on the indicator, it can be good from the strategic to the operational level; of course it is crucial the link with the objectives and the actions to get them.
- Timeliness, they can be also better of the classical non-financial indicators, providing almost real time information.

	Value based	Accrual	Non-Financial	Social media metrics
Completeness	Very good	Good	Depends on the set chosen	*Depends on the set chosen*
Measurability	Low	Good	Very good	*Low*
Long term orientation	Very good	Low	Good	*Good*
Precision	Very good	Good	Low	*Good*
Specific responsibility	Good just at high level	Depends on the indicators	Good at operative level	*Depends on the indicators*
Timeliness	Very low	Low	Good	*Very good*

Table 15: Social media metrics performance

4.1.2.5 Monitoring and reporting

This phase aims to formalize the report and monitor the process regularly; in fact this phase is performing just once the process has been done for the first time. The process generates reports and measures the KPIs defined, allowing the company to analysis trends over time.

4.2 Uses

The major issue in the uses of this data are the reliability of the sources in terms of trustfulness and reputation. This two characteristics influence the accuracy of the metrics.

Even if these analyses are in someway taken in the company, the data are used only in the specific departments in which the analysis is performed. This is one of the main drawbacks of the current situation; in fact the huge amount of data provided by social media can be used for much wider scope. The entire company can take the benefits of this information, influencing not just the operational levels but also tactical and strategic level. The sharing of this data could be crucial in the managing of a company, in fact according with the nature of information can provide in some situation a more complete picture of the situation.

4.2.1 New Uses

The new metrics can provide complementary information that can be very useful to improve the performance of some parts of the cycle of control of a company and also for decision-making process. Particularly, the phase of planning and feedback can get

the greatest benefits of this information exploiting the high timeliness of these indicators.

The impact of KPIs can be evaluated into the four main stages of the control cycle, correlated to the main purposes of such information usage. Moreover the KPIs are grouped with the information purposes, as well as identifying the main differential with the classical approach and providing explicit examples.

	Purpose	Differences	Examples
Planning	Benchmarking	Complementary information that give an higher degree of precision	Number of "Fans" on the Facebook page Number of "Followers" in Twitter
	Scenario analysis Simulation	Introducing of new crucial variables in the creation of a model	Clusters of new ideas generated on SM Features selected according the polls
Measurement	Evaluation of characteristics of the node	Analysis performed basically on third parties sources.	Analysis of a Facebook fan page.
	Evaluation of pattern of information generated by relationship	Measurement metrics defined in more precise way.	Analysis of sentiment about some features of the product/service
Variance analysis	Trend of characteristics	Higher timeliness of information.	Monitoring number of fan upon time (daily report).
	Changing in patterns		Monitoring the polarity of comments related to the object of analysis
Feed-back	Actions on specific object (i.e. internal or external actors)	Corrective actions timely, sometimes anticipative in the prediction of a problem.	Create attractive contents to increase the number of fans and brand loyalty.
	Corrective actions towards model variables (internal purpose)		Improving the match of requirements

Table 16: Impact of SM KPIs on the control cycle

As regards, to overall impact of such indicators on company performance, the information is correlated to different components of NCF of NPV depending on the indicator chosen, for instance the analysis of likes will impact on the brand value. In addition the information is correlate to the risk and time of NPV as well, in fact it allows to mitigate the risk with good timeliness and gives complementary information to create a more reliable model.

4.2.1.1 Planning

This new kind of information can support the planning phase in three main processes:

- Simulation
- Scenarios analysis
- Benchmarking

Basically the planning phase is crucial to set objectives, manage resources and assess risks. The making of a reliable model is crucial to set targets and the presence of social media can support its creation, considering important variables that with traditional tools are evaluated in a rough way.

The planning is referred to each part of the company, so it is better off an integrated approach to link the strategy and the process management to social media.

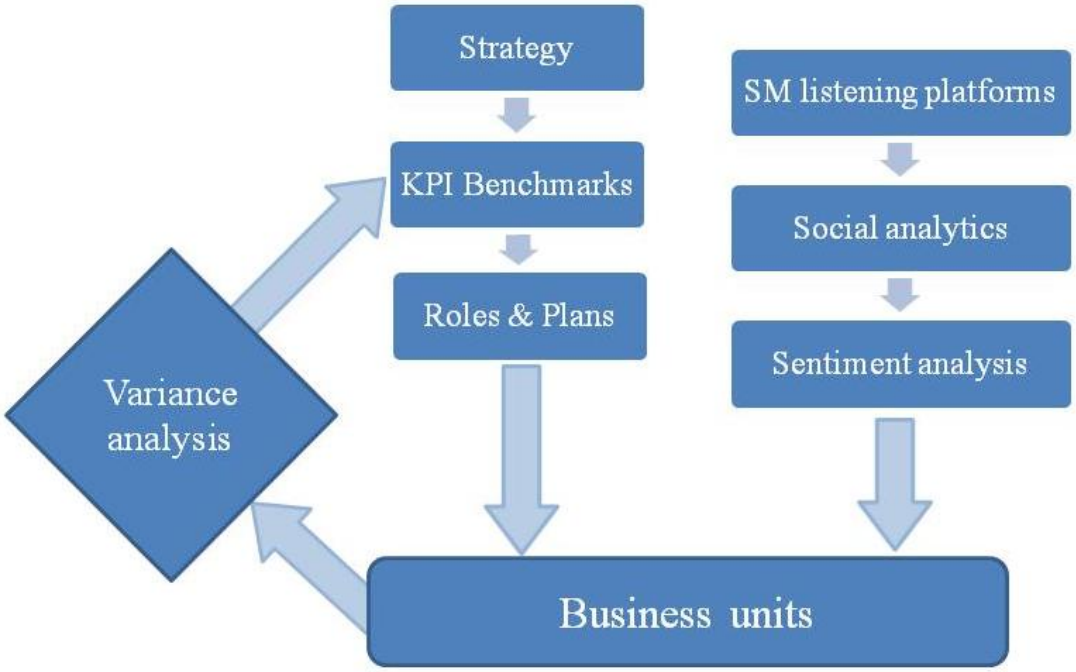

Figure 11: Impact of SM KPIs on the planning stage

Usually the planning stage employs the process of metrics generation entirely, the object of analysis is linked explicetely to the objective of planning. So the metrics chosen depend a lot on the different issues risen in each planning process, this means that can not be used the same indicators for each model due to different variables, organization and strategy of the company.

4.2.1.2 Feedback

This phase reports outcomes from operational activities generating corrective or learning actions. Here the new metrics can be useful to support the operational activities,

so in this case it will be performed sentiment, content and network analysis to get information about the trends.

While in the planning stage there are the situation AS-IS of social media to take strategic decision, now the actions represents more operational decisions influenced by information. The more timeliness are the information the more are effective the corrective and learning actions, so the monitoring of this metrics can provide useful information that must be complementary to the traditional measurement and variance analysis approach.

4.3 Framework overview

The framework proposed is related to the supporting of decision-making, which is one of the main goals of management control systems [Azzone, 2006]. Particularly, we try to extend the classical control cycle (figure # 12) of the company using data coming from Social media.

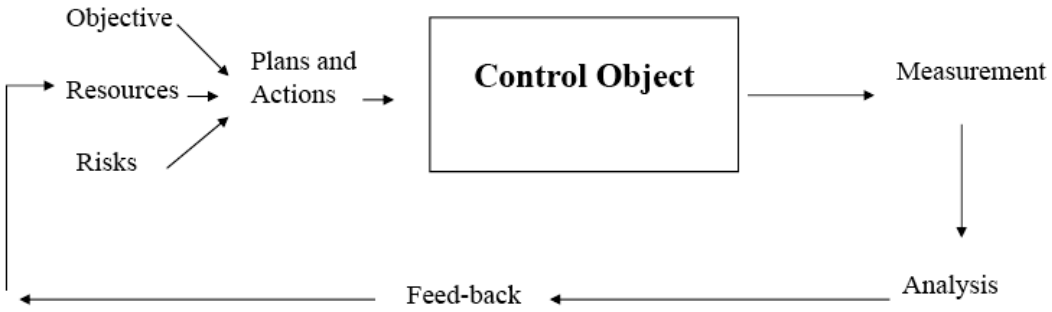

Figure 12: MCS control cycle

Indicators are the main representation of this information, synthesizing it in numbers. To define indicators the first step is to define a control object, than according to literature [Borgatti, 2011; Hanna, Rohm, Crittenden 2011; Wasserman and Faust, 1994], we can adapt three main characteristics of Social media data to classify the KPIs:

- Nature of contents, dividing indicators in punctuated and text-derived.
- Source of information, which can be from Paid, Owned and Earned.
- Network characteristics, orienting the measures on the tie or the node.

Thereon the types of KPIs are defined, that could be obtain from the desired sources, the framework provides a methodology of their measurement, evaluating the performances

of these indicators compared with the traditional ones. The main advantage of such KPIs is the timeliness of information, which provides a real time indication; while the greatest problem lies in the measurability. In fact the absence of a formal metric can generate discrepancies in the measurement appearing in false signals.

Last component of the framework is related to the uses of social media indicators in the control cycle. Actually the KPIs generated by Social media data can influence and support each phase of the cycle of control, in fact they can be useful in scenarios analysis, simulation and benchmarking and once analysed the variance also in the corrective actions.

5 Methodology

The objective of this paragraph is to represent the methodologies we have used for the development of our thesis: literature review, case studies, collecting voices of already existing practices and experimentation for the application of the newly developed framework (in particularly the process of choosing the Euro-Fer as the main case study, the work plan of the data collection and data analysis, methodologies used to perform main activities with our key supporter - Nextbit), together with the detailed analysis of the technology and instruments chosen for the experimentation performance. According to the definition of the methodology as a "guideline system for solving a problem, with specific components such as phases, tasks, methods, techniques and tools[1]", we have followed and combined numerous approaches to achieve the best possible results in our research.

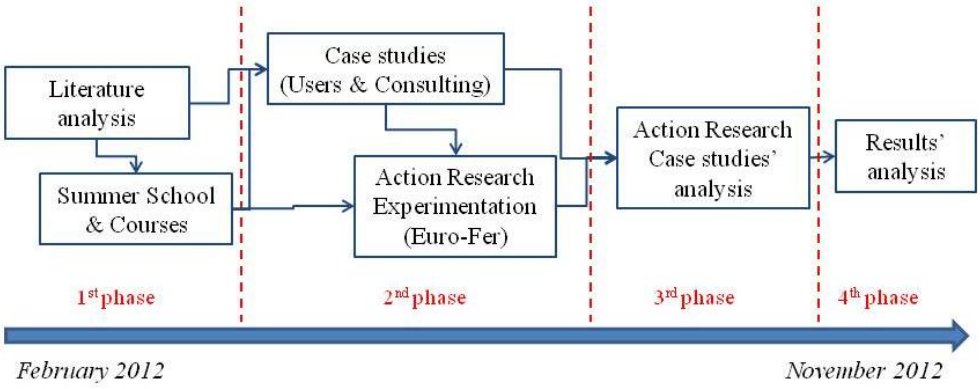

Figure 13: Timeline of project research

In general, the research process could be divided into to the 4 phases, as represented in the graph above (figure # 13): literature overview with the participation in the "Social business process management" summer school and courses devoted to the usage of "R-project", case studies and experimentation, in the third phase we have analyzed the results of the case studies and real experimentation, thus, in the forth we have made analysis of the research results and verification of the proposed framework.

[1] Irny, S.I. and Rose, A.A. (2005) "Designing a Strategic Information Systems Planning Methodology for Malaysian Institutes of Higher Learning (isp- ipta), Issues in Information System, Volume VI, No. 1, 2005.

5.1 Literature review methodology

Out research theme is widely spoken, however in the management control systems only several materials have been published by this time. Therefore, when we decided to make the literature review we have started from all the materials already written complementing them with the relative materials from marketing, strategy and informatics technology fields completing our list of the conceptual[2], primary[3] and secondary[4] sources. Facing these difficulties we have also introduced below the part of the anecdotal[5] sources - practitioners literature, which is represented by the materials and presentations published on the social media platforms and on the websites of the companies providing consulting services in the field of social media strategy, social media marketing and etc. Moreover, we have taken part in the summer school in July 2012 that was devoted to the theme "Social business process management", which helped us to enrich our thesis combining the scientific approach with the practitioners side and providing the newly published books devoted to the topic.

5.1.1 Literature review of the scientific & research papers

The literature review of scientific and research materials we have done following the systematic or "quantitative" approach and in total read more than 90 papers, of which major part is published in 2010-2012, around 20% is written from 2005-2010 and minor part is published earlier than 2005, proving growing interest both of researchers and business in the theme.

In order to collect the materials we have chosen the key-words:

- Management Control System
- Networks
- Social media
- Social Networks
- Enterprise 2.0

---

[2] Papers concerned with the description of theories/ideas and not empirism. Lucio Lamberti Research methodology seminars "Literature review and content analysis"

[3] Report by the original researchers of a study. Lucio Lamberti Research methodology seminars "Literature review and content analysis"

[4] Description or summary by somebody other than the original researcher (e.g. review article). Lucio Lamberti Research methodology seminars "Literature review and content analysis"

[5] Views or opinions that are not research Lucio Lamberti Research methodology seminars "Literature review and content analysis"

for searching articles and papers on the trustful search engines:

- http://www.jstor.org/
- http://scholar.google.it/
- http://www.scopus.com/home.url
- http://www.emeraldinsight.com/

The journals and were chosen based on the ABS ranking systems for ensuring the quality and reliability of the sources used for the literature review.

For the evaluation of our materials we have used together with the ranking system a table, consisting of four main columns: Author of the paper, Title of the paper, Topic and Why we consider this paper useful for our literature review. An abstract of the table # 17 is represented below:

Author	Title	Topic	Why
Eytan Bakshy Winter A. Mason Jake M. Hofman Duncan J. Watts	Everyone's an Influencer: Quantifying Influence on Twitter	Studies about the influence of the behavior of the people focusing on the relationship between individuals	Metrics that can build from social software: driver to motivate people
Thomas Wiedmann Manfred Lenzen	Triple-Bottom-Line Accounting of Social, Economic and Environmental Indicators - A New Life-Cycle Software Tool for UK Businesses	Sustainability and TBL accounts	Benchmarking with social, environmental and economic KPI
Atif Nazir Saqib Raza Chen-Nee Chuah	Unveiling Facebook: A Measurement Study of Social Network Based Applications	Social network and applications development	Metrics to measure the use of social network by using social games or social applications.
Elizabeth M. Daly Mads Haahr	Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs	Study on the information flow in the social network	Metrics to measure the use of social networks
Marjo Brander	Processing Social media Posts in Contact Centers	Thesis about social media as a tool as customer service	Metrics that we can built from social software
Stephen P. Borgatti, Daniel S. Halgin	On network theory	Networks foundations	Classification of networks
Michael H. Zack	Researching Organizational Systems using Social Network Analysis	Influence of social indicators upon organizational structure	Metrics that we can built from social software
Maldini	Enterprise 2.0: The natural evolution of Enterprise	List of paper for enterprise 2.0	List of paper for enterprise 2.0
Robert H. Chenhall	Management control systems design within its organizational context: findings from contingency-based research and directions for the future	MCS	MCS overview

Table 17: Example of literature review analysis of the scientific & research papers

5.1.2 Practitioners literature overview

Collection of the information from the web sources was following the same method as for the scientific and research materials. First, we have developed list of the key-words:

- Social media/Networks
- KPI in Social media
- Social media strategy
- Usage of Social media for management

Secondly, we have chosen the web-sources according to the Google ranking of these web-site in the search by means being relatively adequate and reliable:

- http://www.slideshare.net
- http://www.altimetergroup.com/
- http://www.radian6.com/

Afterwards, we have collected over 20 relevant and comprehensive materials, that were analyzed in the table # 18 represented below, this information from the presentation and materials was used for further research analysis and literature overview.

Author	Title	Topic	Why
Justin Kistner Nick O'Neill http://bit.ly/amafbJustin Kistner Nick O'Neill http://bit.ly/amafb	How the Big Brands do Facebook Marketing	Step by step explanation how to make and measure marketing campaign in FB	Measurement of the success of the campaign in FB
Declan Metcalfe	Social media Masterclass	All perspectives of digital "toolkit"	Social media - Engagement, positioning, developing networks, create/maintain relationships, getting feedback
Richard Pentin, Senior Planner,TMWRichard Pentin, Senior Planner,TMW	Social media - Engagement, positioning, developing networks, create/maintain relationships, getting feedback Social media - Engagement, positioning, developing networks, create/maintain relationships, getting feedback	Usage of Social data and creating framework for controlling company.	IAB framework Guidelines to defining KPIs for the Social Web Examples of KPI for different social networks (blogs, microblogs, app/widget, sharing, podcasting, SM advertising)IAB framework Guidelines to defining KPIs for the Social Web Examples of KPI for different social networks (blogs, microblogs, app/widget, sharing, podcasting, SM advertising)

Sean Moffitt, President & Chief Evangelist– Agent Wildfire Inc.	Monitoring and Measuring Conversations	Introduction of Social media into the company structure	Measurement and management of the social media Social media Value Measurement and management of the social media Social media Value
Brad Little Head of NM Incite, EMEABrad Little Head of NM Incite, EMEA	The Value of Social media	Explanation of social media value and usage for company purposes	Measurable impact of using social media and other web 2.0 technologies
TarushijioTarcízio Silva Directorof Research and Development PaperClioTarushijioTarcízio Silva Directorof Research and Development PaperClio	Social media Metrics	Metrics and Social media: Elements, Tolls and approaches	Detailed set up of metrics. Examples of usage of different tools to analyze social data in order to manage and achieve company goals, strategy.
Steve Latham Spur DigitalSteve Latham Spur Digital	Measuring Value from Social media	Measurement of value and engagement in WEB 2.0	ROI case study with detailed explanation.
Yongfook of Egg Co.	Social media ROI	Definding of success metric (quantitative and qualitative)	Examples of implementing metrics for several KPI (different firms) for achieving the set goal using social media.
InSites Consulting	The Social Dynamics Model	A four step approach of introducing Social media in the company and customer-centric thinking	Step by Step explanation how to introduce social media model into company - detailed roadmap. KPI schema - very well structured for Social media (p. 47)Step by Step explanation how to introduce social media model into company - detailed roadmap. KPI schema - very well structured for Social media (p. 47)
Margaret Francis	Measuring the Future: New Metrics for New Media	"SOCIAL" METRICS	Metrics into KPIs for most of the social networks

Table 18: Example of literature review of practitioners materials

5.1.3 Conference

In July 2012 we have taken part in Summer School "Social business process management" that was held in Como for four days. The general purpose of the school:

"For the Social BPM is defined as the effort of designing and executing business processes collaboratively by leveraging social networking. It aims at: – Exploiting weak ties between people and implicit enterprise know-how to improve activity execution and disseminating of knowledge – Increasing transparency and participation to the decision procedures – Involving (informal) communities in activity execution, thus assigning the execution to a broader set of performers or to find most appropriate contributor within a group."

The Summer school was involving significant Italian and international partners working in the field of implementing solutions of using social media for business processes, such as OMG, BPM4People Project, Future Strategies Inc and Università di Trento. The group of the professors was leaded by Marco Brambilla, who speaking during these days together with Fabio Casati, Piero Fraternali, Bruce Silver, Sandy Kemsley, Richard Soley, Claudio Bartolini, Florian Daniel.

The program of the school was consisting of the following main topics, that further were included into our literature review, as well as the knowledge received into the research performance:

- BPM: technical and business challenges (Fabio Casati, Unitn)
- Mashups, personal processes and task composition on the web (Florian Daniel, Unitn)
- Crowdsourcing foundations (Marco Brambilla, Polimi)
- GWAP techniques and applications (Alessandro Bozzon, Polimi)
- Social BPM: a model-driven approach to design (Marco Brambilla, Polimi)
- The WebRatio tool suite for Social BPM (Emanuele Molteni, WebRatio)
- Social BPM: a business perspective (Sandy Kemsley, Independent consultant, Canada)
- Social in the Enterprise (Claudio Bartolini, HP Research, USA)
- Idea Management and tools for grassroots innovation in the enterprise or local government (Gregorio Convertino, Xerox XRCE, France)

5.1.4 Conclusions

Interpretation of the all collected information were dived into three main parts:

1. Social media overview;

2. Management control systems overview;
3. MCS and performance of social media.

In the first two chapters it is carried out the fundamental overview of current state of art separated and chronolysed internally by the theme (i.e. social media; management control systems). Instead, in the third chapter there are discussed main drawbacks of the systems and future possibilities after the integration of the social media into the management process and identification of the literature gaps.

5.2 Empirical analysis

The empirical analysis is represented by two main methodologies: action research (experimentation study) and case studies approach (interviews).

The action research (AR) methodology was chosen as the most appropriate way in order to reach our personal objective of developing thesis by creating an innovative approach in the management control systems that should bring improvements into the system itself (confirmed by the use of this approach on the visible and reliable study) and follow the ultimate trends in business, benefiting from the progress of the information technologies. We also thought that the action research is the most suitable methodology, as it is widely used for the social science and organizational development: "in these areas it has a particular niche among professional who want to use research to improve their practices" [Denscombe, 1998, p. 57]. Moreover, after the primary step of the literature review we detected the room for the new framework implementation in the MCS, which by now was only partially developed in the marketing sphere and in the internal business processes.

Furthermore, within the AR we have introduced the experimentation approach in order to gather information directly from the company where the new framework have been applied for the phenomenon observation. Observation is one of the three elements of the triangulation model of multiple source of evidence. Also we have introduced the case studies (interviews) methodology for collecting and listening the people experiencing usage of the social media for internal and external purposes inside their companies, as a starting point for research. Last, but not the least we have used archival data which were mostly represented by the "practitioners" materials review and taken into account with the certain degree of reliability, as such sources in our cases are not scientifically

trustful, however reflecting much more clearly true state of the current business practices.

Upon three typologies of the action research: a) technical [6] ; b) practical [7] and c) emancipatory[8] - we have followed the one which is satisfying needs of our research project. Therefore, the Practical action research, as the Euro-Fer case first of all requires understanding of the companies reality and transformation of consciousness of the Euro-Fer management board in order to enter and use the Social media platforms for the objectives set in general strategy of the company development. The general overview of possibilities of the Social media have been carried out from the literature review and cases studies (interviews) in order to have the most definitive background in order to perform experimentation.

In order to implement this task we have carried out the following activities, according to the four steps model introduced in the literature [Grundy and Kemmis, 1981; Kemmis and McTaggart, 188; Perry and Zuber-Skerritt, 1991; Bakerville and Wood-Harper, 1996, Dick, 2002]:

1. *Develop a plan* in order to improve practice, from the general idea to the certain aim - objective of the research.
2. *Act to implement the plan,* by means, to deliberate and controlled process leading to a careful and thoughtful change in practice [Kemmis and McTaggart, 1988]. Each action during the research process is express the plan stage, but also

[6] Technical action research requires the testing of an intervention based on a pre-developed and specified theoretical framework. The intent of the research is to question whether the selected intervention can be applied in a practical setting (Holter and Schwartz-Barcott, 1993).

[7] Practical action research requires the researcher and practitioner to join together with the intention of identifying potential problems, underlying causes, and possible solutions or interventions. The researcher encourages participation and self-reflection of the practitioner (Kemmis and McTaggart, 1988: Holter and Schwartz-Barcott, 1993).

[8] Emancipatory action research require the involvement of all participants equally with hierarchy existing between the researcher and the practitioners. The aim of the researcher is to reduce the distance between the actual problems identified by the practitioner and the theory used to explain and resolve the problems. The researcher facilitates the discussion with the practitioners, so as to identify potential underlying problems and assumption and allow the researcher to become a collaborative member of the group (Kemmis and McTaggart, 1988; Holter and Schwartz-Barcott, 1993).

bears in itself the risk, as the action is taking place under the real circumstances in real life situation.

3. *Observe the action and data collection*. Observation should be flexible and responsive for the representation of the real life situation, probably not planned and together with the data collection stage, leading to reflection stage.
4. *Reflect on the action and re-plan*. The reflection stage is providing as-is situation and inside information which helps to move the research forward for the next level of the spiral.

These steps were implement in whole research path in order to achieve the set results. Likewise, the case studies approach helped to enrich the research project and to avoid possible drawbacks of each methods in order to perform the complete work.

5.2.1 Case studies

Case study research is one of the tree most used strategies of empirical research (such as field experiment research, statistical research or survey). We have chosen the descriptive case study in order to verify that a phenomenon of using social media information as supplementary to the traditional information, being implemented as the social media scorecard section, or alternative instruments of the management control systems together, as well as to see the level of usage of such techniques in different companies. Thus we have introduced the set of interview, as a part of the triangulation model.

The research interview is one of the most important qualitative data collection methods and has been widely used in conducting field studies and ethnographic research [Goldman and McDonald, 1987; McCracken, 1988].

Therefore, conducting qualitative research interviews process, there are many decisions that must be carefully considered, such as who to interview, how many interviewees will be required, what type of interview to conduct, and how the interview data will be analyzed [Doyle, 2004].

A well-planned interview approach can provide a rich set of data. The interview process is an opportunity to explore the meaning of the research topic for the respondent and our framework to be examined for the right to exist and being used in business.

According to Alvesson's framework existing 3 main types of the interview methods:

structured, semi-structured and unstructured interviews.

Structured interviews

The structured interview is where the interviewer asks interviewees a series of pre-established questions, allowing only a limited number of response categories.

Since researchers take a very active role in question design, there is a possibility that they inadvertently or overtly bias data collected. Thus, highly standardized procedures are designed to substantially reduce the probability of the results being influenced by the interviewer.

Unstructured interviews

The unstructured interview process shapes to the individual situation and context, intending to make the interviewee feel relaxed and unassessed [Hannabuss, 1996]. Fontana and Frey (1998) point out that most of the data gathered through participant observation is gleaned from informal conversations in the field. The unstructured interview proceeds from the assumption that the interviewers do not know in advance all the necessary questions.

Semi-structured interviews

The semi-structured interview involves prepared questioning guided by identified themes in a consistent and systematic manner interposed with probes designed to elicit more elaborate responses. Thus, the focus is on the interview guide incorporating a series of broad themes to be covered during the interview to help direct the conversation toward the topics and issues about which the interviewers want to learn. Generally interview guides vary from highly scripted to relatively loose. However, the guides all serve the same purpose, which is to ensure the same thematic approach is applied during the interview. The semi-structured interview enjoys its popularity because it is flexible, accessible and intelligible and, more important, capable of disclosing important and often hidden facets of human and organizational behavior. Often it is the most effective and convenient means of gathering information [Kvale and Brinkmann, 2009].

Based on the following definitions of the interviews and their main characteristics, together with the necessities of our research we decided to implement the structured interview. However, afterwards, during the interview sessions we observe that some interviews were semi-structured, as the participant are experts in the field of the social media applied for business purposes. Typology of questions used in research interviews.

We have carefully selected the participants of our interview by choosing companies that are actively using at least 3 social networks and have significant numbers of "likes", "comments", "shares", as well as have their own blog and maybe forum. We are proud to say that we have interview 8 very important and well-known companies. Generally, the interview could be divided into two parts:

1. Interview with active users of the social media for business purposes (6 interviews)
2. Interview with the companies consulting about usage of the social media (2 interviews)

Furthermore, while selecting the interviewees we have tried to have the right mix between the people from marketing, computer science and management departments. Below we have represented the summary table # 19 of the interviews held:

#	Company	Department	Position	Mode	Date	Duration
1	American International Consulting firm	Marketing and business development	Specialist	via phone	26.10.2012	35 min.
2	American Multinational Computing corporation	Digital Lead, Marketing organization	Head	via skype	02.11.2012	45 min.
3	Data intelligence company		Business Expert	via skype	23.11.2012	15 min.
4	Multinational Manufacturing firm	Digital Media	Director of Digital Media	via phone	05.11.2012	20 min.
5	Italian Spirits company	Marketing and Advertising	Assistant Brand Manager	via skype	01.11.2012	28 min.
6	British Multinational Telecommunications company	OnLine Consumer Division	Specialist	via skype	07.11.2012	32 min.
7	Italina Consulting company	Public Relations and Communications	Content & Community Manager	via skype	29.10.2012	47 min.
8	Multinational Computing and Communications corporation	Laboratories	Research Manager and Principal Investigator	via skype	02.11.2012	30 min.

Table 19: Summary table of the interviews

To ensure the quality of the interview we were heeding suggestions of Shensul et al. [1999, p. 141] paying careful attention to the following three principles:

1. maintaining the flow of the interviewee's story;
2. maintaining a positive relationship with the interviewee; and
3. avoiding interviewer's bias.

We paid special attention to preparation of the questions for this interview and in line with the two types of the interview represented the specific questions for the users of the social media, which should help to understand and evaluate better the "Use" part of the proposed framework, and consulting, giving an alternative vision of the "Use" and future perspective of the place of social media in business.
Questions of the interviews were prepared based on this classification adapted from Kvale [1996, pp. 133-5]:

1. Introducing questions (to kick start the conversation and move to the main interview);

2. Follow-up questions (to direct questioning to what has just been said);

3. Probing questions (to draw out more complete narratives);

4. Specifying questions (to develop more precise descriptions from general statements);

5. Direct questions (to elicit direct);

6. Indirect questions (to pose projective questions);

7. Structuring questions (to refer to the use of key questions to finish off one part of the interview and open up another, or to indicate when a theme is exhausted by breaking off long irrelevant answers);

8. Silence (to allow pauses, so that the interviewees have ample time to associate and reflect, and break the silence themselves with significant information);

9. Interpreting questions (similar to some forms of probing questions, to rephrase an interviewee's answer to clarify and interpret rather than to explore new information);

10. Throw away questions (to serve a variety of purposes, i.e. to relax the subject when sensitive areas have been breached).

Therefore, we mainly concentrated on the structured questions with mix of direct and indirect question for the interview with users, instead we tried to use more probing question and strongholds for the consulting interviewee. As the interview were done in three languages English, Italian and Russian all the questionnaires were prepared accordingly in three languages (the questionnaires in English and Italian, an example is in the Appendix p. 194).

Moreover, during the interview we faced some ethical and confidential problems, therefore it was decided to hide names of the participants in order not maintain the requested by the participants level of security. Furthermore, by request of some participants some of the web-sites used by their companies for analysis were also not shown in the results part of the interviews.

Overall the interview are very important part of our research helping to add necessary information and practices, that are presented in the case study of Euro-Fer.

5.2.2 Experimentation analysis

However, it can be argued that action research through practice-driven and small-scale, should not lose anything by way of rigour. Like any other small-scale research, it can draw on existing theories, apply and test research propositions, use suitable methods, and offer evaluation of existing knowledge [Steven French, Coffs Harbour Centre for Enterprise Development and Research, Southern Cross University, Coffs Harbour, Australia, 2009]. Based on this we have al carried out the experimentation of our new framework on the real case.

Further, taking into account that this is action research study, we foresee that it is interactive and about real change in company management approach that was conducted in the real time form May 2012 to November 2012 together with Euro-Fer and our business tutor Nextbit. As well as implementing all corresponding to the action research steps of diagnosing, action planning, action taking, evaluating of the action and specifying learning.

Comparing all the possible ways of collecting information we took into account significant benefits and disadvantages of this approach, such as that it reflects the reality and following our framework context. In contrary, we were aware that it is costly and time-consuming process, requiring robust selective approach.

In order to perform the experimentation we have followed the methodology presented on the Septembers course for thesis writing by Federico Frattini:

1. Choosing case study;
2. Selecting the cases;
3. Collecting information;
4. Analyzing information;
5. Ensuring quality;
6. Presenting results.

Here and after we have described in details the subsequent steps in the experimentation study's performance.

5.2.2.1 Selecting the cases

Firstly, when we started our work we have made a deep analysis of the literature and current market offer of the new management control systems and social media work through. We found out that in nowadays more and more companies were using social media for internal and external purposes inside the companies. In order to select the proper case study we have identified four main streams of social media usage: marketing perspective, which usually proposed by the consultancy firms for improving and launching new marketing campaigns for the clients; human resource perspective, which is using the social media to find and select personnel, here we also identified that it could be used totally in different manner for the career service of Politecnico di Milano, helping students to find stage or permanent work; strategy perspective, including information from the social media for decision-making process.

After identification of these possibilities we have prepared official proposals (examples of proposal could be find in the appendix on the pp. 196-205) for collaboration that have been sent out to 10 potentially interested partners, such as Ernst & Young, Telecom, Career service of Politecnico di Milano and Euro-Fer.

Furthermore, we have made series of the telephone negotiations with the interested parties and set up face to face meetings.

Secondly, we have held meeting with Telecom Italia in their office in Turin in order to discuss in the details our human resource proposal and their requirements of the project performance. Meeting were very constructive and robust, however Telecom Italia was

requesting that this kind of project were performed also by the computer science student in order to write the concrete solution for the human resource department. Based on this information and tight deadlines in which Telecom was wishing to perform the project, we selected another case study.

Thirdly, we have set up a meeting with the Euro-Fer, medium firm, based in Veneto region of Italy, for the long distance between Como and Veneto we have held meeting by using the skype. During the meetings we discussed Euro-Fer requirements and expectations. On one hand company wish to increase the awareness of its social media channels and launch them with the possibility to use information generated there, on the other hand management are particularly interested in the new Russian market, that they would like to enter, therefore how social media could help them to execute both tasks.

Analyzing all information received on the meeting we have prepare the initial action plan which afterwards was discussed and reviewed with Euro-Fer via skype in the followed week.

As a results of this research stage we have analyzed and chose the appropriate case for applying our framework, moreover prepared state of art phase for the collection of the information.

But before proceeding with the data collection, we have met with Nextbit (*Nextbit is Italian consulting company, which vision is to become a strategic partner for the clients. Since inception, they have built on this vision by delivering high quality consulting projects, created a strong team culture based on respect and collaboration; they have ensured that rewards and objectives are not based on monetary results but on successful projects, satisfied clients and enthusiastic colleagues*) - our business tutor for this research project, as they are working in the same direction with their clients, consulting and implementing instruments based on the social media data. We have agreed to help them with developing solutions of data collection and analysis based on open sources products, implementing these solutions both for Italian and Russian language, for which such analysis are very appreciated as most of the tools are developed for English language. Moreover, we have developed with Nextbit in tight collaboration the methodology for the processing information from the social media

from entry points to discussing the results, which is represented in the conclusion of our research.

5.2.2.2 Collecting information

The methodology of the collection information for such project as we performed for the Euro-Fer consists of the following steps:

- Key words selection
- Sources/fonts
- Web-crawling

Key words are the words mostly characterizing the subject of the research that should used for the identifying the list of sources or fonts that afterwards would be used for the web-crawling. According to the fact that we agreed with Euro-Fer two project: one for Italian market and one for Russian with their specific objective, we have prepared the two specifics list of the key words in Russian and in Italian.

For the first review we have submitted the vast list of the possible key words, that was discussed with Euro-Fer representatives via skype. Than the list of the key words were improved and send for the final approval to the Euro-Fer. Before starting to work with the key words for the sources and fonts search, we have divided into the several levels, identifying the key words for search and taxonomy of the research objects, the final result is represented in the following table # 20:

Italian	**Russian**
Key words for web-crawling	
FERRO BATTUTO	**КОВКА**
INOX	**МЕТАЛЛ**
Tassonomia	
Decoration	
Componenti per arredamento	Металлические / стальные детали/элементы мебели
Componenti per caminetti	Кованные и сварные изделия из металла
Decorazioni in ferro e swarovski	Декоративные элементы/изделия из железа / стали / чугуна
Decorazioni in ferro/inox/ferro battuto	Элементы кованого железа / стали
Lance e Borchie	Элементы интерьера и внешние декоративные изделия из железа

Elementi in ferro forgiato/inox	
Elementi d'arredo interno/esterno in ferro	
Elementi in ferro ricalcato/inox	
Ferro martellato	
2nd level	
maniglie	ручка
cassette postali	почтовый ящик
pali scorritenda	карниз
attaccapanni	вешалка
cimase	петля
ruote	пика
cardini	заклепка
ribattini	клепка
scala	лестница
parapetto	парапет
balcone	балкон
recinzione	корпус
cancello	ворота
ringhiera	перила
corrimano	лестница
partenze	решетка
elementi tubolari	поручень
funi	вылет
torciglioni	прут
fascette e barre	труба
barre forate	конус
barre sagomate	
la pigna	
Products' combination	
vetro	стекло
legno	дерево
swarovski	сваровски
Company	
Euro-Fer	
Mazzacani	
Lyuiforging	
Rik-fer	
Feba	
Fer-bat	
Fatih	
Sahinler-forge	
Lecky	
INDIA	

People	
Architetti	архитектор
Installatori	дизайнер
Fabbri	строитель
	прораб

Table 20: Developed taxonomy of words for data analysis

Source analysis is crucial part of the analysis of the social media as it defining the main input that it will be received in the end. Thus, it is so important that the sources of the data were reliable, efficient.

For the web-sites, we have taken into account the following features:

- Google ranking
- importance of the sphere of the object of research
- number of subscribers
- importance of the sponsors
- number of the post per day
- number of the comments, likes, shares

Moreover, with the initial list of the websites of the direct competitors, suppliers and customers were provided directly by the Euro-Fer for the Italian objective. Instead for the Russian objective we have contacted the designer/architect in Russia, who is directly working with such products and she provided the comprehensive list of the forums and journals that we have used in our analysis.

Overall, we have collected 44 websites, as the entry points. Before, processing with the next stage of the web-crawling, we have ensured that all of this sources are comprehensive and created a table of entry points analysis, which includes such columns as name of the source, tag (competitor, forum and etc.), tag2 (where this website is coming from), language, to be used or not and note, why we have used or not this website.

Name	Tag	Tag2	Language	Uso	Note
http://www.arteferro.com/arteferro	Competitor	Euro-Fer	English/Italian	si/no	solo il blog
http://www.fatih.com.tr	Competitor	Euro-Fer	English	no	lingua non di nostro interesse
http://www.sahinler-forge.com	Competitor	Euro-Fer	English	no	lingua non di nostro interesse
http://www.lecky.it/default.asp	Competitor	Euro-Fer	English	no	niente forum
www.luyiforging.com	Competitor	Euro-Fer	English	no	niente forum
www.rik-fer.com	Competitor	Euro-Fer	English/Italian	no	niente forum
http://www.feba.it	Competitor	Euro-Fer	English/Italian	no	niente forum
http://www.fer-bat.it/	Competitor	Internet	English/Italian	no	solo social network e pochi commenti
http://www.mazzacani.it	Competitor	internet	English/Italian	no	articoli non commentati da consumatori, troppo di parte
http://www.ferroearte.it/	articles &Blogs	internet	Italian	no	utilizziamo forum sotto
http://www.archiproducts.com/	articles &Blogs	Euro-Fer	Italian	da valutare	login collegato a edilportale vedi sopra

Table 21: Example of the entry points analysis

After the analysis of all source we have chosen 4 websites for Italian language and 3 websites for Russian language, for which we have proceed with the web-crawling.

- http://blog.lavorincasa.it/ (articles)
- http://blog.lavorincasa.it/ (forum)
- http://www.arredamento.it/
- http://wwwferroearteit.globalfreeforum.com/viewforum.php?f=1
- http://forum.homeideas.ru/
- http://forum.ivd.ru/
- http://360.ru/

A Web crawler is a computer program that browses the World Wide Web in a methodical, automated manner or in an orderly fashion. In other terms for Web crawlers are ants, automatic indexers, bots, Web spiders, Web robots, or—especially in the FOAF community—Web scatters. This process is called Web crawling. Many sites, in particular search engines, use spidering as a means of providing up-to-date data. Web crawlers are mainly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches. Crawlers

can also be used for automating maintenance tasks on a Web site, such as checking links or validating HTML code. Also, crawlers can be used to gather specific types of information from Web pages, such as harvesting e-mail addresses (usually for sending spam) [Wikiperdia, 2012].

Working with the open source programs we have always to be attentive, as a huge number of new programs are developed, as well as the new versions are entering the web. Therefore, in order to have the best possible results, we have tried to use the following web-crawler.

- Rcurl package of R-project
- Rafabot
- Blugrab
- Firebag
- XML package of R-project.

As described further in technology part we have tried to use and evaluate each of these tools to perform the work. So, after the evaluation, we have used the XML package of R-project that was giving good results both in Italian and Russian. We have also used the Firebag in order to tag the web-pages for analysis.

Furthermore, in order to perform the crawling we were tightly working with the Nextbit, discussing possible solutions with them and sharing codes in R.

Afterwards, we have written down 28 codes in R each for the font and for the key word used, as the crawling depending on the structure of each website and downloading information according to the key word, please see below the examples of Italian and Russian codes in the appendix from page 202 to page 225.

It is important to mention that the language problem of Italian and in more of Russian was overcame by using specific encoding UTF-8, which required rewriting of all codes.

Therefore, we have downloaded according to the key words from each of the website chosen in total **128,864 comments**. Specifically in this part we have followed the action research model planning, reviewing and implementing a new plan for each activity. Moreover, this phase of the data collection took the significant time being management engineering students, but we have learn a lot and understand the process of the crawling.

5.2.2.3 Analyzing information

The phase of the analyzing information were performed the 14 steps, that focalized on ensure the information quality and reliable results, moreover provide text mining and sentiment analysis, giving important results for the case study and research itself:

1. Pre-filter
2. Structured texts (metadata)
3. Filtering texts
4. Washing texts/ Preparing texts
5. Ttokenize: metadata at the level of the words/phrases
6. POS tagging
7. Stemming/Lemming
8. Reconstruction of the lemmed phrases
9. Phrases matching
10. Reduction of the dimension
11. Word matching
12. Summarising matching at the words and phrase level
13. Sentiment analysis rule-based
14. Text mining

After the web-crawling results, we have imported information from the R to the excel format in order to analyze the received data. Then we have made data filtering and seen that for the web-site - http://wwwferroearteit.globalfreeforum.com/viewforum.php?f=1, for "acciaio inox" number of comments are not sufficient lower than 1000, therefore we have eliminated it from the further analysis. Moreover, we have made data washing by using the commands of R, specifically adapting the code for receiving better results, nonetheless the process of the data washing is one of the most important in the whole process that ensuring the quality of the final results, helping to focalize on the objective of the research (codes could be found in appendix devoted into Russian and Italian cases on pp. 202-225).

Later on we have started to work on tokenizing and tagging data, this was really difficult and time-consuming task, as the Italian and especially Russian language using different logic and alphabet of the language, thus could not be simply translated with the Google translator and after the English rules of the language applied. Hence, we have to

search for the specific instruments to perform this task which are required both text mining and sentiment analysis, the detailed explanation of tools could be find in the technology part of empirical experimentation methodology. However, thanks for collaboration with Nextbit, we have performed lemming both for Italian and Russian data, using the algorithm of the TreeTagger and dictionary together with subset for Italian language provided by Marco Baroni's and for Russian language by Serge Sharoff (codes could be found in appendix devoted into Russian and Italian cases on pp. 202-225).

As soon as we have received the washed and lemmed data we have made the sentiment analysis using the R- project package (tm) according to the rule-based algorithm, as well as we have written the code also based on the rule-base algorithm. Thus, we have tried the both methods to make sentiment analysis and both of them giving good results for Russian and for Italian data. For the large amount of the comments we have used our code, that was performing work faster and did not apply the numerous number of parameters of tm package. But it is important to mention that the tm.plugin.sentiment package could provide a more precise information with implementation of all the parameters, but requires as well the significant time for computation process.

Following the sequence of the action name above, we have performed the text mining, which gave from the start good results and let us to create the word cloud, clustering, frequency of words and frequency of taxonomy words, as well as association with the key-word, however the last function is requiring the high computation time, which was not possible to perform for each font (codes could be found in appendix devoted into Russian and Italian cases on pp. 202-225).

As the result of this phase we have received the detailed analysis of the data represented in the chapter # 7.

It is important to point that all the process of the data analysis was flow of activities in order to find the best possible solutions in terms of the results and time, that could be afterwards implemented for the business purposes. Moreover, we have followed the action research approach evaluating actions taken and reflecting and replanning our activities.

5.2.2.4 Technology, instruments & tools

During our research work we have used, tried and compared more than 60 different tools and instruments of the information technology in order to develop our project and reach the posed results.

In nowadays the information technology is highly developed and usage of certain instruments are familiar to all users of computers. However, for our research we have applied linear approach of evaluation of the technology instruments and tools, following logic of the usability of these tools for a real company, such as Nextbit, but also aiming to use open-source products in one way more easy in terms of material side, in other way more easy to be modified according to the specific requirements of our project and could be applied as tailored decisions for the business.

Therefore, we have taken part in the course of R-project (http://www.r-project.org/) to receive the necessary knowledge about this platform of the open source instruments that could be used for different purposes, but most importantly the several packages specifically devoted to the Web-crawling, Text mining and Sentiment Analysis.

5.2.2.4.1 Corse of R-project

In the beginning of June 2012 we have taken part in the course of the R-project, that was crucial and starting point of the technological analysis of the information collected during the research. The course aimed to provide an overview of the basic R environment and its applications. As well as this course was a starting point for any future development with R, gain awareness of the basic R language, importing/exporting data and manipulate data using R and exploratory data analysis, perform basic statistics and build plots using R.

The course were leaded by the professional team having large experience in using such open-source software for different purposes. We have received the detailed knowledge about R data objects, data import and export, using R functions, R graphics, writing a function, introduction to statistical models with R, tables with R and etc. Being team of two management engineers, it was highly important know how to use and modify program according to our research needs and general logic of our case study, furthermore this course was important step in our project.

5.2.2.4.2 Instruments

The analysis of the instruments is following the general methodology of the work and work flow of the data analysis. Thus, it is divided into the four main steps, together with the sub-stages (such as lemming, stemming, work with documents and words, etc.), represented in the following table # 22:

Phase of Analysis	Instrument
Source analysis	By-hand
	Mentioning
	Sysomos
Web-crawling	BlueCrab
	Rafabot
	Navicrawler
	Dapper
	TM.plugin.webcorpus
	Rcurl
	TwitteR
	XML
Text-mining	TM
Sentiment analysis	Sentiment
	TM.plugin.sentiment
	koRpus
	OpenNLP
	Rstem
	Snowball
	TreeTagger
	Morph-it!
	WEKA
	TM.plugin.tags
	SentiWordnet
	NLTK

Table 22: Stages of data analysis and instruments

From here and below we will describe the main tools used in our research abiding these main features of each tool in order to have the best possible results of the data collection, selection and have clear idea of the tools characteristics and their possible usage in this or in any other work, which are represented by:

- name of the instrument;
- short description (objective, owner, for which operational systems could be used);
- function;

- benefits,
- disadvantages.

5.2.2.4.2.1 Source analysis

By-hand

The traditional way of searching of the sources using the key-words, detected together with the client for the subject of the interest, no specific technology is not required, the different search engines could be used such as Google, Yahoo, Yandex and Rambler (for Russian fonts).

Main functionality of this kind of the fonts collection is allowing to include all possible solutions and make raw web-sites collection, that further could be analyzed by other tools.

Benefits: large fonts set, first possible categorizations based on the position in the search results in the search system.

Disadvantages: large data set, required further accurate analysis, not final fonts list, in some cases could be too large. The possible way of resolving such drawback is to make the key words more precise.

Mentioning

More accurate way of searching for the fonts, starting from the page of the client or its' direct competitors, moreover, could be used the list of suppliers and customers. This method has also no limits, could be used with any operational system and with any search engine. Most often the owner of such lists are the clients of the project knowing their environment.

Functionality is that this kind of list are collected by the players during their work on the market and are not immediate

Benefits assurance is more precise search of the fonts that could lead to better results in terms of the final information, assurance of theme of the interest mentioning.

Disadvantages significant level of the information predisposed, by means that players are talking about the research object in particular way, not always reflecting the final

consumers opinion about the product/service. However, this information is important for the benchmark analysis.

Sysomos (http://www.sysomos.com/)

The general object of this platform is to monitor the social media, make analysis and manage the social networks. This tool could be used not only for the fonts analysis, but also after the in the stages of the information analysis. Owner is Marketwire Inc. System is working online with the specific login credentials for each user, no operation systems limitations. The Demo version is free of charge.

Functionality to know precisely in which social media is spoken about the subject of the interest, who and what is saying.

Benefits the ready to use engine with the predisposed entry points of all social media, that can provide the whole list of the fonts and analysis of it.

Disadvantages being a whole product it has limitation of the usage as well as limitation of the modification, therefore it is covering most of the social networks, but not all the forums and blogs of the potential interest.

5.2.2.4.2.2 Web-crawling

BlueCrab (http://www.limit-point.com/BlueCrab/BlueCrab.html)

BlueCrab is a versatile application that could be used to download the contents of a website in whole or in part, is owned by the Limit Point L.L.C. The trial of the program if free of charge and working for the 30 days which was enough for our work performance, but could be crucial for the day to day work inside company. BlueCrab has a strong limitation as working only on the MAC operation system.

Functionality is downloading files with no internet connection, relinking and remapping links in the downloaded files, can proceed the batch of URLs at once from the browser, archiving the websites and collecting resources, saving the specific types of the information such as images or email addresses using configurations. Furthermore the techniques yield very compete downloads and dynamic links.

Benefits are the easy interface and usability, many options and possibility to download the information from the same sources periodically.

Disadvantages the output of the data in the 'csv' format is quite difficult, so the whole process of the further usage of the data is quite complicated, moreover it is not possible to use with the Microsoft operational system and gave problems with usage of the Cyrillic letters set as the key words for search.

Rafabot

The general purpose is the web-crawling of the website from a starting URL, search engine results or web. It is property of the Spandix Software, as well as with the BlueCrab the trial version is working 30 days, but the whole version should be bought. Rafabot is working only with the windows operation systems and could not be installed on MAC.

The main functions are similar to BlueCrab, could work offline, browse several URL's for the downloading the data, download websites form the input URLS, extract the information.

Benefits are the simplicity of usage and good menu for operating, a lot of functions for performing the analysis.

Disadvantages not possible to use for all languages, giving errors of downloading data in Russian and not possible to use for MAC, in Italian difficulties of downloading the data for further analysis.

Navicrawler

The Navicrawler is monitoring links of the navigated pages. This tool is owned by the Webatlas and is compatible with Windows, Mac requiring the installment of the Mozilla.

Navicrawler is permitting to crawl, export information in the 'csv' files with the filters, as well as permit to create the list of the URL's of the entry points.

Benefits are the user friendly interface and commodity of the usage.

Disadvantages are the technical documentation only in French language and necessity to resave data for further analysis in other engines.

Dapper

Dapper is the web application presented by Yahoo, permitting to select the display content from HTML code given as a input source.

Dapper is the mainly representing the content, the crawl function is working not entirely.

Benefits are that Dapper identifies the parts of the menu from html, very convenient, makes you want to export data in html.

Disadvantages is not possible to make automation, required therefore the significant time to make crawl of the information that should be downloaded

TM.plugin.webcorpus

TM.plugin.webcorpus is one of the package introduced in the open source R-project platform in order to perform the crawling. It is possible to use on every operation system after downloading the R-project.

A set of functions to be tested and downloaded for reading news feeds. contains also function for extracting text starting from the whole pages.

Benefits are functions to retrieve data from news feeds like Google and Yahoo news, useful to get ideas for the crawl itself.

Disadvantages: specific to the web news only and hardly adaptable for other kind of information.

Rcurl

Another package of open source R-project presented by another writer, that also could be used by any operation system on the platform of the R. Rcurl is very popular tool that sends HTTP requests to the server and reading the results pages by Google.

Functions are to crawl according the URL's, download the comments from the web-sites.

Benefits easy to use, good results of the crawling, easy to be further analyzed with other instruments of R-project.

Disadvantages: Rcurl does not follow the links on the web-pages, in order to have comprehensive crawl the precise list of the URL's should be provided.

TwitteR

TwitteR is the specific package of the R-project made to download the comments from the Twitter, according to the chosen theme or subject.

Functions: crawling of the Twitter, downloading comments in the 'csv' or 'txt' format with the possibility to download information according to the filters.

Benefits are the simplicity of this package and convenience to use on ongoing bases, the easiness to work with the results.

Disadvantages are that for each key-word the single crawl should be done and for the different languages the code of R required specific adaptation in encoding.

XML

XML is the package of the R-project with the all appropriate characteristics of R-package. Functionally used for the extraction from HTML pages text and transforming into XML the post and performing crawling.

Benefits are the complete for the management for the making crawling and have a lot of the command to make the comprehensive crawling.

Disadvantages are the interface and difficulties of writing the additional and connective commands in natural language.

5.2.2.4.2.3 Text mining

TM package of R-project

TM package of the R-project is performing the text mining from studying the frequency of the terms. However, it is required to download the stop-word list and also make analysis of the specific stop list for the subject, as well as choose the most comprehensive libraries.

Benefits: the good performance of the package, possibility to make clustering and large number of the command varieties for wider analysis.

Disadvantage are that for non-English analysis a lot of libraries are required to be found and downloaded and quality of the open source libraries should be checked.

5.2.2.4.2.4 Sentiment analysis

Sentiment

Sentiment package of the R-project is the sentiment analysis using classification algorithms type 'naive bayes', which are based on a default lexicon rules. This package permit to classify emotions into one of the sixth emotions, classify polarity using 'naive bayes', moreover to classify the information based on the trial data extraction based on Janyce Wiebe's subjectivity lexicon.

Benefits: allows to perform sentiment analysis of the words from the comments.

However, the main disadvantage that is not using the content of the phrase or sentence and impossible to introduce the simplest denial 'not'. Also do not support any other languages but English.

TM.plugin.sentiment

TM.plugin.sentiment package of the R-project calculates the sentiment of a topic, its series and represent it in the graphic format allowing to classify emotions of the text. Important that for making functioning this package it is necessary to make text-mining, have the document with the term matrix, dictionary of the sentiment-laden words, make the stemming (could be done for English with usage of the snowball program).

Benefits of this package is the large number of opportunities and functions.

Disadvantages are the high quality of the input data, including the dictionary for each language and problems of usage snowball program for other languages. Also this package do not use the contest of the phrase or sentence.

5.2.2.4.2.5 Tools for lemming, stemming and working with words and data

koRpus

KoRpus is the package of the R-project developed for the analysis of the texts. Functionalities of this package are language detection; readability, frequency analysis, lexical diversity. Also it can use the TreeTagger as tokenizer and POS tagger for identification of the word class (name, adjective, verb and etc.). Inside this package could be done lemming.

Benefits: broad opportunities and rather high quality of the results.

Disadvantage the necessity to insert the correct dictionaries for the appropriate level of the results.

OpenNLP

OpenNLP is package of the R-project that allows to make the sentence detection and tokenize data. The following functions are integrated in this package: language detection, readability, frequency analysis, lexical diversity. It is also possible to use the TreeTagger as a tokenizer and POS tagger in order to make lemming.

Benefits of this package are that could be with different languages, could be done in one package lemming and POS tagging.

The main disadvantage for the business is not possible to usage for commercial purposes of the TreeTagger and absence of the similar tool for the commercial use.

Rstem

Rstem one of many R-project packages used for stemming. It is using the interface for the Snowball tool of Porter's word stemming algorithm. Main function is stemming of the words for further analysis.

Benefits are the rather good results for stemming of the words and ease of usability.

The disadvantage is that could be used only for English language, the Snowball tool as well as Porters word stemming algorithm were developed specifically for English, and working with other languages with non-sufficient results.

Snowball

Snowball is the tool used in R in order to perform the stemming. It is Porter's algorithm for stemming the words of the data set.

Benefits and Disadvantages are the same for the Rstem package.

TreeTagger

TreeTagger is a tool for annotating text with part-of-speech and lemma information. It was developed by Helmut Schmid in the TC project at the Institute for Computational Linguistics of the University of Stuttgart.

TreeTagger performance is very good and the main benefit is compatibility with large number of languages including most of the European languages, as well as Swahili, Russian and Chinese.

The disadvantage is the impossibility of using TreeTagger for the commercial purpose.

Morph-it

Morph-it is a free morphological resource for the Italian language. It is a lexicon of inflected forms with their lemma and morphological features.

The main benefits that it has large library inside for lemming the words and is free of charge and could be used also for the commercial purposes.

It is specific tool and could be used only for Italian language.

WEKA

WEKA is a suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. The WEKA workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality. It supports several standard data mining tasks: data preprocessing, clustering, classification, regression, visualization, and feature selection.

Providing rather good results, but the data should be imported in a proprietary format which is rather difficult, so it is not so easy to build a model.

TM.plugin.tags

TM.plugin.tags is the package of the R-project used together with the TM for tagging categories of the words employed.

As benefit also containing the sentiment lexicon that if required could be done lemming.

The disadvantage is that this package is not using the meaning.

SentiWordNet

SentiWordNet is a lexical resource for opinion mining. It assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity.

Provide quite good results, but only for English language.

NLTK

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning.

Benefits: includes significant number of languages and user friendly.

Disadvantage usage of the snowball algorithm for stemming which are not accurate.

This overview of the instruments were prepared together with the Silvia Gigante and Fabio Decontardi of Nextbit.

5.2.2.5 Ensuring quality

The quality of the data set was ensured by the quantity of the data set and specific action taken to wash the data. For each fonts and key words, we had at least 1000 comments, that considered to be very good results and sufficient for the reliable results of data analysis. Further more the whole process and methods used for identifying sources and key words, provided results coherent with the objectives of all the two sub-projects of the case study.

5.2.2.6 Presenting results

The overall results are presented in the specific format in the Results chapter of the Thesis. However, during the case study performance we have also developed and made the presentation of the Social media Strategy for Euro-Fer, which was done in September 2012 in their office.This presentation was the outcome of all our work for the case study, giving advice and proposing several metrics and uses of the social media, moreover developing in the objectives of the both projects of Euro-Fer for Italian and Russian market.

5.3 Analysis of results

The results of the research are represented according to the structure of our new framework proposed for the MCS and divided into the two big parts "Indicators" and "Uses". For results' analysis of each part we have implemented a defined treatment approach used in the experiments. Out of many ways to execute this, such as Explanation building [Yin, 1994], Constant comparison [Glaser and Straus, 1967], we have followed the Pattern matching [Yin, 1994], based on its main characteristics suiting perfectly our experimentation phase of the research process, i.e. finding evidence in the case that effects association for each causal path in the theoretical model under consideration was of non-zero value and was of the expected sign [Yin, 1994].

In other words, in both parts of analysis we introduced *process tracing* (the effect association is the results of a causal process and does not merely reflect spurious association), which were mostly implemented for the description of the Euro-Fer case study and *congruence testing* (it requires the selection of pairs of cases which are identical in causal type, except for the difference of one independent variable) that were crucial method for analyzing the interviews performed during our work before. Taking into account difficulties faced in the phase of the collecting data, programming in R-project and learning ICT component of our work, we focused on the controlled observation (researcher attempts to control for effects by looking at shift in magnitude or even valence, on key variables), slightly touching critical incident technique (collecting a large number of brief, factual reports on critical incidents, researchers attempt to identify common factors associated with effective outcomes) and time series technique (researcher also attempts to establish not only the existence, sign, and magnitude of each model link is as expected, but also the temporal sequence of events relating the variables in the model).

The results of our research were categorized, analyzed by implementing synthetic evidence tables and discussed in the conclusion part of the thesis, also predicting the possible ways of improving and overall findings by comparing the proposed framework with the case study findings and experimentation analysis.

6 Case Studies

This chapter presents a series of case studies regarding the use of Social media in the company comparing giving starting point for the empirical experimentation. At the beginning of the chapter we have discussed each case in the detailed providing thorough information of the Social media usage inside company (users and uses), data analysis and instruments, well as its advantages and disadvantages Afterwards, it has been done a summary table of the main evidences coming from cases studies, together with the general overview of the current usage of the Social media.

6.1 Case # 1

The firm taken as case study is a multinational company, operating in the manufacturing sector, based in Milan, Italy. It is present in over 160 countries, has 20 manufacturing sites around the world and a network of around 10,000 distributors and retailers.

6.1.1 Social media uses

The firm employs mainly external Social media to support the marketing department in the operating activities. The main objectives is to trying to create awareness, drive the sales, promotion, reputation integrity, crisis management, customer service and insight collection. To pursue these objectives it has been chosen a set of social platforms distributed upon earned, paid and earned media: the main actor is the personal profile on "Facebook", in which there are references to other social networks such as "Twitter" and "Instagram". Besides the popular social network the company has a presence on some professional blogs and forums. The activities and updates of contents on the Social media are made daily, basically depending on the topics and the news from the markets in which the company works; actually the contents delivery is done in strategic moments in which it is likely that the fans can see and interact with them.

6.1.2 Data analysis

The monitoring of the activities is performed directly by marketing department with the support of an external agency to create reports. The tools employed to support the analysis of data is "Radian 6".

The main goal of the analysis is trying to identify approximately the potential value creation of a fan of "Facebook". The tracking of the fans activities and the evaluation of

comments are two components for the computation of value. In addition it is performed a benchmarking analysis with the main competitors about fans, people engaged, campaign results, interaction with applications and participation at events. This kind of analysis is not that structured due to the high cost of implementation.

The reports are done monthly for the ordinary monitoring, while they are done weekly in case of extra-ordinary situations for instance campaigns, contests or special offers. The information is represented with charts, pies and dashboards for each source in order to show the trends and the values of the period.

6.1.3 Uses of data in the company

The reports created are used essentially by the marketing department, sales department and also by the country manager. The other departments have not access to this kind of information; such integrated approach is not used because the company is not ready yet to manage. In fact even if this data are immediate and pure, the main drawbacks is the absence of a systematized method to manage the information, this can bring to bad signals sometimes just because there is not a way to categorized the data, for instance there is not a unique way to weight comments polarity.

6.2 Case # 2

The firm is a well-established Italian company operating in the spirits sector, founded over 165 years ago in Milan. Since its establishment, the company has been characterized by internationality: a pioneer in the creation of a bitter liqueur made from herbs and roots from all over the world and in its marketing abroad, as the product was already present on several international markets in 1860.

The mix of product portfolio and strategic decisions taken by the present Chairman allow the company, now present in more than 160 countries, to compete in Italy and abroad with the most important multinational companies.

6.2.1 Social media uses

The company employs Social media to support the marketing department especially in the launch of new products. The main objective is to make the product known among the customers through a marketing campaign, so the communication through Web 2.0 is targeted towards the direct end-users with low costs. The resources are a set of Social media chosen for this purpose: the main one is “Facebook”, used as transmitter, assisted

by more specific platforms such as “Youtube”, “Instagram” and “Twitter” to communicate with videos, pictures and text sentences. The activities and updates are made daily with a frequency of twice per day.

6.2.2 Data analysis

The data coming from the internet campaign are gathered and analyzed using simple tools that are provided in each Social media and some applications such as “Facebook Insight” and “Twitter Counter”.

The main goal of the analysis is to evaluate the outcomes of the campaign in order to make changes about the product strategy promptly. In addition the competitors profiles are monitored to get new idea of communication and involving, of course it is also used to compare product perception, in case of similar object of control. The competitors’ analysis is useful also for benchmarking purpose

The reports are done monthly through presentations and pie charts, using the basic dashboard of indicators provided by the applications and tools chosen.

6.2.3 Uses of data in the company

Marketing department is the only one that uses report and analysis coming from Social media. Actually there are still some problems in the entire process of gathering and analysis of information. It is hard to understand the real why of behavior of people, even if Social media gives lot of information, but not the motivation that drives the people. Beyond this, it is needed a deeper analysis regarding the relation between user and Social media in a qualitative way, what the real factors are that moves the user to interact with Social media.

The main advantages in using this data is for sure the great possibility to change strategy in a very short time; while the disadvantage is to have just quantitative report reliable, in fact the qualitative component is still too much untrustworthy.

6.3 Case # 3

The firm is a British multinational telecommunications company headquartered in London and with its registered office in Newbury, Berkshire. It is the world's second-largest mobile telecommunications company measured by both subscribers and 2011 revenues (in each case behind China Mobile), and had 439 million subscribers as of

December 2011. It owns and operates networks in over 30 countries and has partner networks in over 40 additional countries.

6.3.1 Social media uses

The company uses Social media as listening channel in the marketing department, in human resources for employer branding purpose and in the management of the company.

Marketing department uses an own forum to listen opinions of the customers, complains and ideas in order to provide better satisfaction level in terms of price plans and customer services. It has got a presence also in other Social media such as "Twitter" and "Facebook" in which it provides a basic customer features. Actually the real purpose of this platform is to cover the function of broadcast transmitter of topic and contents that can be interested for the user. The personal profile on the main social platform is updated daily and managed by a team of three people (Social media Strategist) that plans a communication strategy in order to create traffic on the page.

The human resources department uses their own platform to communicate with possible new employers in which people share their experience and answer questions of students and/or people interesting in the company.

The management of the company uses an internal platform to communicate with people of the firm and also asking for experts opinion. Anyway the decision making is structured in the classical way, but for certain kind of markets is used the Social network.

6.3.2 Data analysis

All the process of data gathering is managed by an external party while for the analysis is employed an internal software. All the information are systemized in three main groups of indicators that are broken down depending on kind of social platform and geographic area:

- Brand visibility
- Sentiment polarity
- Contents

For each group, beyond the classical trend analysis, it is performed a benchmark analysis as well.

The analysis is formalized in a daily reports consisting in pie charts and direct links to the sources, without the integration of a balance scorecards system.

6.3.3 Uses of data in the company

The main users of this analysis are the whole marketing department and some level of management. The great advantage of this data is that is possible to make benchmarking because all the data are free to on the web, so creating appropriate metrics I can compare two companies, regarding topic, price plans and complains. So it can be considered as a real good listening platform but just for a part of the customers usually in the segment of age between 15-40 years old.

This great opportunity of listening brings the risk to increase false alarms about some complains, in fact there is not a reliable method to weight comments depending on some unpredictable factor of people such as the mood.

6.4 Case # 4

The company is an American multinational corporation headquartered in Washington that develops, manufactures, licenses and supports a wide range of products and services related to computing. The company was founded in seventies years of XX century. It is the world's largest software maker measured by revenues. It is also one of the world's most valuable companies traded on the NASDAQ.

6.4.1 Social media uses

The social media is in this company for several purposes, such as marketing, benchmarking, human resources and managerial. Social media is developing rapidly, and is the key area, where companies have possibility develop and provide a lot of activities all over the world.

In the marketing department currently you cannot operate without SM. Usage of it permit company to involve easily all key players into its marketing activities at one time, especially if we are talking about consumers. The SM is used as extend of the campaigns and should be detected adequately to the target audience. Any campaign launched in SM is to create awareness, to get the attention of the community to the

specific product, but also to all the existing product range, with purpose to create loyal customers' approach. Mainly the SM is the way of intercommunication with clients about new products and launched so-called support center, the portal where people can ask any question about the product, direct their feedbacks and get clarification of the products features.

For the benchmarking SM is mostly used by the product department in order to make comparison with the main competitors products and being aware of the situation on the market.

Likewise for the human resource department it is used as communication channel, providing information to the target audience (for an example - students) about new working possibilities within the company. Here, the SM instrument is chosen for campaign based on the specifics of the target audience taking into account the particularities of Russian market. In other countries the SM resource is used much wider for the human resource department, however Russian market is not so well prepared and not all people are aware about these possibilities.

In relation to the social media platforms, company is using all known sources: Facebook, Youtube, LinkedIn and etc.; together with the particular Russian ones: Odnoklassniki and Vkontakte, which is widely used both for B2C purposes and B2B activities. Of course being a large company, they have their own blogs, special platform for communication with clients and big internal social network for intercompany communication purposes, which are used for co-creation purposes, decision making and other managerial objectives.

The information updates regularly on each of the SM sources according to the each specified product department and their campaign. Overall the information is provided on the daily basis. Company is using account for each big company and several official accounts for each product group, their connected in some way and all important announcement covering all the communities, instead the specifics messages covering their product and target groups.

6.4.2 Data analysis

SM it is not the place where everything is controlled, that is why ongoing analysis of the information appearing is crucial and should be managed in proper way. Therefore,

company is using several "listening platforms" to monitor the sentiment of the comments and provided prompt response. Information from the SM is direct feedback from audience, the piece of data that can be analyzed. The quality of data is ensured by the statistical analysis implemented into the information analysis.

For such data analysis, company is using its own programs developed specifically for some purposes, but also some open-source application as:

- http://www.socialbakers.com/
- http://klout.com/home/
- https://www.twentyfeet.com/

As well as, commercial listening platforms:

- http://sysomos.com
- http://www.radian6.com (http://www.radian6.com/resources/)

Most of the reports are generated automatically, together with the measurement analysis, graphs, charts and other visual elements. Company also pay specific attention to the performance, such as indicators showing the traffic on the SM platforms. The generic reports could be generated anytime are required for each platform and product, according to the tailored developed metric for the each line. However, some specific reports are also prepared on the weekly, monthly basics, according to the needs of the departments.

6.4.3 Uses of data in the company

The main users of this information are the managers, that are permanently working with the SM sources for the specific department/product. As well as the human resource department is using this information for their own purposes, together with the head of the departments that using this information in aggregated way together with other indicators and already included in their balanced score card.

6.5 Case # 5

This firm is one of the largest professional service firms in the world and one of the "Big Four" accounting firms. It is a global organization of member firms in more than 140 countries with the headquartered in United Kingdom. Likewise it was highly ranked

by Forbes magazine as the one of the largest private company in the United States in 2011.

6.5.1 Social media uses

The SM is used by the company for the business purposes, but being a consulting firm it is used for communication, news transmission, company branding and recruitment. In Moscow office of the company is actively using Facebook, Twitter, LikedIn, less we use Youtube, on the international level also Google+.

For branding, the firm uses Facebook in order to maintain public relationship and tell more about the company itself in friendly way. Furthermore, the Facebook, LinkedIn, Google+ is used for the benchmarking with the competitors.

In other perspective, company is using social media for recruitment purposes to announce new vacancies, Olympiads, open positions, careers days on their page in Facebook and in LinkedIn. As well as, provides a lot of information about the company for candidates acknowledgment.

Company is posting in Facebook twice in a day in order to maintain the flow on the fan page. Facebook has more social aspect, than business function, company is sharing current news, conferences. As well as trying to represent the firm with the true person faces and show all creativity of personnel and all charity activities performed. Twitter is used less by department of marketing, however it is used for support of the real activities (conferences, business schools, seminars, publications).

Company is following the general purposes of the social media, and do not use any special metric for them.

The reports generated from the social media is done in sequence with the activities in order to announce and to track their performance. Overall reports are done irregularly.

6.5.2 Data analysis

The data is actually collected inside the company from the SM sources, instead it has being monitored by calculating number of "likes", "shares" and "comments", as well as the responsible person is actively participating in discussion and of course paying attention, who is commenting and how (positively/negatively) to the content in SM, trying to create significant traffic on the companies pages. Based on the manual data

collection company is also shaping its content in order to be more moving for action to receive the feedback. The main problem that it is new tool for the company and the usage and implementation is rather difficult.

However that the most of the information is collected manually, but company is also using Google analytics.

The SM is more true resource of information, as information is public and be easily commented by people, more direct, but is difficult for the company's business, directly bound up with the confidentiality problems and privacy law.

6.5.3 Uses of data in the company

The main user of the SM information is the marketing department, being a new tool for the company right now is being implemented. Moreover, in the big companies with the core business in the consulting sphere implementation of this kind of instruments taking time, but marketing department is seeking to implement and promote them.

6.6 Case # 6

This firm is an American multinational information technology corporation headquartered in the United States. It provides products, technologies, software, solutions and services to consumers, small- and medium-sized businesses and large enterprises, including customers in the government, health and education sectors.

It is one of the world's leading PC manufacturer. It specializes in developing and manufacturing computing, data storage, and networking hardware, designing software and delivering services. Major product lines include personal computing devices, enterprise, and industry standard servers, related storage devices, networking products, software and a diverse range of printers, and other imaging products. Its products to households, small- to medium-sized businesses and enterprises directly as well as via online distribution, consumer-electronics and office-supply retailers, software partners and major technology vendors. It also has strong services and consulting business around its products and partner products.

6.6.1 Social media uses

The social media sources are widely used by the company both for internal and external prospective. The firm has its own internal laboratory in the intranet of the company, as

well as "Jive" social business platform for communication, collaboration purposes, that also could be used for collaboration with clients in order to co-create and provide higher level of the support. Therefore, social business process management is used for the improving internal performance of the company. Moreover, company is using "Twitter", "LinkedIn" and "Yammer" customer satisfaction for receiving and processing feedbacks of the clients and resolve problems with the products, services provided.

Additionally, company is using the SM platforms for recruitment and acknowledgement about the current campaigns, as well as in order to promote company itself, creating long lasting relationships and loyalty of the brand.

The updates of the information is done 5-10 a day per feed, being a big multinational company with different service lines, the SM sources are managed internally of each service line, but important messages are spread all over the SM platforms. The reporting is done adequately and are provided on the daily and weekly basis to the person in charge.

6.6.2 Data analysis

Information received from the SM platforms is analyzed and proceed for the further interpretation and action taking in order to improve the sentiment and give the necessary responses to the clients questions. Moreover, for the internal purposes the special systems of "Bright idea" in order to find the best solution for the case is implemented, which analyzing all the proposed solutions and choosing the best one. As well as information collected from internal and external instruments could be represented in graphics, charts and proceed with the clustering, depending on the requirements of the department or task objective. The specific metrics are implemented according to the activities and the indicators are calculated.

6.6.3 Uses of data in the company

The main users of external SM information are the sales and marketing department directly working with these platforms for the client satisfaction and consumer services. Instead the information from the internal SM is very important for the internal processes, decision making and creation of new product or resolving current problems, Social business management process is inserted in the company for improving efficiency

inside of multinational company in order to be updated all over 100 countries where it is operating.

6.7 Case # 7

The firm is an Italian reality founded in 2006, that offers services related to information technology from design of web applications, management of community up to social media analysis. Barilla, Fiat, Microsoft, Nokia and Unilever are just some examples of the customers of this firm that works in international context.

6.7.1 Social media uses

The company employs Social media for projects with customers, to support marketing, sales and human resource department. The main objectives of projects is to work properly in relevant aspect of customer business in Social media, that could been for example the monitoring, marketing or web application development; while the marketing and sales department use this channel to promote and sell services, instead human resources employs Social media for recruitment purpose.

Depending on the purpose it is employed a different Social media platform. The most used are “Facebook”, “Twitter” and “Youtube”, but on demand or driven by the topic also more specific platform such as “Foursquare”, “Pinterest” or “Instagram”. The source nature is 60% from social network, 30% forum and 10% blogs.

The activities and updates on Social media platforms are daily, usually done in the morning or in the late afternoon with an average of 4 updates per day.

6.7.2 Data analysis

The data coming from Social media are gathered and analyzed separately for each platform. Actually the main tools of analysis are the applications and software integrated with the most known social network (i.e.“Twitter” and “Facebook”) such as “Crowdbooster” and “Topsy”, while it is done manually for the one that are not supported (i.e. “Pinterest” and “Instagram”).

The information usually is represented by the metrics that are different for each social platform. The KPIs have different metric depending by the source of measurement and the information nature that the platform is able to provide:

Source	KPI	Metric
Facebook	Potential viewer	Number of fan
	Talking about	Likes + Shares + Comments
Twitter	Potential viewer	Number of followers
	Talking about	Retweets + Favorites tweets + Replies
	Reach impression	Retweets + Favorites tweets + Replies + Followers
Youtube	Potential viewer	Number of subscriptions
	Talking about	Views + Likes + Comments

Table 23: Case # 7 - KPIs' metrics for social networks

It is performed a qualitative benchmarking on competitors profiles to get new ideas on marketing campaigns and contents communication. This kind of analysis is not structured and it is done manually depending on the necessities.

The reports of this analysis are basically of two types:

- Ordinary activities, which is done monthly and gives basically information about content trend in Social media.
- Extra-ordinary activities (i.e. marketing campaign for a product), which is done at the end of the activities giving indication about the sentiment of the users about the activity performed.

6.7.3 Uses of data in the company

Marketing and product department uses the data coming from Social media. The first one is able to drive the selling with low costs, promoting the company on the web, while the second one provide design and services up-to-date especially in teasing activities and digital PR.

The main advantage of the adoption of platforms is to have data about sentiment and opinions of the user of the web with a good level of integration with data analytics tools, on the other side there are loads of interesting data up other platforms that are unparsed due to the lack of universal analysis tools. In addition Web 2.0 is 24/7 service, this means high effort in crisis management and coordination among departments.

6.8 Case # 8

The firm taken as case study is a data intelligence consulting company, based in Milan, Italy. The company provides solutions and services related to marketing models, customized dashboards, data modeling, risk models and social CRM.

6.8.1 Social media uses

The firm uses Social media to support the projects with clients providing and analysing information. The main objective is to perform analysis whose outcomes are useful to support marketing and communication process. To pursue these objectives it has been chosen a set of social platform: blogs, new channel for experts opinion, dedicated forums, to evaluate opinions of people and Facebook, to listen to customers needs and comments. In addition in Italy it is growing up as listening platform Twitter that enables the share of contents through tags.

6.8.2 Data analysis

The monitoring of the activities is performed through own developed tools and open source programs. Classical approaches of data gathering is going to be substituted by the innovative one, represented by information coming from Social media. Appropriate metrics of measurement are developed depending on the goals and objectives that the client wants to achieve and of course depending on the object of control. It is quite difficult to define standard KPIs for the evaluation of communication process, because it depends a lot on the nature of information and the relation with the object of control that I can get from Social media.

The reports are usually weekly or monthly depending on the needs of the client, showing the results of analysis with indicators previously arranged. The frequency of the reports is correlated to the specific characteristic of social media information; basically the data get updated in brief in a relevant way.

6.8.3 Uses of data in the company

The consultant projects are performed mainly in manufacturing, distribution and financial company, usually for marketing and management. Actually not all companies look for a precise solution; most of them want to get an overview of potentiality of a structured approach of information analysis.

The main benefit is to obtain new kind of information, basically unique and high valuable, which can influence strategy of communication of the company. Unfortunately the absence of standard methodology for the manipulation and the analysis of data is a significant drawback, which creates misalignment of the final results of the analysis. Moreover, the associated commercial tools are not so consolidated, which means that instrument of analysis should be suited for each company.

6.9 Synthetic evidence: case studies

The table # 24 represents the synthetic analysis of the research done, specifically the 8 interviewers that had been carried out for current use of the social media inside the companies, taking into account the direct and consulting use:

Case #	Field	Market	SM used	Metric		Use		
				Objective	KPI	Department	Report	User of data
1	Manufacturing	Multinational	Facebook Twitter Instagram	Sales forecasting Benchmarking	Traffic Sentiment	Marketing	Weekly / Monthly	Marketing Country manager
2	Beverage	Multinational	Facebook Youtube Instagram Twitter	Product promotion Public awareness	# "Likes" # "Fan"	Marketing	Monthly	Marketing
3	Telecommunication	Multinational	Internal SN Facebook Twitter	Product co-creation Product promotion Customer service Recruitment Benchmarking	Sentiment Opinion mining # "Likes" # "Fan" # "New ideas"	Marketing Human resources Sales Management	Daily/ Weekly	Specialists Managers Department's head
4	IT conglomerate	Multinational	Internal SN Facebook Youtube LinkedIn Odnoklassniki VKontakte	Product co-creation Product promotion Listening platform Customer service Recruitment	Sentiment Opinion mining # "Likes" # "Fan" # "New ideas"	Marketing Human resources Sales Management	Daily/ Weekly/ Upon request	Specialists Managers Department's head

5	Consulting	Multinational	Facebook Youtube Twitter LinkedIn	Sales forecasting Listening platform Recruitment	# "Likes" # "Fan" Sentiment	Marketing	Randomly	Marketing
6	IT corporation	Multinational	Internal Lab Internal SN LinkedIn Yammer Twitter	Product co-creation Product promotion Listening platform Customer service Recruitment Social BPM	Sentiment Opinion mining # "Likes" # "Fan" # "New ideas"	Marketing Human resources Sales Management	Daily/ Weekly	Specialists Managers Department's head
7	IT agency	Italian	Facebook Twitter Instagram Pinterest Foursquare	Customer service Sales forecasting Product promotion Listening platform	Potential viewer Talking about Potential viewer Talking about Reach impression Potential viewer Talking about	Consulting Marketing Human resources Sales Management	Monthly/ Upon request	Customer Specialists Managers Department's head
8	Data intelligence	Italian	Facebook Forum Blog Twitter	Marketing Communication	Tailored by client's needs	Marketing Management	Weekly/ Monthly	Marketing Management

Table 24: Synthetic evidence of the interviews

The most of the companies interviewed are big multinational corporations, working in the different spheres of the business. However, we can see the trend that quantity of the purposes for which the social media platforms are used, are positively correlated with the information technology and telecommunication background of the companies, that use such platforms not only for analysis of the external sources, but also for the more efficient collaboration inside the companies taking into account the geographical distance between decade of their offices all over the world. Instead the small and medium companies are focalized on the external analysis and product promotion of their goods.

According to the social media platforms used we can clearly see from the table that top four are "Twitter", "Facebook", "Youtube" and "LinkendIn". In major part of the cases the purpose of the social network and its' usage are correlated positively, but more important issue that they are used as a set of instruments and "Facebook" are used as an information transmitter for all other social media. As regards, to the metrics they are following the purposes of the companies, therefore, mainly the listening platforms, product promotion and customer service, as well as the company brand image and loyalty of the consumers. The KPIs identified by the interview based on the simple quantity of "likes", "comments" and etc., but also on the information collected by the specific tools measuring "positive" and "negative" sentiment, helping companies to identify problems and react on time. Therefore, the reporting inside the firms are done on the basis of internal necessity and campaign timing and could vary from hourly to monthly delivery.

As mentioned before, the department which is using the most social media is marketing for right now, but sales and human resource departments following it, profiting from the specified instruments offered by "LinkedIn" and "Twitter". Overall, we can see that information from the social media is gaining positions in the balanced score cards of the companies, being innovative and providing up-to-date information in real time. Therefore, the main users of such information are the marketing specialists, managers and departments' heads.

As well as we would like to introduce the following graph and diagram that are representing of the usage of indicators focalizing on their structural characteristics used inside the company and giving general overview on the company's social media KPIs.

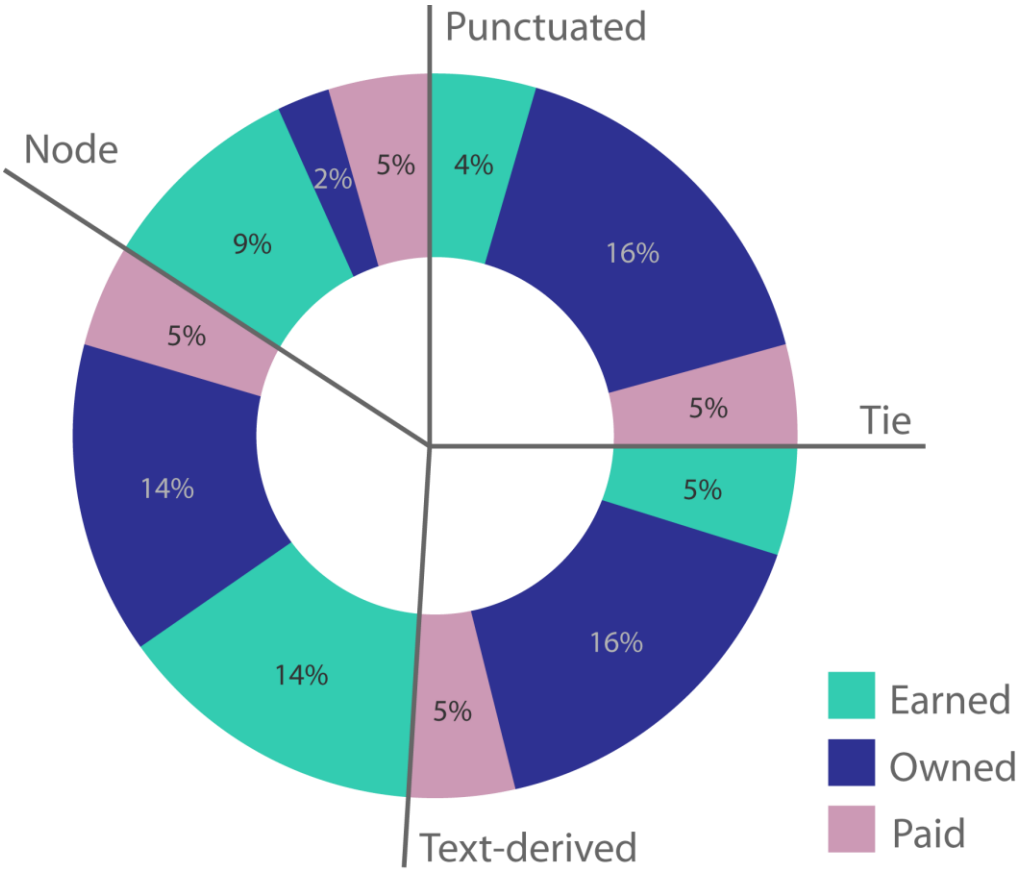

Figure 14: Indicators usage map

This indicators usage map is confronting and classifying indicators according to the proposed framework dividing them according to the structural characteristics in four main areas, such as Punctuated - Tie, Text-derived-Tie, Text-derived-Node and Punctuated-Node with the further division according to the source of the social media: Earned, Paid and Owned. Therefore, we can sum up that the Tex-derived-Node indicators have the leading positions in usage of the interviewees with the overall 31%, followed by the Text-derived-Tie indicators, according to the source the leading position is "Earned" with overall 48%, then "Owned" with 32% and last is "Paid" with 20%.

7 Experimentation

This chapter aims to present the results coming from empirical experimentation to apply the framework presented in a practical way.

7.1 Company overview

Euro-Fer dedicates its work, fruit of the passion and competence of the workers of a company with more than fifty years of experience, to everyone who can appreciate the works of quality and refined elegance, to everyone who loves the pureness of line of wrought iron and to everyone who wants to realize works that acquire value in time. Euro-Fer was born in the early 1960s as a small artisan's workshop business. To forge wrought iron, to work it, to shape it; these are the activities that the hand of man performs egregiously since the dawn of time. Euro-Fer was amongst the forerunners of this activity and the first to go abroad to propose its products. In the beginning several hundreds of articles were manufactured, but now we are talking about thousands of different pieces to create railings, gates, balconies and stairways. The combinations are infinite and can be harmonized with any architectural style. In 2012, Euro-Fer has celebrated forty-five years of activity as an industrial reality, although it existed as an artisan's workshop from the beginning of the 1960s. Thanks to the intuition of its founder, Tomaso Pretto, Euro-Fer is today amongst the world leaders of the sector, top seller in Italy and various European countries. The investments that Euro-Fer makes are continuous and go from the perfecting of existing technologies to the creation of new machinery, from project studies that are effectuated by well-known designers to the implementation of new structures that improve logistics: for example, just one new automatic warehouse for all articles. The production and the storage area for finished and semi-finished products have already reached 35.000 square metres. There are approximately two hundred employees at Castelgomberto and this data testifies to the company's intention to maintain and increase its production in Italy because its product is "Made in Italy".

7.2 Euro-Fer requirements

The metrics are intended to support the marketing department to take decision on two different issue: the first one is the decision to get, so the developing of Social media strategy in terms of platform and contents; while the second one is focused on the

possibility to enter in a new market. They are two decisions to be taken that can be supported by the analysis of the two markets.

Here it is been defined two main object of control: the Italian and the Russian market, evaluating not only the business opportunity, but also the potential market and the products feedback. The metrics created are somehow similar, but referring to different object of control independent one from each other. For this reason the analysis has been carried on separately.

7.3 Italian market: setting up

This phase set the parameters for the following analysis, so for the Italian market the object of the analysis is the “wrought iron” couple with its alternative “Stainless steel”. The language for the analysis is Italian; this brings to the definition of keywords and taxonomy useful for the analysis performed.

According to the business of Euro-Fer has been selected as keywords: “ferro battutto” and “acciaio inox”, while to generate the taxonomy has been defined 5 clusters of interest: kind of decoration, name of company and main competitors, final products, possible combination with other material and people involved in the selling process.

Cluster	Features
Decoration	arredamento
	caminetto
	ferro
	inox
	lancia
	borchia
	forgiato
	arredo
	interno
	esterno
	ricalcato
	inox
Company	Euro-Fer
	Mazzacani
	Lyuiforging
	Rik-fer
	Feba
	Fer-bat
	Fatih
	Sahinler-forge
	Lecky
	INDIA
	Euro-Fer
Products	maniglia
	cassetta postale
	scorritenda
	attaccapanni
	cimasa
	ruota
	cardine
	ribattino
	scala
	parapetto
	balcone
	recinzione
	cancello
	ringhiera
	corrimano
	partenza
	tubo
	fune
	torciglione
	fascia
	barra
	barra forata
	barra sagomata
	pigna
Combination	Vetro
	Legno
	Swarovski
People	Architetto
	Fabbro
	installatore

Table 25: Taxonomy of Italian words

7.4 Italian market: data gatheing

The source analysis come up with 39 sources listed in the table belonging to different kind of social media: 10 are websites owned by the competitor, 17 are articles or blogs, 7 are forums and 4 are social networks.

Name of Source	Type of Social media
http://www.timelesswroughtiron.com/	Competitor website
http://www.arteferro.com/arteferro	Competitor website
http://www.fatih.com.tr	Competitor website
http://www.sahinler-forge.com	Competitor website
http://www.lecky.it/default.asp	Competitor website
www.luyiforging.com	Competitor website
www.rik-fer.com	Competitor website
http://www.feba.it	Competitor website
http://www.fer-bat.it/	Competitor website
http://www.mazzacani.it	Competitor website
http://www.ferroearte.it/	articles &Blogs
http://www.edilportale.com/prodotti/arredamento-e-design/766	articles &Blogs
http://www.archiproducts.com/	articles &Blogs
http://news.archiexpo.it	articles &Blogs
http://architettura24.com	articles &Blogs
http://www.designmag.it	articles &Blogs
http://mobili-arredamento.blogspot.it	articles &Blogs
http://www.ideare-casa.com	articles &Blogs
http://blog.lavorincasa.it/	articles &Blogs
http://blog.lavorincasa.it/	Forum
http://www.info-arredamento.com/arredare/scale_in_ferro_battuto.html	articles &Blogs
http://ferrobattutoarredamento.blogspot.it/	articles &Blogs
http://blog.libero.it/FerroePassione/	articles &Blogs
http://rogiamblog.blogspot.it	articles &Blogs
http://www.undicilandia.it	articles &Blogs
http://www.abitare.it	articles &Blogs
http://atcasa.corriere.it	articles &Blogs
http://www.casa24.ilsole24ore.com	articles &Blogs
http://www.luxgallery.it/tende-da-sole-di-lusso-34803.php	Forum
http://www.ferrobattuto.net/sedie-ferro-battuto-bar-trento-centro/	Forum
http://www.arredamento.it/	Forum
http://tempiemodi.com/2012/05/24/open-office/	Forum
http://wwwferroearteit.globalfreeforum.com/viewforum.php?f=1	Forum
http://www.ciao.it/	Forum
www.pinterest.com	Social network
www.linkedin.com	Social network
www.twitter.com	Social network
http://48ers.com/	Social network

Table 26: Italian fonts selection

Not all the sources has been used for the crawling, but depending by the quantity and quality of data, it has been selected 4 out of 39 sources giving the following results:

1. http://blog.lavorincasa.it/ - Blog, this source provides opinion by expert such as architect upon internal and external design products. The keyword "ferro battuto" offered 282 articles while "acciaio inox" 653. The final amount of articles coming from this website is 935.
2. http://blog.lavorincasa.it/ - Forum, this source provides user opinion instead of expert about internal and external design products, giving the opportunity of discuss on the solutions and possible uses. While the blog is more guided by professional author, here the community is the main character identifying key author. The keyword "ferro battuto" offered 625 articles while "acciaio inox" 1861. The final amount of articles coming from this website is 2486.
3. http://www.arredamento.it/ - Forum, one of largest forum for internal and external design in Italy. The keyword "ferro battuto" offered 25921 articles while "acciaio inox" 39600. The final amount of articles coming from this website is 65521.
4. http://wwwferroearteit.globalfreeforum.com/viewforum.php?f=1 - Forum, this source is focused on the products in wrought iron, here there is the presence both of the end user and the producer. The keyword "ferro battuto" offered 654 articles while "acciaio inox" 12. The final amount of articles coming from this website is 666.

The total number of crawled comments for the keywords "ferro battuto" is 27482 while for "acciaio inox" is 42126. The comments are arranged in a metadata table such as the one show in the picture [] that represent the map of the whole data:

ID	FONTE	TITOLO	TESTO	DATA	AUTORE
FB1	http://blog.lavorincasa.it/	Ferro battuto	Si definisce ferro battuto il ferro sottoposto a lavora	NA	Arch. Carmen Granata
FB2	http://blog.lavorincasa.it/	Letti in ferro battuto	Un tocco di romanticismo d'altri tempi. E' quello che	NA	Barbara Machin
FB3	http://blog.lavorincasa.it/	Complementi in ferro battuto	Il giardino e la terrazza sono spazi da vivere che dive	NA	NULL
FB4	http://blog.lavorincasa.it/	Restauro di ringhiere in ferro battuto	Le ringhiere di ferro battuto, che contraddistinguono	NA	Arch. Emanuele Distefano
FB5	http://blog.lavorincasa.it/	Sedie Savonarola contemporanee	Probabilmente cercando in soffitta o in qualche casa	NA	Arch. Valentina Caiazzo
FB6	http://blog.lavorincasa.it/	Montare una tenda da interno	La tenda che arreda!Molto spesso la tenda da intern	NA	NULL
FB7	http://blog.lavorincasa.it/	Arredamento in arte povera	Arredare gli ambienti della casa con mobili in arte po	NA	Tiziana Loprete
FB8	http://blog.lavorincasa.it/	Maniglie per interni	Le maniglie per gli interni sono oggetti semplici ed es	NA	NULL
FB9	http://blog.lavorincasa.it/	Abat-jour classiche e moderne	Eleganti, classiche, decorate o dallo stile semplice. In	NA	Tiziana Loprete
FB10	http://blog.lavorincasa.it/	Stile etnico - Cow Boy	L'America è sicuramente il continente che mostra la	NA	NULL
FB11	http://blog.lavorincasa.it/	Tipologie di letto	Esistono diverse tipologie di letto e la scelta va fatta,	NA	Arch. Carmen Granata
FB12	http://blog.lavorincasa.it/	Bagno classico	Stanchi della pulizia estrema e del minimalismo asso	NA	Francesca Gobbo
FB13	http://blog.lavorincasa.it/	Stile etnico	Dopo un viaggio in qualche meta esotica ci si porta a	NA	NULL
FB14	http://blog.lavorincasa.it/	Arredamont a Longarone	Baricentro tutto spostato a nord, in provincia di Bell	NA	Arch. Nora Santonastaso
FB15	http://blog.lavorincasa.it/	Tende da interni	A volte non si da' troppa importanza alla scelta delle	NA	NULL
FB16	http://blog.lavorincasa.it/	Cucina su misura	Disposizione delle varie zoneUna cucina ben organiz	NA	NULL

Figure 15: Example of crawled comments for Italian fonts

Then applied filters properly, the number of comments has been reduced with the quantity reported in the table # 27 divided by sources and keywords:

ID source	Source	Number of comments	
		"ferro battuto"	"acciaio inox"
IT1	http://blog.lavorincasa.it/ - Blog	30	342
IT2	http://blog.lavorincasa.it/ - Forum,	267	726
IT3	http://www.arredamento.it/ - Forum	1402	3157
IT4	http://wwwferroearteit.globalfreeforum.com - Forum	589	0
	Total number of comments	**2288**	**4225**

Table 27: Quantity of comments per font in Italian

7.5 Italian market: data analysis

All the analyses performed are reported in disaggregated way due to the different nature of the sources and the quantity of comments. In addition the sources are few, this make the aggregate analysis just a gross approximation of the 4 analyses separated.

7.5.1 Analysis of the source "Http://Blog.Lavorincasa.It/ - Blog" - IT1

The analysis is divided in two parts, the first one is referred to the keywords "ferro battuto" while the second one is about "acciaio inox".

7.5.1.1 Keyword "Ferro Battuto"

Here there is the top ten of most frequent terms that identifies the main topics:

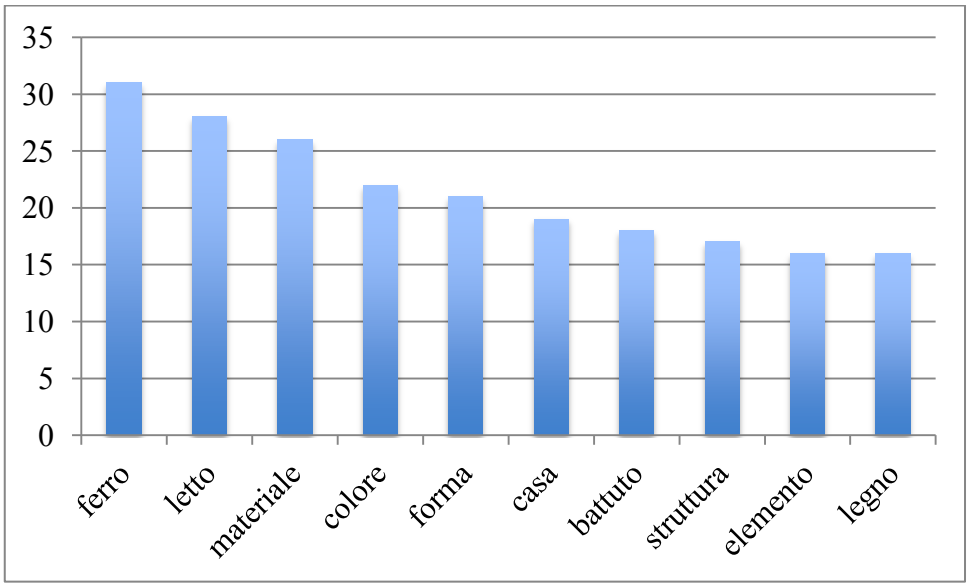

Figure 16: 10 most frequent terms for the keyword "ferro battuto" in IT1 font

The word cloud plotted gives at once the main themes:

Figure 17: Word cloud for the keyword "ferro battuto" in IT1 font

The study of words association identified two words with a good level of correlation with the keywords “ferro battuto”: “lamiera” and “riccioli”. While in the taxonomy it has been matched 42 terms in “Decoration” and 12 terms in “Products”.

Cluster	Feature	Number of matches
Decoration	arredamento	2
	ferro	31
	arredo	9
Products	maniglia	1
	cardine	1
	parapetto	2
	recinzione	2
	cancello	1
	ringhiera	2
	partenza	1
	tubo	1
	barra	1

Table 28: Taxonomy clusters for the keyword "ferro battuto" in IT1 font

It is been created clusters trying to identify patterns and topics.

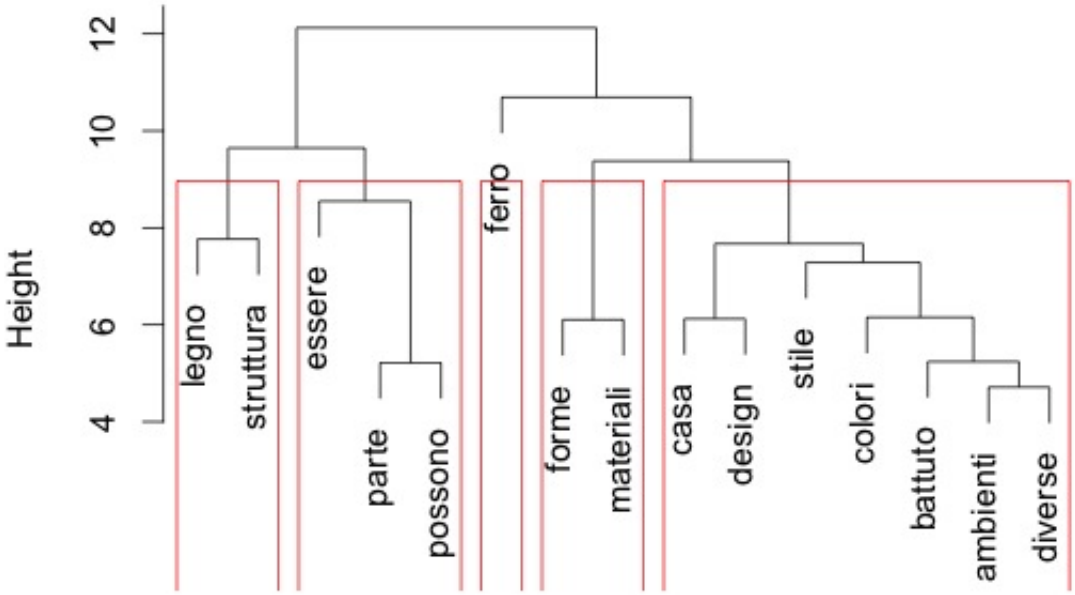

Figure 18: Clusters for the keyword "ferro battuto" in IT1 font

Regarding sentiment analysis, the polarity means of all the documents 0.5767 in a range between [-1,1], so the source talks in a positive way about "ferro battuto". The number of sentences in which it has been evaluated successfully is 564 (432 positive, 64 neutral, 68 negative) out of 800.

7.5.1.2 Keyword "Acciaio Inox"

Here there is the top ten of most frequent terms that identifies the main topics:

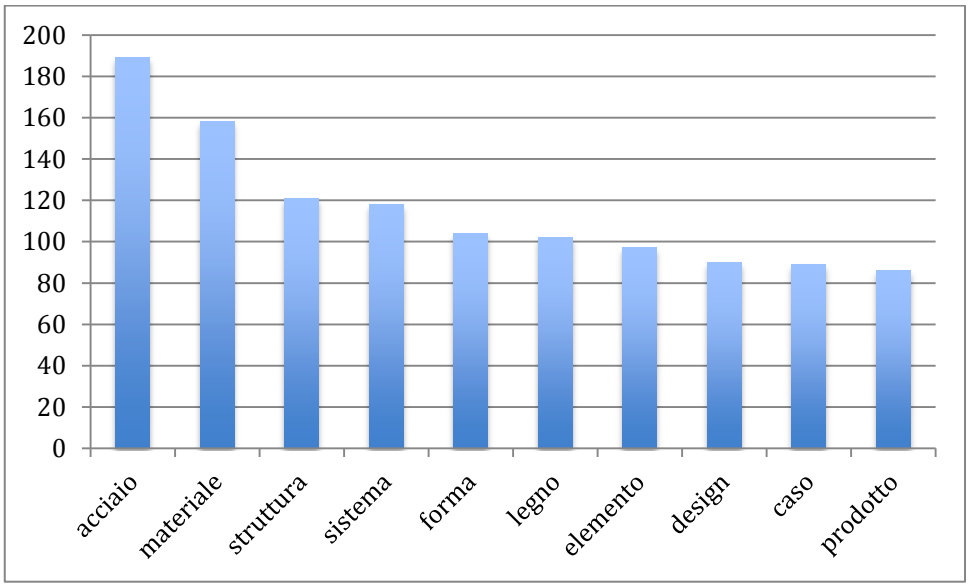

Figure 19: 10 most frequent terms for the keyword "acciaio inox' in IT1 font

The word cloud plotted gives at once the main themes:

Figure 20: Word cloud for the keyword "acciaio inox" in IT1 font

The study of words association doesn't identify any words with a good level of correlation with the keywords "acciaio inox". While in the taxonomy it has been matched 60 terms in "Decoration", 125 terms in "Products" and 1 term in "People".

Cluster	Feature	Number of matches
Decoration	arredamento	7
	caminetto	1
	ferro	28
	lancia	1
	arredo	23
Products	maniglia	17
	attaccapanni	2
	ruota	9
	scala	46
	parapetto	3
	balcone	3
	recinzione	7
	cancello	1
	ringhiera	5
	partenza	1
	tubo	17
	fascia	5
	barra	9
People	installatore	1

Table 29: Taxonomy clusters for the keyword "acciaio inox" in IT1 font

It is been created clusters trying to identify patterns and topics.

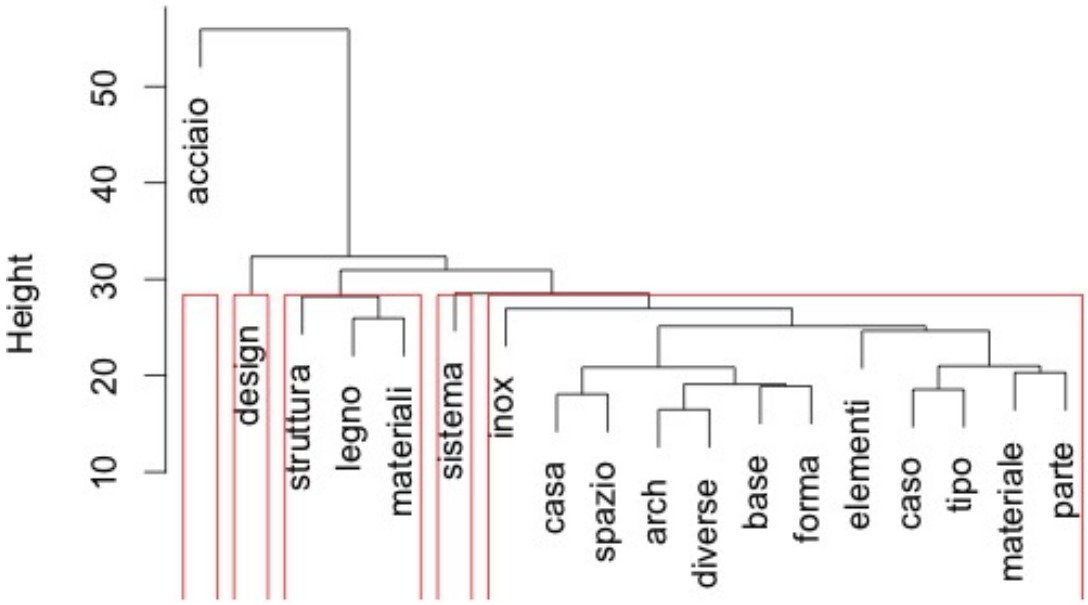

Figure 21: Clusters for the keyword "acciaio inox" in IT1 font

Regarding sentiment analysis, the polarity means of all the documents 0.5336 in a range between [-1,1], so the source talks in a positive way about “acciaio inox”. The number of sentences in which it has been evaluated successfully is 4365 (3223 positive, 503 neutral, 639 negative) out of 6130.

7.5.2 Analysis of the source “Http://Blog.Lavorincasa.It/ - Forum” - IT2

The analysis is divided in two parts, the first one is referred to the keywords “ferro battuto” while the second one is about “acciaio inox”.

7.5.2.1 *Keyword "Ferro Battuto"*

Here there is the top ten of most frequent terms that identifies the main topics:

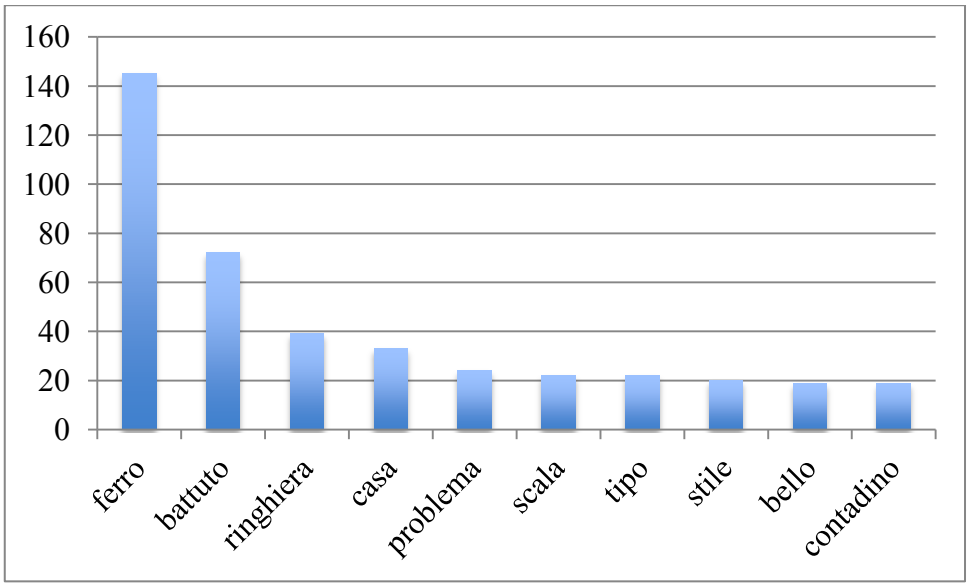

Figure 22: 10 most frequent term for the keyword "ferro battuto" in IT2 font

The word cloud plotted gives at once the main themes:

Figure 23: Word cloud for the keyword "ferro battuto" in IT2 font

The study of words association identified two words with a good level of correlation with the keywords “ferro battuto”: “brand” and “classico”. While in the taxonomy it has been matched 149 terms in “Decoration” and 92 terms in “Products”.

Cluster	Features	Number of matches
Decoration	arredamento	4
	ferro	145
Products	maniglia	1
	scala	22
	parapetto	4
	balcone	8
	recinzione	2
	cancello	9
	ringhiera	39
	corrimano	2
	tubo	4
	barra	1

Table 30: Taxonomy clusters for the keyword "ferro battuto" in IT2 font

It is been created clusters trying to identify patterns and topics.

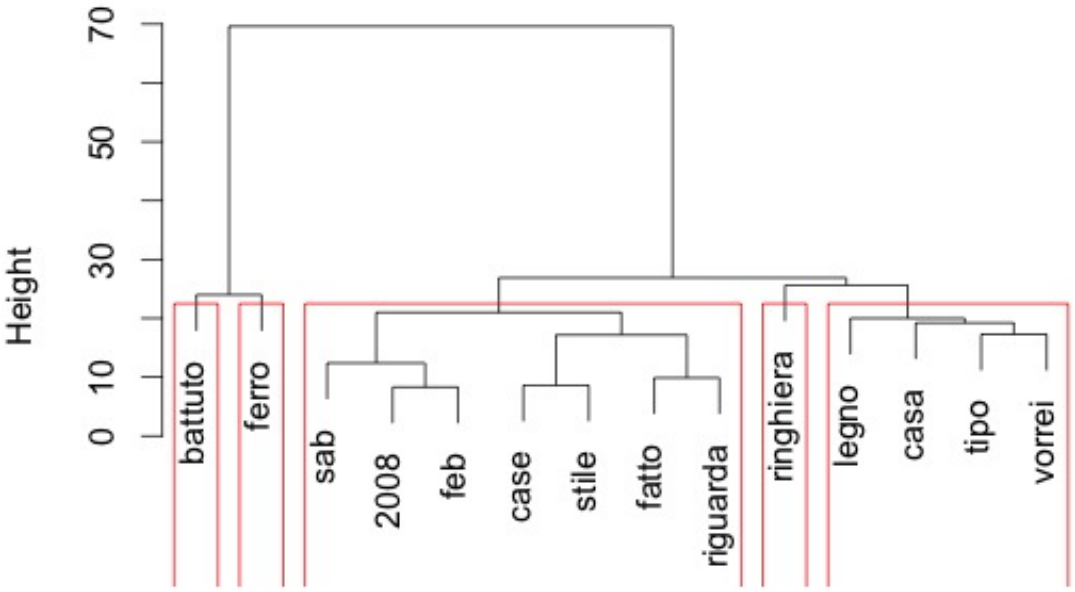

Figure 24: Clusters of the terms for the keyword "ferro battuto" in IT2 font

Regarding sentiment analysis, the polarity means of all the documents 0.0847 in a range between [-1,1], so the source talks in a neutral way about “ferro battuto”, in fact the number is so near to the 0, this means that there are both positive and negative comments. The number of sentences in which it has been evaluated successfully is 725 (352 positive, 85 neutral, 288 negative) out of 1643.

7.5.2.2 *Keyword "Acciaio Inox"*

Here there is the top ten of most frequent terms that identifies the main topics:

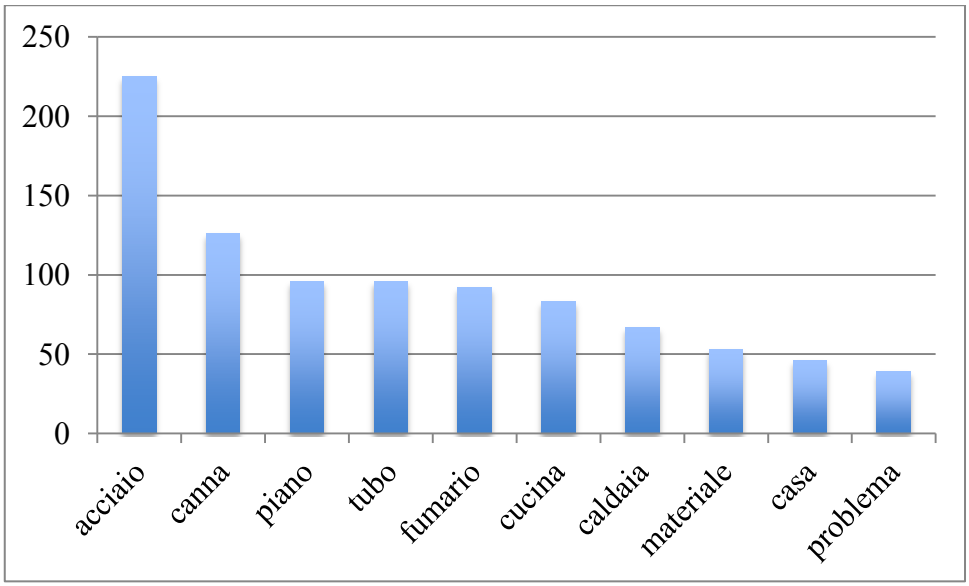

Figure 25: 10 most frequent terms for the keyword "acciaio inox" in IT2 font

The word cloud plotted gives at once the main themes:

Figure 26: Word cloud for the keyword "acciaio inox" in IT2 font

The study of words association doesn't identify any words with a good level of correlation with the keywords "acciaio inox". While in the taxonomy it has been matched 34 terms in "Decoration", 122 terms in "Products" and 2 terms in "People".

Cluster	Features	Number of matches
Decoration	caminetto	5
	ferro	27
	arredo	2
Products	ruota	2
	scala	7
	balcone	5
	ringhiera	6
	corrimano	3
	tubo	96
	fascia	2
	barra	1
People	installatore	2

Table 31: Taxonomy clusters for the keyword "acciaio inox" in IT2 font

It is been created clusters trying to identify patterns and topics.

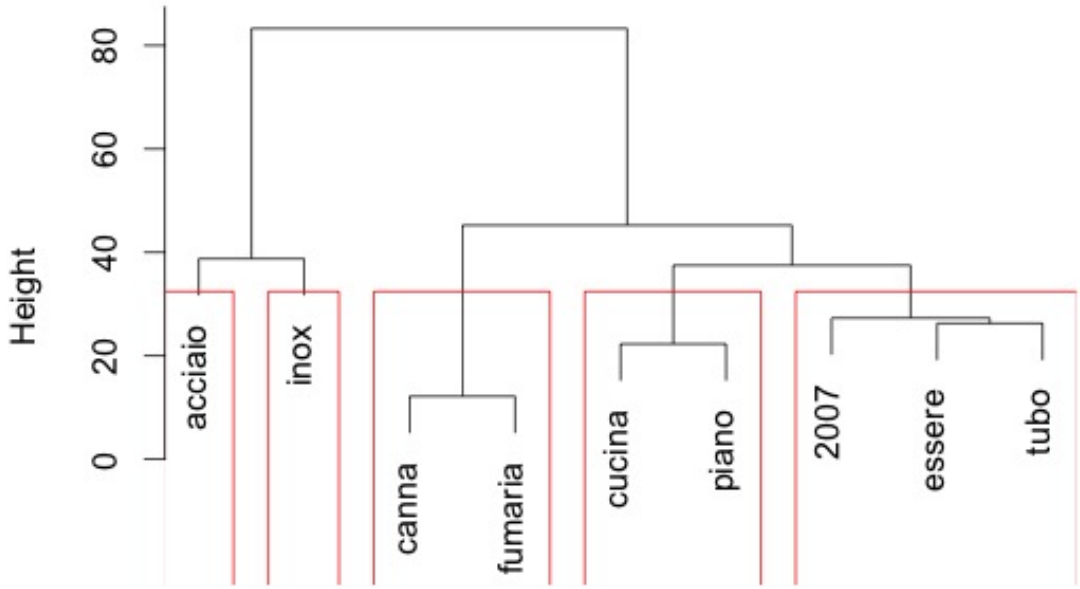

Figure 27: Clusters for the keyword "acciaio inox" in IT2 font

Regarding sentiment analysis, the polarity means of all the documents 0.2818 in a range between [-1,1], so the source talks in a positive way about "acciaio inox". The number of sentences in which it has been evaluated successfully is 1559 (935 positive, 223 neutral, 401 negative) out of 2989.

7.5.3 Analysis of the source “Http://Www.Arredamento.It/ - Forum” - IT3

The analysis is divided in two parts, the first one is referred to the keywords “ferro battuto” while the second one is about “acciaio inox”.

7.5.3.1 Keyword “Ferro Battuto”

Here there is the top ten of most frequent terms that identifies the main topics:

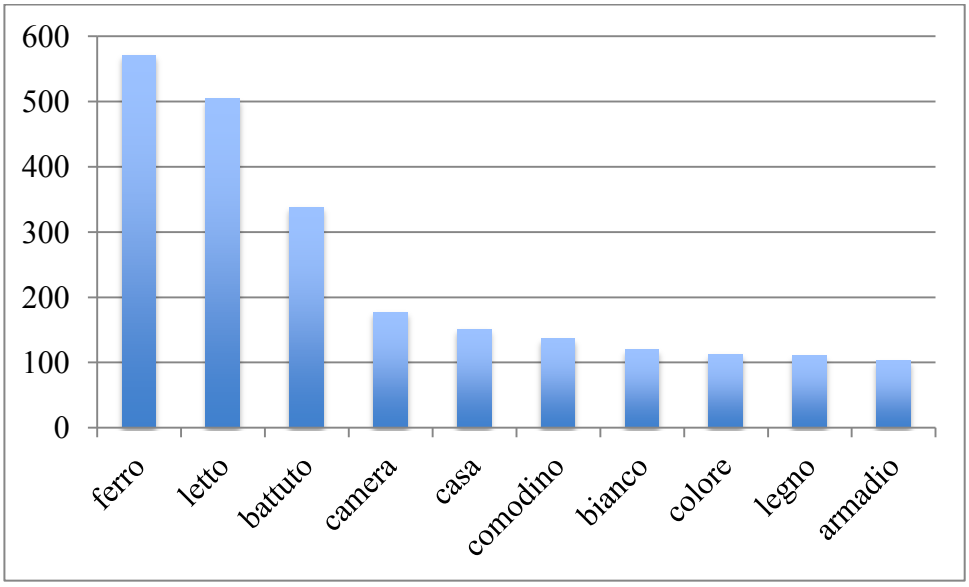

Figure 28: 10 most frequent terms for the keyword "ferro battuto" in IT3 font

The word cloud plotted gives at once the main themes:

Figure 29: Word cloud for the keyword "ferro battuto" in IT3 font

In the taxonomy it has been matched 601 terms in “Decoration” and 120 terms in “Products”.

Cluster	Features	Number of matches
Decoration	arredamento	22
	caminetto	1
	ferro	571
	arredo	7
Products	maniglia	13
	attaccapanni	1
	scala	38
	parapetto	3
	balcone	7
	cancello	1
	ringhiera	34
	corrimano	13
	tubo	3
	fascia	4
	barra	3

Table 32: Taxonomy clusters for the keyword "ferro battuto" in IT3 font

It is been created clusters trying to identify patterns and topics.

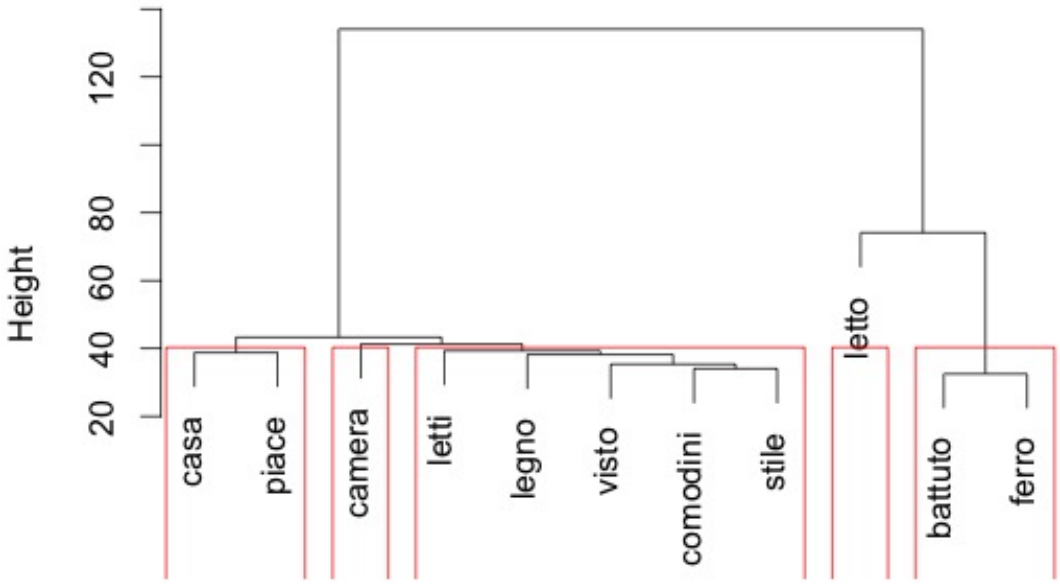

Figure 30: Clusters for the keyword "ferro battuto" in IT3 font

Regarding sentiment analysis, the polarity means of all the documents 0.117 in a range between [-1,1], so the source talks in a neutral way about "ferro battuto", in fact the number is so near to the 0, this means that there are both positive and negative comments. The number of sentences in which it has been evaluated successfully is 3816 (1925 positive, 504 neutral, 1387 negative) out of 8850.

7.5.3.2 Keyword "Acciaio Inox"

Here there is the top ten of most frequent terms that identifies the main topics:

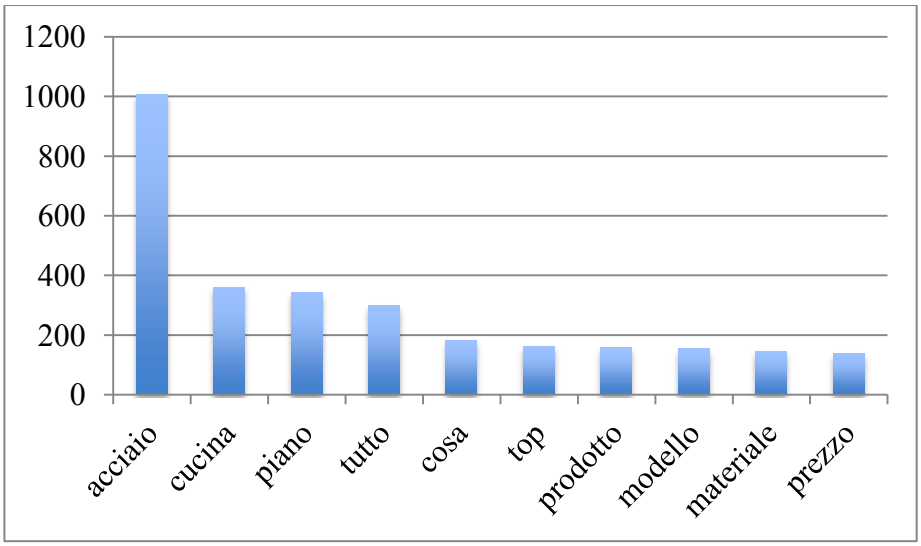

Figure 31: 10 most frequent terms for the keyword "acciaio inox" in IT3 font

The word cloud plotted gives at once the main themes:

Figure 32: Word cloud for the key word "acciaio inox" in IT3 font

The study of words association doesn't identify any words with a good level of correlation with the keywords "acciaio inox". While in the taxonomy it has been matched 176 terms in "Decoration", 248 terms in "Products" and 4 terms in "People".

Cluster	Features	Number of matches
Decoration	arredamento	4
	ferro	57
	borchia	2
	arredo	7
	interno	64
	esterno	42
Products	maniglia	57
	cardine	1
	scala	66
	parapetto	16
	balcone	3
	cancello	4
	ringhiera	44
	corrimano	3
	partenza	6
	tubo	21
	fune	1
	fascia	13
	barra	13
People	installatore	4

Table 33: Taxonomy clusters for the keyword "acciaio inox" in IT3 font

It is been created clusters trying to identify patterns and topics.

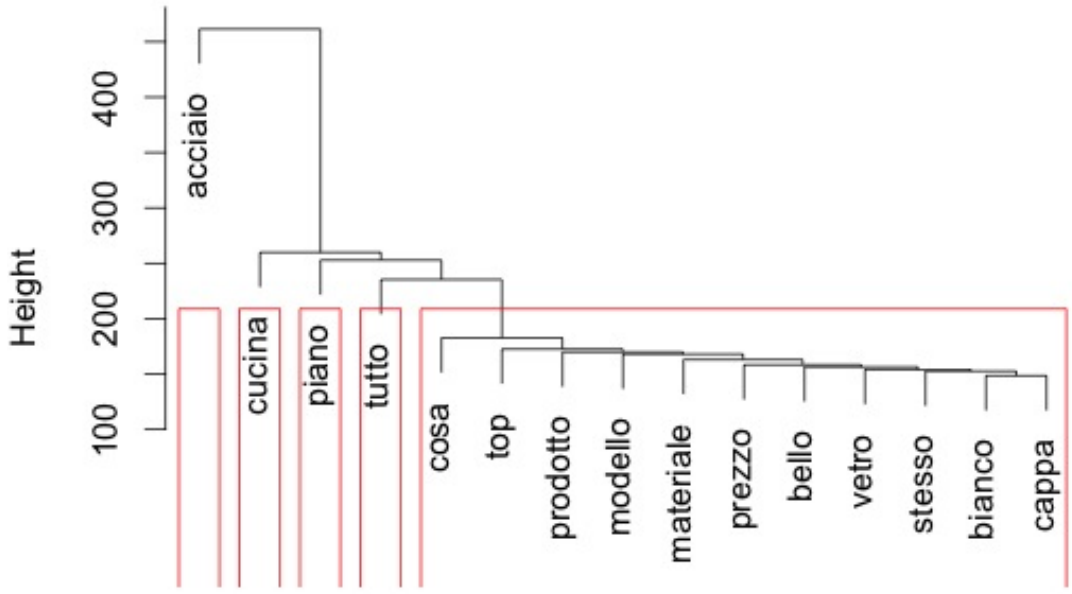

Figure 33: Clusters for the keyword "acciaio inox" in IT3 font

Regarding sentiment analysis, the polarity means of all the documents 0.345 in a range between [-1,1], so the source talks in a positive way about “acciaio inox”. The number

of sentences in which it has been evaluated successfully is 6230 (3919 positive, 732 neutral, 1579 negative) out of 14308.

7.5.4 Analysis of the source "Http://Wwwferroearteit.Globalfreeforum.Com" - IT4

Below provided the analysis just for a single keyword because the data for the reliable analysis of the second word if not sufficient.

7.5.4.1 Keyword "Ferro Battuto"

Here there is the top ten of most frequent terms that identifies the main topics:

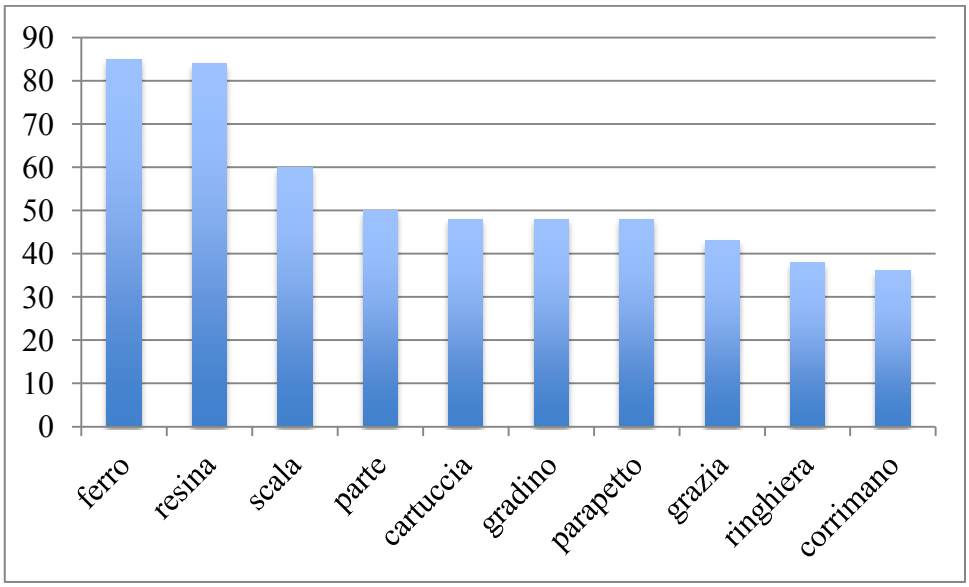

Figure 34: 10 most frequent terms for the keyword "ferro battuto" in IT4 font

The word cloud plotted gives at once the main themes:

Figure 35: Word cloud for the keyword "ferro battuto" in IT4 font

The study of words association identified two words with a good level of correlation with the keywords “ferro battuto”: “tradizione” and “esposizione”. In the taxonomy it has been matched 86 terms in “Decoration” and 290 terms in “Products”.

Cluster	Features	Number of matches
Decoration	ferro	85
	lancia	1
Products	cardine	26
	scala	60
	parapetto	48
	cancello	22
	ringhiera	38
	corrimano	36
	partenza	12
	tubo	36
	barra	12

Table 34: Taxonomy clusters for the keyword "ferro battuto" in IT4 font

It is been created clusters trying to identify patterns and topics.

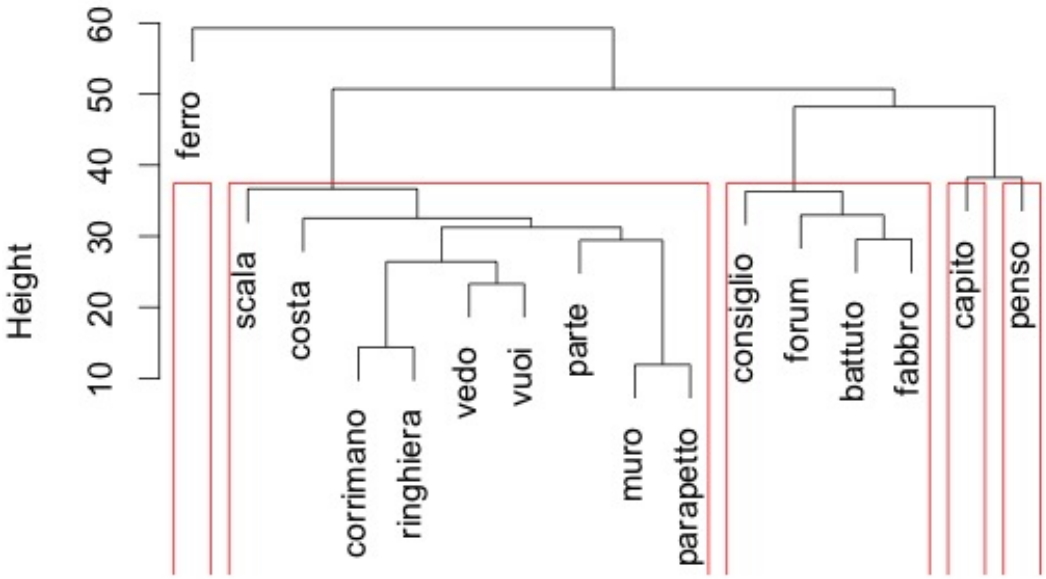

Figure 36: Clusters for the keyword "ferro battuto" in IT4 font

Regarding sentiment analysis, the polarity means of all the documents 0.2006 in a range between [-1,1], so the source talks slightly in a positive way about “ferro battuto". The number of sentences in which it has been evaluated successfully is 951 (508 positive, 110 neutral, 333 negative) out of 2180.

7.5.5 Italian market: synthetic evidence

The table # 35 represents the synthetic analysis of the data collection and analysis for the Italian market for the keyword "Ferro battuto":

Source	Comments	Filtered comments	Phrases	Sentiment	Absolute frequency		Taxonomy frequency		Cluster	
					Word	#	Word	#	Name	#
Http://Blog.Lavorincasa.It/ - Blog	282	30	564	0,5767	ferro	31	ferro	31	Decoration	42
					letto	28	arredo	9		
					materiale	26	arredamento	2		
http://blog.lavorincasa.it/ - Forum	625	267	725	0,0847	ferro	145	ferro	145	Decoration	149
					battuto	72	ringhiera	39		
					ringhiera	39	scala	22		
http://www.arredamento.it/ - Forum	25921	1402	3816	0,117	ferro	571	ferro	571	Decoration	601
					letto	504	scala	38		
					battuto	338	ringhiera	34		
http://wwwferroearteit.globalfreeforum.com - Forum	654	589	951	0,2006	ferro	85	ferro	85	Products	290
					resina	84	scala	60		
					scala	60	parapetto	48		

Table 35: Data synthetic evidence for keyword "ferro battuto"

Overall, analysis of the data for the keyword "ferro battuto" gave rather good results. We have collected and analyzed more than 26 thousands of comments with outcome of 5 thousands phrases. The sentiment of the "ferro battuto" on these sources is positive, but value is quite slight. Further, it is important to point out that during the clustering we can see domination of the "Decoration" cluster in the frequency analysis, that provides values from 42 to 601 terms in the analyzed phrases. The most frequent word for both absolute and taxonomy analysis is "ferro", which confirms the search result of the data set. High frequency of the word "scala" is indicating on the most popular use of the "ferro battuto" in these blogs and forums, but providing contradicting result with the cluster, being included into the "Products" cluster. Nevertheless, we could preserve that for the such products as "letto", "ringhiera" these channels could be used also.

The table # 36 represents the synthetic analysis of the data collection and analysis for the Italian market for the keyword "acciaio inox":

Source	Comments	Filtered comments	Phrases	Sentiment	Absolute frequency		Taxonomy frequency		Cluster	
					Word	#	Word	#	Name	#
Http://Blog.Lavorincasa.It/ - Blog	653	342	4365	0,5336	acciaio	189	scala	46	Products	125
					materiale	158	ferro	28		
					struttura	121	arredo	23		
http://blog.lavorincasa.it/ - Forum	1861	726	1559	0,2818	acciaio	225	tubo	96	Products	122
					canna	126	ferro	27		
					piano	96	scala	7		
http://www.arredamento.it/ - Forum	39600	3157	6230	0,345	acciaio	1008	scala	66	Products	248
					cucina	360	interno	64		
					piano	342	ferro	57		
http://wwwferroearteit.globalfreeforum.com - Forum	12	0	0	0	-	-	-	-	-	-

Table 36: Data synthetic evidence for keyword "acciaio inox"

Instead, for the keyword "acciaio inox" we have analyzed more than 40 thousand of the comments from the 3 sources, as the FerroArteIt forum was giving insufficient results in terms of data quantity. However, number of the filtered and finished phrases is around 12 thousands phrases, which are providing quantity reliable results. Generally, the sentiment of the "acciaio inox" is positive and the value slightly higher then for the "ferro battuto" keyword. And also here we can see the domination of one cluster "Products" and subsequent high frequency of the words "scala", "tubo" and "materiale". As well as the prevailing frequency of "acciaio" in both absolute and taxonomy results showing correlation with the objective of the data set collection.

7.6 Russian market: setting up

This phase set the parameters for the following analysis, so for the Russian market the object of the analysis is the “metal furniture and decorations”. The language for the analysis is Russian; this brings to the definition of keywords and taxonomy useful for the analysis performed.

According to the business of Euro-Fer has been selected keywords: “Ковка” and "Металл" while to generate the taxonomy has been defined 4 clusters of interest: kind of decoration, final products, possible combination with other material and people involved in the selling process.

Cluster	Features
Decoration	мебель
	камин
	чугун
	сталь
	пика
	заклепка
	кованый
	внешний
	внутренний
Products	ручка
	почтовый ящик
	карниз
	вешалка
	петля
	пика
	заклепка
	клепка
	парапет
	балкон
	корпус
	ворота
	перила
	лестница
	решетка
	поручень
	вылет
	прут
	труба
	конус
Combination	дерево
	сваровски
	стекло
People	дизайнер
	строитель
	прораб
	архитектор

Table 37: Taxonomy of Russian words

7.7 Russian market: data gathering

The source analysis come up with 13 sources listed in the table belonging to different kind of social media: 5 are websites owned by the competitors, 2 are forums, 1 is a Blog and 5 are social networks.

Name	Type of Social media
http://www.timelesswroughtiron.com	Competitor website
http://www.fatih.com.tr	Competitor website
http://www.sahinler-forge.com	Competitor website
http://www.lecky.it/default.asp	Competitor website
www.luyiforging.com	Competitor website
http://forum.homeideas.ru/	Forum
http://forum.ivd.ru/	Forum
http://360.ru/	Blog and Articles
www.pinterest.com	Social network
www.linkedin.com	Social network
www.twitter.com	Social network
www.vk.com	Social network
http://48ers.com/	Social network

Table 38: Russian fonts selection

Not all the sources has been used for the crawling, but basing on the quantity and quality of data, it has been selected 3 out of 13 sources giving the following results:

- http://forum.homeideas.ru/ - this source is leading forum for the interior and external design layout and design of building materials, equipment and products. It provides users' opinion of experts and ordinary people about internal and external design products, giving the opportunity to discuss the solutions and possible uses. The keyword “Ковка” offered 790 articles while “Металл” 700. The final amount of articles coming from this website is 1490.
- http://forum.ivd.ru/ - this source is devoted to discussion of renovation, remodeling interior and external design, technologies and products to make renovation. The articles of this source provide user opinion and expert opinions about internal and external design products, house, apartment renovations, giving the opportunity of discuss on the solutions and possible uses. The

keyword “Ковка” offered 3442 articles while “Металл” 14222. The final amount of articles coming from this website is 17664.

- http://360.ru/ - this source is blog and catalogue of products for internal and external design from the best world manufactures. It provides the detailed description of the products and giving opportunity to discuss its characteristics, possible uses, advantages and disadvantages. Of course this blog and catalogue is guided by professional team of authors, but give possibility to comment. The keyword “Ковка” offered 145 articles while “Металл” 138. The final amount of articles coming from this website is 283.

The total number of crawled comments for the keyword “Ковка” is 4377 and for keyword "Металл" is 15060. The comments are arranged in a metadata table such as the one show in the figure # 37 that represent the map of the whole data:

Id	Source	vtitolo	vdata	vautore	vbody
1	http://360.ru/	кованое зеркало+столик	05 Апрель 2012 - 11:54	IZM	Отправлено 05 Апрель 2012 - 11:543еркало в кованой раме Ш100*В122 в самой
2	http://360.ru/	NA	16 Апрель 2012 - 09:17	IZM	Отправлено 16 Апрель 2012 - 09:1730000=ipb.global.registerReputation('rep_post
3	http://360.ru/	NA	14 Август 2012 - 12:33	IZM	Отправлено 14 Август 2012 - 12:333аинтересована отдать в ближайшее время, т
4	http://360.ru/	Продам Люстру Chiaro Версаче	13 Август 2012 - 18:10	robotspb	Отправлено 13 Август 2012 - 18:10Новая, не подошла к общему интерьеру. Цена
5	http://360.ru/	NA	15 Август 2012 - 14:15	robotspb	Отправлено 15 Август 2012 - 14:15цена 15 000 руб . торг уместен.Сообщение отр
6	http://360.ru/	Продам Люстру Chiaro Версаче	13 Август 2012 - 18:10	robotspb	Отправлено 13 Август 2012 - 18:10Новая, не подошла к общему интерьеру. Цена
7	http://360.ru/	NA	15 Август 2012 - 14:15	robotspb	Отправлено 15 Август 2012 - 14:15цена 15 000 руб . торг уместен.Сообщение отр
8	http://360.ru/	Двушка-новостройка, самое нача	28 Апрель 2012 - 12:26	motechka	Отправлено 28 Апрель 2012 - 12:263дравствуйте, уважаемые форумчане!Как и м
9	http://360.ru/	NA	28 Апрель 2012 - 12:42	motechka	Отправлено 28 Апрель 2012 - 12:42Этаж 5-ый из 12-ти, плита - электрическая. кв
10	http://360.ru/	NA	28 Апрель 2012 - 13:33	motechka	Отправлено 28 Апрель 2012 - 13:33Ой, голова моя дырявая)))Совсем забыла ука
11	http://360.ru/	NA	28 Апрель 2012 - 14:23	ket-i-a	Отправлено 28 Апрель 2012 - 14:23motechka (28 Апрель 2012 - 12:26) писал:.....
12	http://360.ru/	NA	28 Апрель 2012 - 14:33	motechka	Отправлено 28 Апрель 2012 - 14:33ket-i-a (28 Апрель 2012 - 14:23) писал:Тогда в
13	http://360.ru/	NA	28 Апрель 2012 - 15:01	ket-i-a	Отправлено 28 Апрель 2012 - 15:01Сорри, это я зашла не по тому пути и не обра
14	http://360.ru/	NA	28 Апрель 2012 - 15:08	motechka	Отправлено 28 Апрель 2012 - 15:08ket-i-a (28 Апрель 2012 - 15:01) писал:Сорри,
15	http://360.ru/	NA	28 Апрель 2012 - 23:43	motechka	Отправлено 28 Апрель 2012 - 23:43Эх, не отвечает никто((Продолжу пока сама с
16	http://360.ru/	NA	28 Апрель 2012 - 23:54	motechka	Отправлено 28 Апрель 2012 - 23:54Или второй вариант: оставить дверь на её ро
17	http://360.ru/	NA	02 Май 2012 - 10:58	motechka	Отправлено 02 Май 2012 - 10:58Форумчане, пожалуйста, не бросайте меня совс
18	http://360.ru/	Двушка-новостройка, самое нача	02 Май 2012 - 16:31	innusya	Отправлено 02 Май 2012 - 16:31Гардеробная это, конечно. хорошо, кладовая то
19	http://360.ru/	NA	02 Май 2012 - 20:46	motechka	Отправлено 02 Май 2012 - 20:46innusya (02 Май 2012 - 16:31) писал:Гардеробна
20	http://360.ru/	NA	04 Май 2012 - 22:21	motechka	Отправлено 04 Май 2012 - 22:21Ура!!!Я его запобедила и дотумкала наконец ка
21	http://360.ru/	NA	04 Май 2012 - 22:44	Krapivka	Отправлено 04 Май 2012 - 22:44motechka (04 Май 2012 - 22:21) писал:Критика и
22	http://360.ru/	NA	04 Май 2012 - 22:54	motechka	Отправлено 04 Май 2012 - 22:54Krapivka (04 Май 2012 - 22:44) писал: Первое ,чт

Figure 37: Example of crawled comments for Russian fonts

Then applied filters properly, the number of comments has been reduced with the quantity reported in the table # 39 divided by sources:

ID Source	Source	Number of comments	
		"ковка"	"металл"
RU1	http://forum.homeideas.ru/	310	116
RU2	http://forum.ivd.ru/	2042	7139
RU3	http://360.ru/	54	46
	Total number of comments	**2406**	**7301**

Table 39: Quantity of comments per font in Russian

7.8 Russian market: data analysis

All the analyses performed are reported in disaggregated way due to the different nature of the sources and the quantity of comments. In addition the sources are few, this make the aggregate analysis just a gross approximation of the 3 analyses separated.

7.8.1 Analysis of the source “Http://Forum.Homeideas.Ru/” - RU1

The analysis is divided in two parts, the first one is referred to the keywords “Ковка” while the second one is about “Металл”.

7.8.1.1 Keyword “Ковка”

Here there is the top ten of most frequent terms that identifies the main topics:

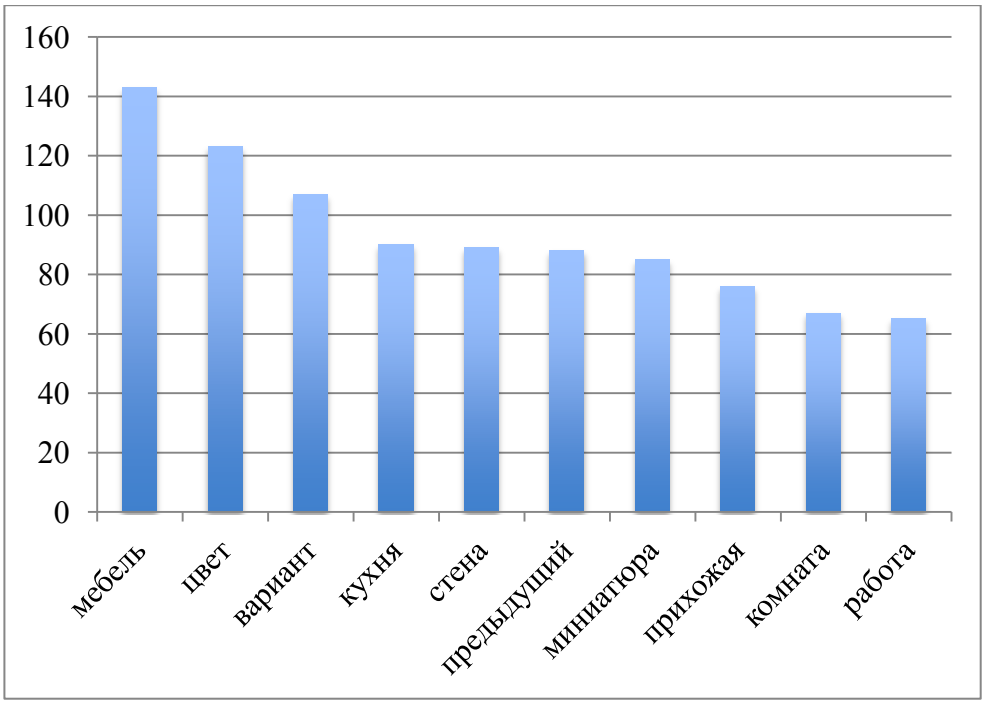

Figure 38: 10 most frequent terms for the keyword "ковка" in RU1 font

The word cloud plotted gives at once the main themes:

Figure 39: Word cloud for the keyword "ковка" in RU1 font

The study of words association has not identified the relevant word with good level of correlation with the key word "ковка". In the taxonomy it has been matched 166 terms in “Decoration”, 50 terms in "Combination", 3 terms in "People" and 47 terms in “Products”.

Cluster	Features	Number of matches
Combination	стекло	20
	дерево	30
Decoration	кованый	2
	внутренний	6
	внешний	5
	камин	10
	мебель	143
People	строитель	1
	прораб	2
	архитектор	3
Products	труба	1
	карниз	4
	решетка	4
	вешалка	5
	лестница	7
	ворота	7
	ручка	9
	балкон	10

Table 40: Taxonomy clusters for the keyword "ковка" in RU1 font

It is been created clusters trying to identify patterns and topics.

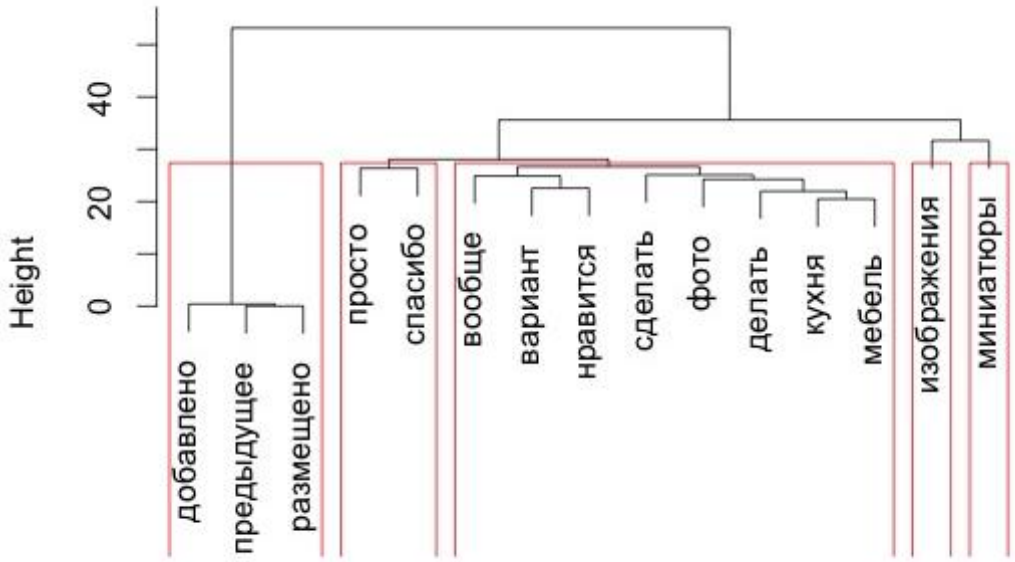

Figure 40: Clusters for the keyword "ковка" in RU1 font

Regarding sentiment analysis, the polarity means of all the documents 0.5433 in a range between [-1,1], so the source talks in a positive way about “Ковка”. The number of sentences in which it has been evaluated successfully is 1016 (758 positive, 68 neutral, 190 negative) out of 3242.

7.8.1.2 Keyword “Металл”

Here there is the top ten of most frequent terms that identifies the main topics:

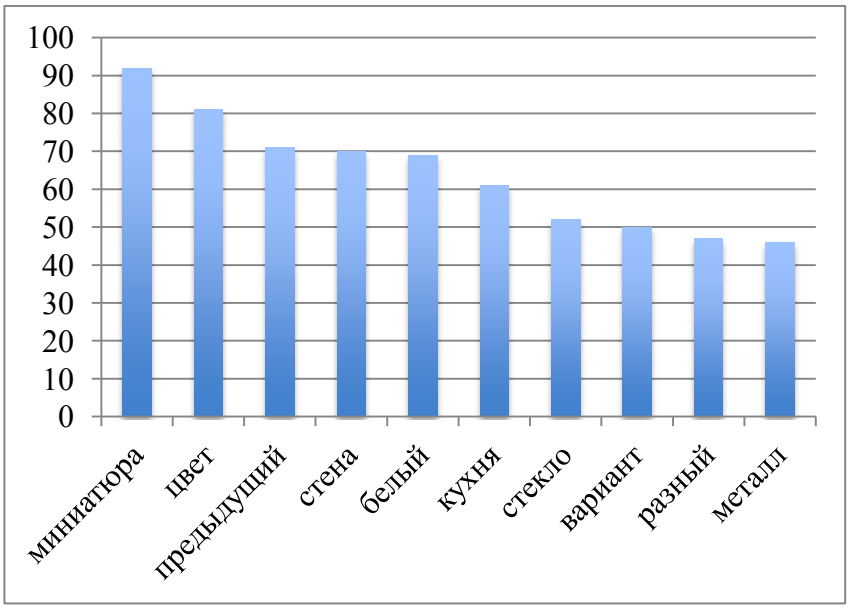

Figure 41: 10 most frequent terms for the keyword "ковка" in RU1 font

The word cloud plotted gives at once the main themes:

Figure 42: Word cloud for the keyword "металл" in RU1 font

The study of words association has not identified the relevant word with good level of correlation with the key word "металл". In the taxonomy it has been matched 23 terms in “Decoration”, 79 terms in "Combination", 6 terms in "People" and 79 terms in “Products”.

Cluster	Features	Number of matches
Combination	дерево	27
	стекло	52
Decoration	внешний	1
	внутренний	3
	камин	1
	кованый	1
	мебель	17
People	архитектор	2
	прораб	2
	строитель	2
Products	балкон	16
	вешалка	3
	карниз	3
	ручка	17
	труба	40

Table 41: Taxonomy clusters for the keyword "металл" in RU1 font

It is been created clusters trying to identify patterns and topics.

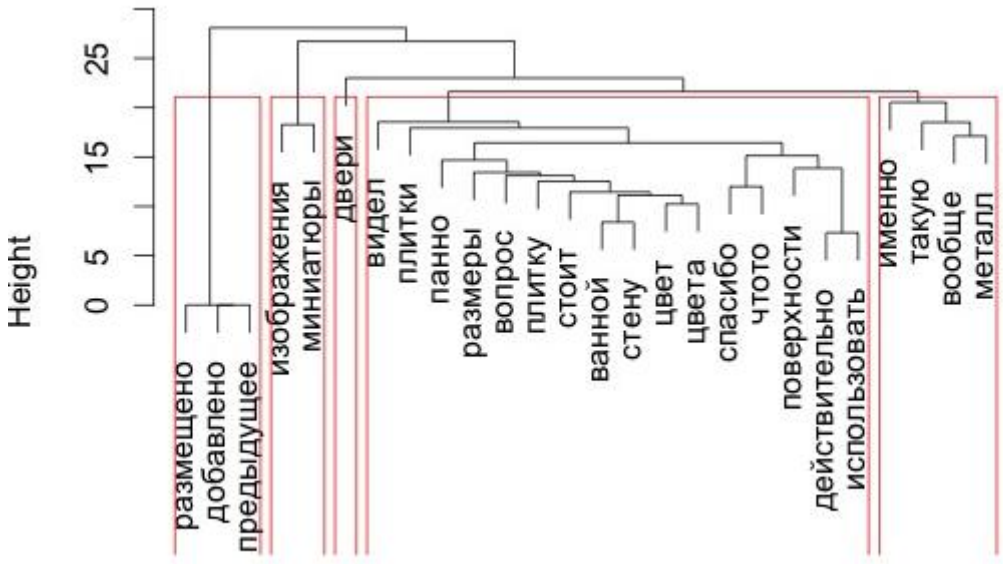

Figure 43: Clusters for the keyword "металл" in RU1 font

Regarding sentiment analysis, the polarity means of all the documents 0.5720 in a range between [-1,1], so the source talks in a positive way about “металл”. The number of sentences in which it has been evaluated successfully is 910 (688 positive, 63 neutral, 159 negative) out of 3041.

7.8.2 Analysis of the source “Http://Forum.Ivd.Ru/” - RU2

The analysis is divided in two parts, the first one is referred to the keywords “Ковка” while the second one is about “Металл”.

7.8.2.1 Keyword “Ковка”

Here there is the top ten of most frequent terms that identifies the main topics:

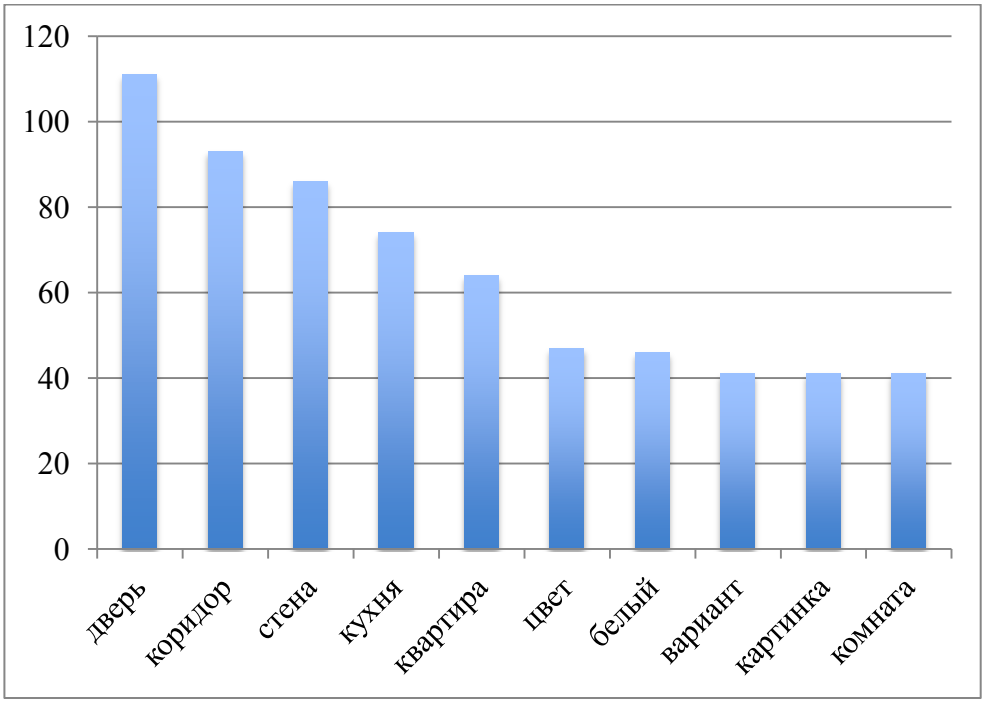

Figure 44: 10 most frequent terms for the keyword "ковка" in RU2 font

The word cloud plotted gives at once the main themes:

Figure 45: Word cloud for the keyword "ковка" in RU2 font

The study of words association has not identified the relevant word with good level of correlation with the key word "ковка". In the taxonomy it has been matched 6 terms in “Decoration”, 15 terms in "Combination", 1 terms in "People" and 41 terms in “Products”.

Cluster	Features	Number of matches
Combination	дерево	11
	стекло	4
Decoration	внешний	1
	внутренний	2
	камин	1
	кованый	2
People	строитель	1
Products	балкон	1
	вешалка	3
	карниз	1
	лестница	25
	решетка	1
	ручка	3
	труба	7

Table 42: Taxonomy clusters for the keyword "ковка" in RU2 font

It is been created clusters trying to identify patterns and topics.

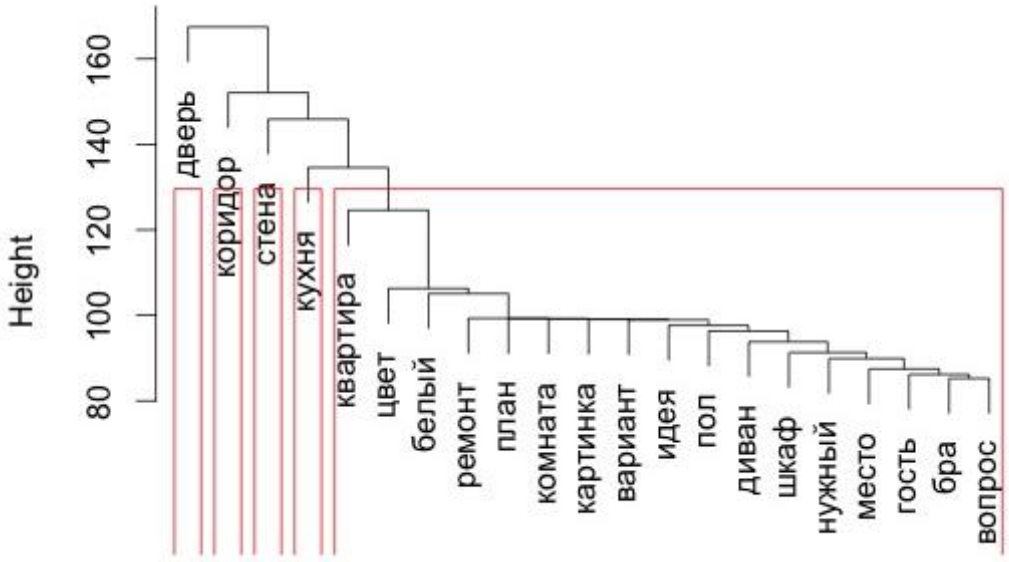

Figure 46: Clusters for the keyword "ковка" in RU2 font

Regarding sentiment analysis, the polarity means of all the documents 0.559 in a range between [-1,1], so the source talks in a positive way about “Ковка”. The number of sentences in which it has been evaluated successfully is 639 (474 positive, 54 neutral, 111 negative) out of 4640.

7.8.2.2 Keyword “Металл”

Here there is the top ten of most frequent terms that identifies the main topics:

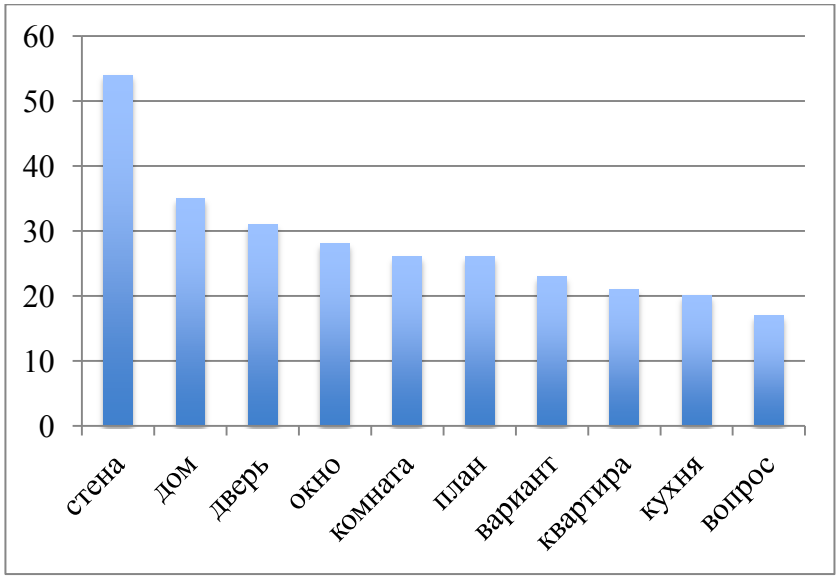

Figure 47: 10 most frequent terms for the keyword "металл" in RU2 font

The word cloud plotted gives at once the main themes:

Figure 48: Word cloud for the keyword "металл" in RU2 font

The study of words association has not identified the relevant word with good level of correlation with the key word "металл". In the taxonomy it has been matched 18 terms in “Decoration”, 7 terms in "Combination", 3 terms in "People" and 15 terms in “Products”.

Cluster	Features	Number of matches
Combination	дерево	5
	стекло	2
Decoration	внешний	3
	внутренний	5
	камин	1
	мебель	9
People	архитектор	2
	строитель	1
Products	балкон	6
	лестница	5
	перила	4

Table 43: Taxonomy clusters for the keyword "металл" in RU2 font

It is been created clusters trying to identify patterns and topics.

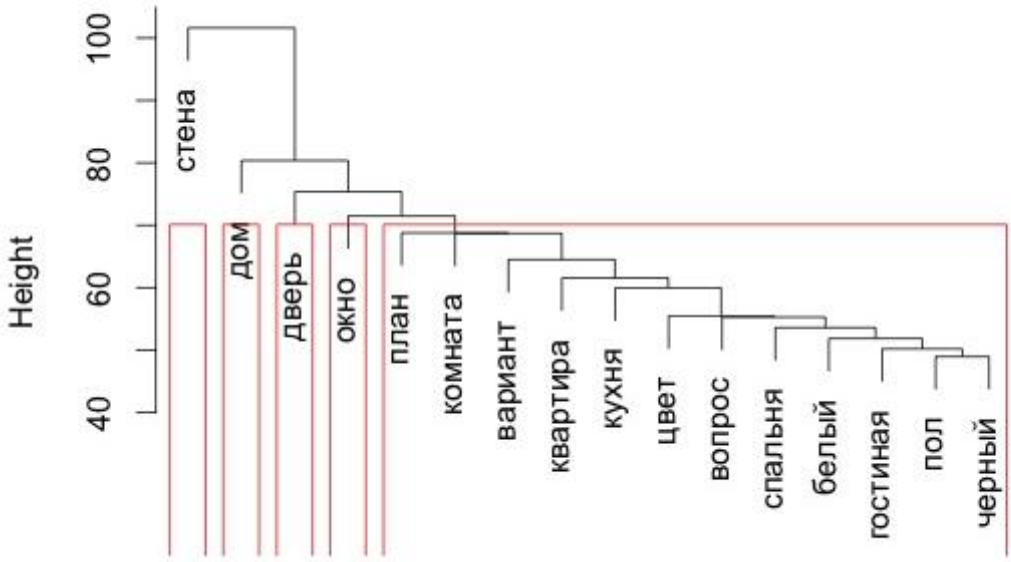

Figure 49: Clusters for the keyword "металл" in RU2 font

Regarding sentiment analysis, the polarity means of all the documents 0.5248 in a range between [-1,1], so the source talks in a positive way about “металл”. The number of sentences in which it has been evaluated successfully is 206 (155 positive, 7 neutral, 44 negative) out of 1562.

7.8.3 Analysis of the source “Http://360.Ru/” - RU3

The analysis is divided in two parts, the first one is referred to the keywords “Ковка” while the second one is about “Металл”.

7.8.3.1 Keyword “Ковка”

Here there is the top ten of most frequent terms that identifies the main topics:

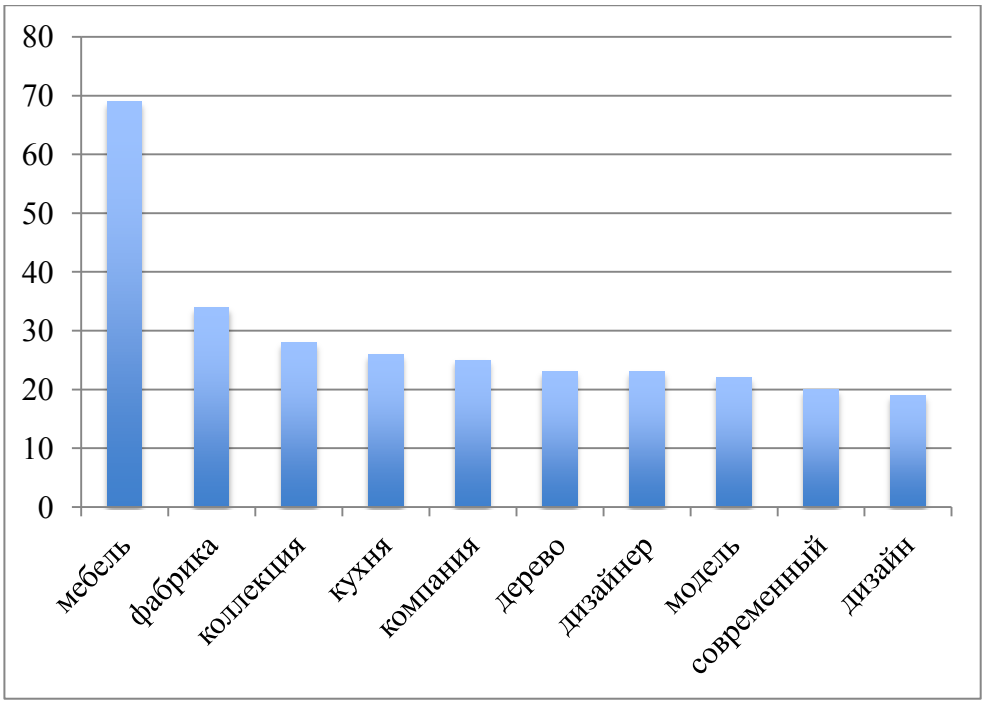

Figure 50: 10 most frequent terms for the keyword "ковка" in RU3 font

The word cloud plotted gives at once the main themes:

Figure 51: Word cloud for the keyword "ковка" in RU3 font

The study of words association has not identified the relevant word with good level of correlation with the key word "ковка". In the taxonomy it has been matched 80 terms in “Decoration”, 23 terms in "Combination", 3 terms in "People" and 19 terms in “Products”.

Cluster	Features	Number of matches
Combination	дерево	23
Decoration	внешний	3
	внутренний	5
	мебель	69
	сталь	1
	чугун	2
People	архитектор	3
Products	корпус	2
	лестница	5
	петля	4
	решетка	1
	ручка	7

Table 44: Taxonomy clusters for the keyword "ковка" in RU3 font

It is been created clusters trying to identify patterns and topics.

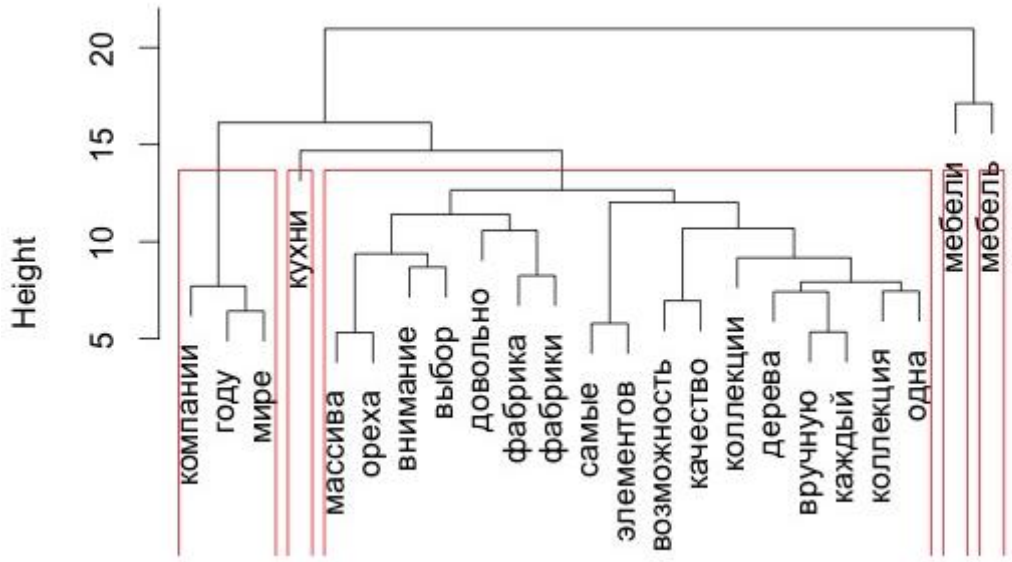

Figure 52: Clusters for the keyword "ковка" in RU3 font

Regarding sentiment analysis, the polarity means of all the documents 0.7167 in a range between [-1,1], so the source talks in a positive way about “Ковка”. The number of sentences in which it has been evaluated successfully is 240 (202 positive, 17 neutral, 21 negative) out of 437.

7.8.3.2 *Keyword "Металл"*

Here there is the top ten of most frequent terms that identifies the main topics:

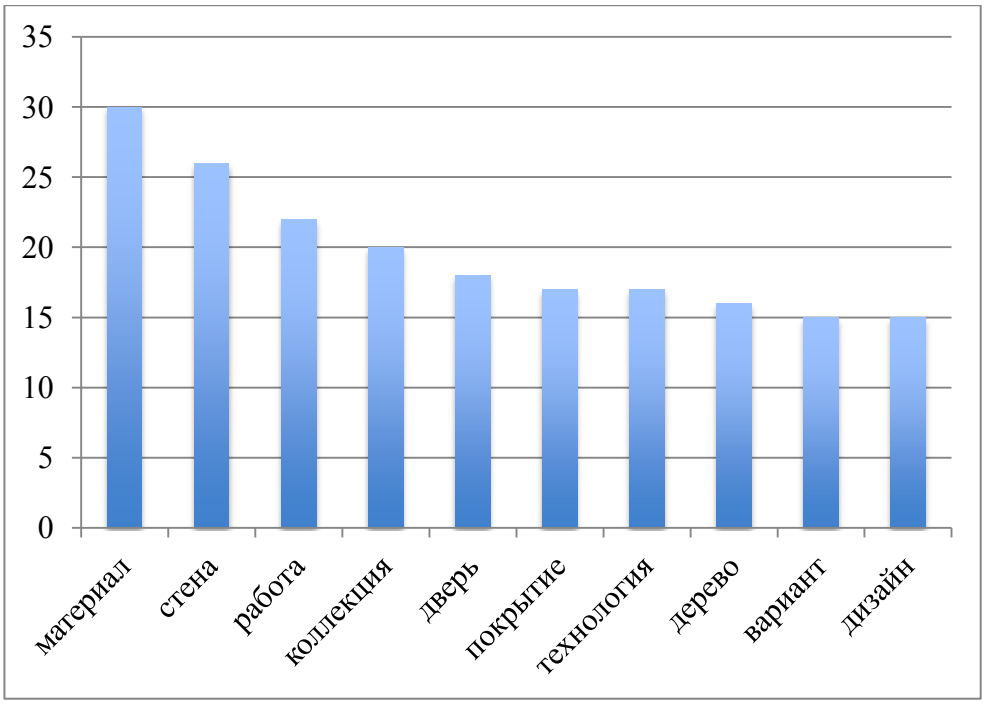

Figure 53: 10 most frequent terms for the keyword "металл" in RU3 font

The word cloud plotted gives at once the main themes:

Figure 54: Word cloud for the keyword "металл" in RU3 font

The study of words association has not identified the relevant word with good level of correlation with the key word "металл". In the taxonomy it has been matched 17 terms in “Decoration”, 21 terms in "Combination", 6 terms in "People" and 15 terms in “Products”.

Cluster	Features	Number of matches
Combination	дерево	16
	стекло	5
Decoration	внешний	5
	внутренний	6
	мебель	5
	сталь	1
People	архитектор	3
	строитель	3
Products	карниз	2
	прут	3
	труба	10

Table 45: Taxonomy clusters for the keyword "металл" in RU3 font

It is been created clusters trying to identify patterns and topics.

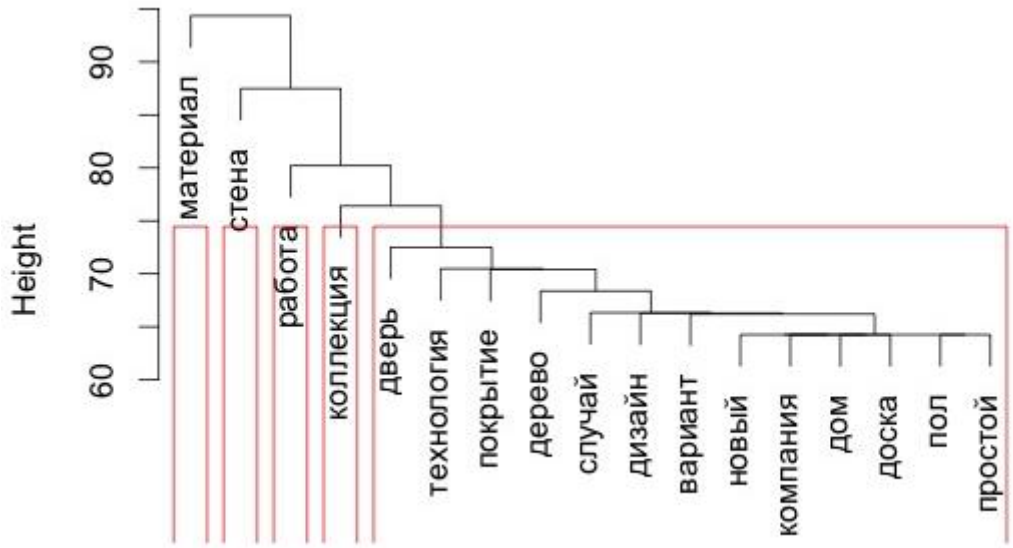

Figure 55: Clusters for the keyword "металл" in RU3 font

Regarding sentiment analysis, the polarity means of all the documents 0.6268 in a range between [-1,1], so the source talks in a positive way about “металл”. The number of sentences in which it has been evaluated successfully is 237 (184 positive, 25 neutral, 28 negative) out of 423.

7.8.4 Russian market: synthetic evidence

The table # 46 represents the synthetic analysis of the data collection and analysis for the Russian market for the keyword "Ковка":

Source	Comments	Filtered comments	Phrases	Sentiment	Absolute frequency		Taxonomy frequency		Cluster	
					Word	#	Word	#	Name	#
http://forum.homeideas.ru/	790	310	1016	0.5433	мебель	143	мебель	143	Decoration	166
					цвет	123	дерево	30		
					вариант	107	стекло	20		
http://forum.ivd.ru/	3442	2042	639	0.559	дверь	111	мебель	25	Products	41
					коридор	93	дерево	11		
					стена	86	труба	7		
http://360.ru/	145	54	240	0.7167	мебель	69	мебель	69	Decoration	80
					фабрика	34	дерево	23		
					коллекция	28	ручка	7		

Table 46: Data synthetic evidence for keyword "ковка"

General analysis of the data for the keyword "ковка" - "worked metal" gave very good results. We have collected and analyzed around 4 thousands of comments with outcome of 2 and half thousands phrases. The sentiment of the "ковка" based on these sources is positive and value is quite significant. Moreover, it is important to stress that during the clustering we can see domination of the "Decoration" cluster in the frequency analysis, that provides values from 41 to 166 terms in the analyzed phrases. The most frequent word for both absolute and taxonomy analysis is "мебель" - "furniture", which reflects the sequence of the word "ковка" with the "мебель". Thus possible usage of these sources, as channels for the new market exploration. The high frequency of the association words as "дерево", "стекло", "цвет" is correlated with the "decoration" cluster dominance. We also should add, that nonetheless the number of comments and phrases from the "360" blog is significantly less than from other source, the quality of the information and its reliability is very high and important for the whole analysis.

The table # 47 represents the synthetic analysis of the data collection and analysis for the Russian market for the keyword "Металл":

Source	Comments	Filtered comments	Phrases	Sentiment	Absolute frequency		Taxonomy frequency		Cluster	
					Word	#	Word	#	Name	#
http://forum.homeideas.ru/	700	116	910	0.572	миниатюра	92	стекло	52	Combination Decoration	79
					цвет	81	труба	40		
					предыдущий	71	дерево	27		
http://forum.ivd.ru/	14224	7139	206	0.5248	стена	54	мебель	9	Decoration	15
					дом	35	балкон	6		
					дверь	31	внутренний	5		
http://360.ru/	136	46	237	0.6268	материал	30	дерево	16	Combination	21
					стена	26	труба	10		
					работа	22	внутренний	6		

Table 47: Data synthetic evidence for keyword "металл"

For the keyword "металл" - "metal" we have analyzed around 15 thousands of the comments with outcome of only 1 and a half thousand phrases, which is indicating that the implemented approach of the data is not perfect, but reliable in terms of the data used for further analysis i.e. text mining and sentiment. Overall, the sentiment of the "металл" is positive and the value slightly lower than for the "ковка" keyword. Here we can see the domination of two clusters "Decoration" and "Combination" with subsequent high frequency of the words "дерево", "стекло", "внутренний", "внешний". As well as the prevailing frequency of "материал" - "material" is showing the sequence of the keyword "metal" and dataset collection. Moreover, the numerous evidence of the phrases analyzed represent that the "HomeIdeas" forum should be considered as important entry point for the market analysis for Euro-Fer, and reflect the objective of this part of the research.

7.9 Company uses

To resume the analysis performed it has been chosen three metrics to make a dashboard, giving immediate and summarized information:

- Relevant sources; this metric show the number of sources in the web that talk in a relevant way about the a certain topic. Usually the topic is related to one or more keywords employed for the research. This measure should be taken each three or six months.
- Sentiment polarity; this metric is the weighted mean of sentiment of the sources depending on the sentences in each source in a range of [-1,1] where a positive value indicates a positive sentiment and instead a negative one shows a negative pattern. This measure frequency depends on the kind of topic, but it is suggested to run it monthly or weekly.
- Relevant topics; this metric evaluate the most talked topics depending on the taxonomy defined across each source. The frequency of this measure should be taken weekly or monthly

The table # 48 represents a dashboard for the Italian market.

ITALY	"Ferro Battuto"	"Acciaio Inox"
Relevant Sources	4	3
Sentiment Polarity	0,169	0,405
Relevant Topics	Decoration 63% Products 37%	Products 64% Decoration 35% People 1%

Table 48: Italian market dashboard

For the keyword "Ferro Battuto" the sentiment of the comments is slightly in positive and the most talked topic is about decorations, while for the keywords "Acciaio Inox" the sentiment is definitely positive especially referring to products topic.

The table #49 represents a dashboard for the Russian market.

RUSSIA	"Kovka"	"Metall"
Relevant Sources	3	3
Sentiment Polarity	0,571	0,574
Relevant Topics	Decoration 61% Combination 19% Products 18% People 2%	Products 38% Combination 37% Decoration 20% People 5%

Table 49: Russian market dashboard

For the keyword "Kovka" the sentiment of the comments is positive and the most talked topic is about decorations, while for the keywords "Metall" the sentiment is positive in a comparable way but referring more to products and combination topic.

Moreover in this case the two project are quite comparable in term of topic, so it can be performed a benchmark between the two markets that shows that Russian people appreciate more than Italian ones basically the decoration and the products that company proposes.

Euro-Fer has decided to use the information about the Italian market to understand what contents propose in its profile on Social media, for now limited to the opening of a "Facebook" and a "Youtube" personal page. In the future with the indications given the presence can be enlarged including other Social media platforms. For the measurement part Euro-Fer, will decide which metric are more suitable for the purpose of monitoring the activity on Social media.

The Russian market KPIs are used as complementary information about the product positioning and proposition in the new market. This is crucial for the marketing plan in order to add a new level of analysis and understanding the habits of the Russian population.

For both the projects, it is fundamental the outcomes of the monitoring stage in order to have feedbacks that can introduce corrective actions timely.

8 CONCLUSION

This chapter aims to provide concluding remarks on the research project. In particular it reviews the goals achieved, the methodology used and their contributions. It also highlights the project limitations, together with the possible solutions to overcome them, and foresees future development of this work.

Social media are representing the new channel of communication, nowadays most of the companies have a presence on Twitter or a Facebook fan page. The promotion through a fan page is just the starting point for a series of opportunities related to direct interaction with people, users and customers. The natural user of this data is the marketing department, using this channel for its campaigns, brand monitoring, product advertising and customer care. In the recent years social networks has been used also for human resource purposes (i.e. recruitment) and in R&D department, as source of new ideas and knowledge sharing. Usually the use of social network is monitored in each department in order to control the activities through the performance indicators which synthesize in numbers the activities. The analysis of monitoring and controlling includes quantity KPIs, such as the number of fan or likes, and qualitative KPIs, regarding the content and its sentiment.

The features of the Social media are attractive for the management control system of the company as well: big amount of data, matched with the possibility to get extremely updated, and diversified information. The information can be synthesized in performance metrics useful to support the four main goals of a MCS: decision making, motivation and external accountability [Azzone, 2006].

In this light the goal of the research project was to analyse the role of social media within MCS and propose a framework which can support the decision-making influencing each phase of the control cycle of a company. Specifically, the goal is formed by three main issues:

- Role of Social media in the MCS
- Possible uses for management purpose
- Measurement of performances

To achieve the proposed goal a literature was conducted focusing on the Social media, MCS and measurement of performance. The Social media topic is expanded providing acknowledgement on the terms such as Web 2.0, social network and social software, describing the main classification done and the uses in the company [Kaplan, 2010; Cook, 2011; O'Reilly, 1999]. The second part of the literature introduces the MCS in the company, its evolution and the open issue, expanding the theme of indicators [Azzone, 2006]. Actually the topic of performance measurement of Social media has been tackled dividing it into two main types of metrics: KPIs related to the network activities and nature [Wasserman & Faust, 1994]; and KPIs measuring the use of Social media [Doerflinger, 2012]. Secondly, we propose a framework, taking in consideration the gap of the literature, in which indicators measuring social media performance are systematized in order to link them to uses that support decision-making process. In the end, it has been conducted two folded methodology, consisting of an action research with the empirical experimentation and case studies. Case studies analysis was carried out as the starting point to get a patchwork of the uses, performing eight interviews with different companies. While the empirical experimentation tested the current framework on a real study with the Euro-Fer company, performing two projects for Italian and Russian market respectively.

The finding of our research provided three main area of results, which are related to the starting issues and goals: the present use of social media data in companies; the challenge of the integration within MCS cycle; the misalignment between the theoretical potentiality and actual implementation.

The first area of result is related to the actual use of social media in the planning and control cycle. It is important to highlight that the case are not intended as representative of a population, but on the contrary they were selected as exemplary in the use, which was detected by a preliminary analysis. The following table # 50 summarises the cross-case analysis on four dimension: the objective, indicators (according to the framework classification), the owner of social media analysis, the role of controller.

Case #	Metric		Use	
	Objective	**Classified KPIs**	**Owner**	**Controller**
1	Sales forecasting Benchmarking	Number of comments Traffic generated Views Number of users Number of subscribers Number of articles Sentiment Popularity	Marketing	Marketing Country manager
2	Product promotion Public awareness	Traffic generated Number of subscribers Frequency of most used terms	Marketing	Marketing
3	Product co-creation Product promotion Customer service Recruitment Benchmarking	Traffic generated Number of subscribers Sentiment Frequency of most used terms Number of new ideas	Marketing Human resources Sales Management	Specialists Managers Department's head
4	Product co-creation Product promotion Listening platform Customer service Recruitment	Traffic generated Number of subscribers Sentiment Frequency of most used terms Number of new ideas	Marketing Human resources Sales Management	Specialists Managers Department's head
5	Sales forecasting Listening platform Recruitment	Traffic generated Number of subscribers Sentiment	Marketing	Marketing
6	Product co-creation Product promotion Listening platform Customer service Recruitment Social BPM	Traffic generated Number of subscribers Sentiment Frequency of most used terms Number of new ideas	Marketing Human resources Sales Management	Specialists Managers Department's head
7	Customer service Sales forecasting Product promotion Listening platform	Number of comments Traffic generated Views Number of users Number of subscribers Number of articles Frequency of most used terms Frequency of keywords terms	Consulting Marketing Human resources Sales Management	Customer Specialists Managers Department's head
8	Marketing Communication	Sentiment Frequency of most used terms Frequency of keywords terms Number of new ideas Rank of features Popularity	Marketing Management	Marketing Management

Table 50: Cross-case analysis

Overall, the case studies show that social networks are used in the marketing department, followed by sales and human resources. The companies use them to monitor marketing online activities, represented in the reports that are composed by KPIs. The metrics are highly diversified, each firm uses its own version linking KPIs to company's object, still we can see that the usage and the nature of indicators are tending to cover the same areas. This is why we preserve the possibility to implement to general methodology which could permit to improve and enlarge the usage of social media KPIs specifically on the managers level.

The second area of results is instead related to technical implementation of the process; this result stemmed both from the cases and the experimentation through action research.

Based on the information collected during the case studies, we have found that the social media information analysis', dealing with specific type of data and big amount, is requiring peculiar competences and tools. For this reason there is an evident comparative advantage of companies, which have internal IT areas, while the other firms prefer to outsource this type of analysis. Actually the case studies show how multinational company with IT background starts the implementation of internal solution focalized on needs of each department. However, we also noted the discrepancy between the implementation of specific social media solutions and its integration with the whole performance system. In consultancy and manufacturing cases the social media is used for the communication and marketing purposes, but still not included in the R&D department, though the usage is limited and implemented like a new source of information for all departments. Even in the ICT and telecommunication cases this information is hardly used by controllers.

Moreover, the important part in the social media information analysis is playing the language. By developing the empirical experimentation we have faced the challenges of analysing information in Italian and Russian, which finished in implementation of two different projects derived not only by different objectives, but by the requirements of the language analysis. This is why we have introduced the specific codes, libraries, lists reflecting languages' peculiarity for sustainable data analysis. This diversity could be preserved as a limitation for big multinational companies working on the different markets, as the analysis of many languages and different societies are requiring

dedicated approach (i.e. for Russian market we have to include Russian social networks, Facebook is officially forbidden in China) and could be very costly.
There is no a unique methodology to get sustainable data. This is why it is necessary to implement patchwork approach to ensure the reliability of data [Abdesslem, Parris, Henderson; 2011]. Likewise during the experimentation we have developed our own methodology of the analysis, which has been collaborated together with Nextbit, that could be seen as one of the possible ways to gather social media information. Anyway this methodology includes the instruments for ensuring quality of data, it is still an open issue, as there is no theoretically and practically consolidated approach.
Another important issue is privacy of the social media information and that in some cases the usage of it should be open only to researches, not for the business [Abdesslem, Parris, Henderson; 2011]. This contradicting issue has several outcomes, as the information from the Facebook is not reachable for the business purposes until the user personally did not give permission. Instead, forums, blogs and Twitter give wider possibilities for business, and lower level of personal details. However, the general problem of information disclosure is tightly correlated to the countries' internal laws and could be a big threat for the business.

The final area of results is instead related to the triangulation between the empirical evidence and the theoretical framework provided:

The rich basis of the management control system's theory having the significant gaps in terms of the usage of the social media information, being a new variable in MCS information model. This is an example where the usage of the information advance the theory background. Based on the experimentation and mainly on the case studies overview, we have discovered the whole new world of social media usage, being very diversified, but often pursuing the same goals. Moreover, the possibility of benchmark comparison between the companies requires the solid theoretical background to be implemented.

Actually the literature proposes several indicators regarding social media, but just few are implemented. Furthermore, the lack of consolidated measurement methodology is one of the main constraints that limits the adoption of this practise in a more integrated way inside the companies.

Overall research project develops a part of the new framework regarding social media in the companies, which is composed by two main parts:

- Indicators, properly classified (Nature of contents, Source of information and Network characteristics) with a methodology of measurement.
- Uses in the control cycle.

Adoption of the current framework inside a company could increase its flexibility and time to react to changes. The extension of the framework to other MCS goals could provide a more coherent MCS package for company mitigating historical issues such as being out dated, out-of-touch, too aggregated, internally focused, not fostering value creation, and not supporting effectively strategy implementation [Eccles, 1991; Johnson & Kaplan, 1987; Kaplan & Norton, 1996a; Stewart, 1991]. To achieve the full potentiality of the framework, we need to consolidate the methodology of measurement improving the measurability issue of indicators coming from social media.

Moreover, we decided to review in details the two dimensions of the framework identifying qualitative level. The "use" part can be synthetized in three main categories of adoption of social media: random use, representing company which have a presence in social media without a precise reason or strategy; object oriented, referring to firms that link the presence and use of social media to one or more objects (company, issue, product, service); refined use, referring to companies to employ social media in a more integrated a process for the process of control. Instead, the indicators are evaluated regarding the nature of contents, which is the one of classification characteristics of social media. This feature is the most critical concerning the reliability of the indicators; in fact the measurability of KPIs is strictly correlated with a correct definition of data analysis and consequently metric definition. The firms can decide to monitor the activities with undetermined metrics, punctuated metrics or the combination of punctuated and text derived metrics. The last level covers the all the data coming from social media providing complete pictures of overall performances.

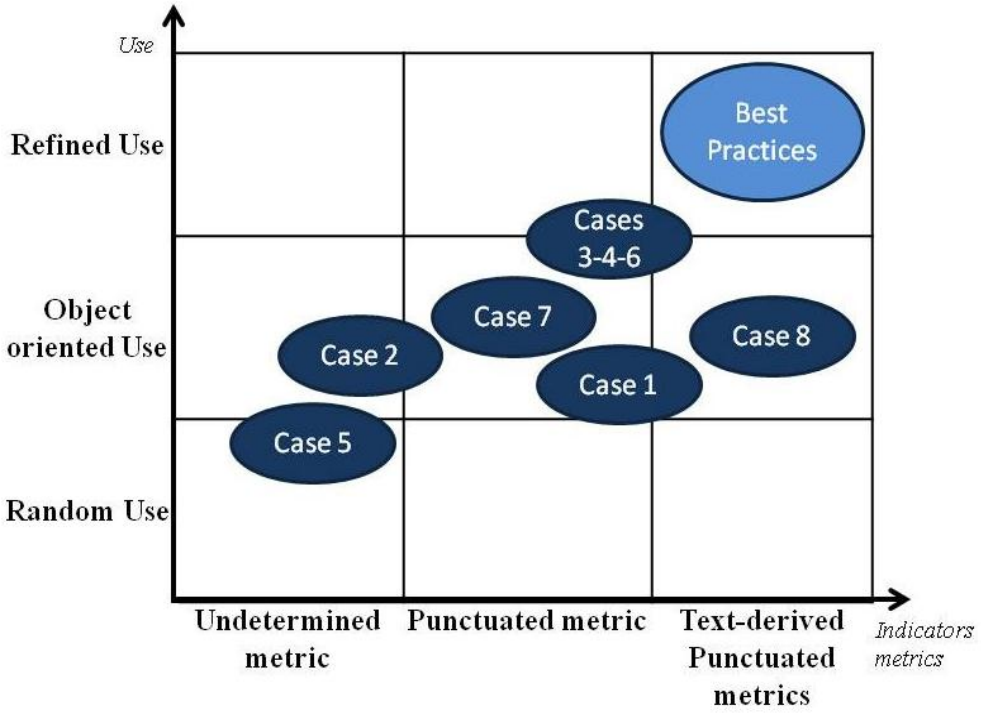

Figure 56: SM usage by cases

We identify as best practise the refined use of punctuated and text derived KPIs, which is confirmed from the direction taken by companies towards it. This process needs time to develop dedicated competencies and knowledge inside the company in order to manage properly the measurement and use stages of social media.

Furthermore, during our action research we have implemented our framework and achieved significant result for Euro-Fer, that allowed to move the position of company from random use and undetermined metric of social media information to the object oriented usage, applying punctuated and text-derived metrics, as shown on the figure # 57 below:

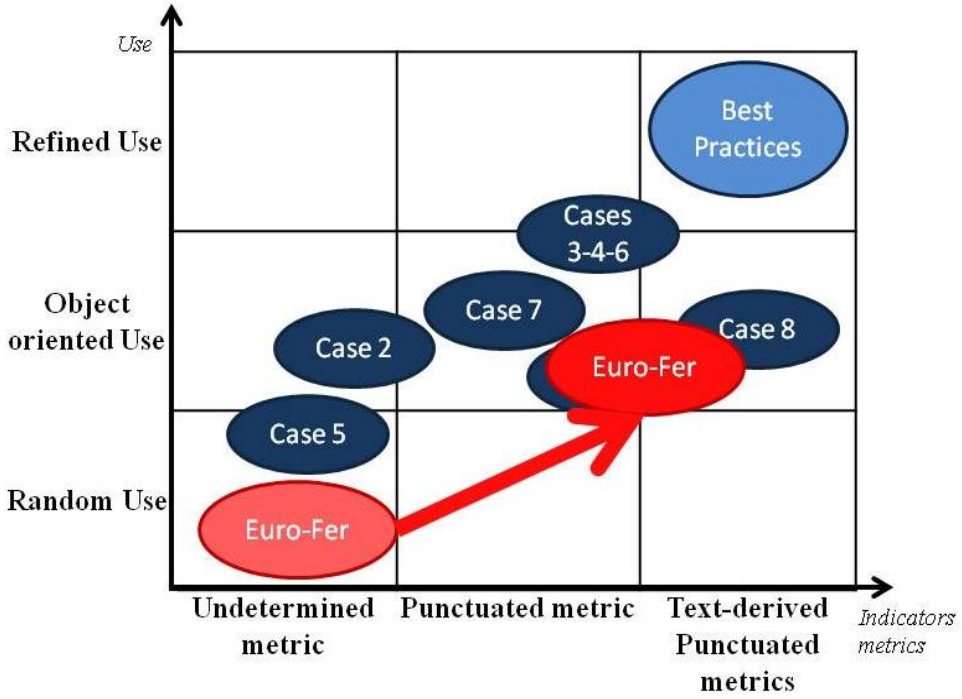

Figure 57: Euro-Fer change of SM usage

Likewise, we have reviewed the framework itself according to the action research and case studies analysis. The implementation of the "External" and "Internal" perspective to the framework is necessary for better development of motivation and learning goals of MCS, focalizing on the internal perspectives of the Social media. While the external perspective were mainly tackled during our research.

The concluding remarks of our research project are complemented by the limitations and future research perspectives. The limitations of the project had been significant challenges during our work. Instead, the future perspectives we consider to be significant starting point for further development.

The first important limitation on academic side is the topic, in fact the papers and articles regarding Social media are not numerous, this because, as it has been already said, the academics are investigating this theme currently, each month it is likely wise the release of a new contribution. In addition it is interdisciplinary topic, which means that the required knowledge is really diversified and making the theme very complex. Therefore, we preserve that a significant improvement could be done if such kind of

research are also done in the interdisciplinary team in order to have complete overview on the subject.

As regards to the proposed framework, we do not have possibility to test all 4 perspectives of the MCS, as we took a decision to focus on one - decision-making goal in order to explore it in full and have comprehensive results, by testing it in the different modes for receiving reliable results. Consequently, we foresee that other 3 goals of the MCS could be explored in further research.

The main limitation of our research project is the managing of the data, in fact the data coming from the web is really hard to handle, as well as the data sets collected from the web is not standardized. The process of gathering data of good quality is quite long taking into account the absence of the unique methodology to proceed it. Thus, it is showing poor measurability and reliability of the Social media KPIs according to the requirements defined in KPI performance [Azzone, 2006]. We understand that social media sources are diversified and following different logic inside of them, however we believe that further work on this theme could bring significant results, as well as the possible methodology of gathering information was proposed in this research and of course could be improved and completed in order to ensure the measurability and reliability of the KPI performance.

Another important limitation is devoted to the empirical experimentation in the sentiment analysis part. In our project we have tested the rule-based approach [Liu, 2004] which is only one out of three possible solutions for this phase of analysis. This is an open field of IT research that is growing up very fast. In the current years it is coming out with advanced and hybrid techniques ensuring good quality of results, but require significant involvement of IT specialist. This research could be a starting point to explore in the details this method and compare it with all other approaches in order to find the best methodology to perform such kind of tasks.

As regards to the technology, this project contains numerous codes and high variety of tested instruments and tools for information analysis coming from Social media, which have been reviewed and systematized in the methodology part of our research. However, the final choices, related usage of the particular instruments and tools, were made for “R project” packages mainly to avoid data compatibility issues. By ourselves we have

created several codes in order to perform necessary tasks for data gathering and analysis both in Italian and Russian languages. Although we see the possibility of the research in order to create a new package that could be included in the official "R-project" library performing the above mentioned procedure of data gathering and analysis in a standardized way.

REFERENCES AND BIBLIOGRAPHY

Scientific and research contribution

Abdesslem, Parris, Henderson; 2011; Reliable Online Social Network Data Collection; School of Computer Science, University Of St Andrews, Elsevier Springerlink.

Alligan, Shah; 2010; Inbound Marketing; John Wiley & Sons, Inc..

Andersen, Henriksen, Aarseth; 2006; Holistic performance management: an integrated framework; International Journal of Productivity and Performance Management, Vol. 55 Iss: 1 pp. 61 – 78.

Azzone, Bertelè; 2005; L'impresa: Sistemi Di Governo, Valutazione Controllo; Etas.

Azzone; 2006; Sistemi Di Controllo Di Gestione: Metodi, Strumenti E Applicazioni; Etas.

Backstrom, Boldi, Rosa, Ugander, Vigna; 2012; Four Deegree Of Separation; .

Bajaj, Russel; 2010; Awsm: Allocation Of Workflows Utilizing Social Network Metrics; Decision Support Systems 50 (2010) 191–202.

Bakshy, Marlow, Rosenn, Adamic; 2012; The Role Of Social Networks In Information Diffusion; Www '12 Proceedings Of The 21st International Conference On World Wide Web, Pages 519-528 .

Bakshy, Mason, Hofman , Watts ; 2011; Everyone's An Influencer: Quantifying Influence On Twitter; Wsdm'11, February 9–12, 2011.

Bakshy; 2012; The Role Of Social Networks In Information Diffusion; Www '12 Proceedings Of The 21st International Conference On World Wide Web, Pages 519-528 .

Benevenuto, Rodrigues, Cha, Almeida; 2009;Characterizing User Behavior in Online Social Networks; IMC'09, November 4–6.

Bennett, Owners, Pitt, Tucker; 2010; Workplace Impact Of Social Networking; Property Management Vol. 28 No. 3, 2010 Pp. 138-148.

Bider, Johannesson, Perjons; 2010; A Strategy For Merging Social Software With Business Process Support; Bpms2'10 Proceedings.

Bielenberg, Zacher; 2005; Groups In Social Software: Utilizing Tagging To Integrate Individual Contexts For Social Navigation; .

Borgatti, Foster; 2003; The Network Paradigm In Organizational Research: A Review And Typology; Journal Of Management 29(6) 991–1013.

Borgatti, Halgin; 2011; On Network Theory; Organization Science, Articles In Advance, Pp. 1–14,.

Boyd, Ellison; 2008; Social Network Sites: Definition, History, And Scholarship; Journal Of Computer-Mediated Communication (2008) 210–230.

Bradbury; 2011; Data Mining With Linkedin; Computer Fraud & Security.

Brander; 2011; Processing Social media Posts In Contact Centers; .

Brogan; 2010; Social media 101; John Wiley & Sons, Inc..

Brzozowski; 2009; Watercooler: Exploring An Organization Through Enterprise Social media; Www Conference 2009.

Bughin; 2008; The Rise Of Enterprise 2.0; Journal Of Direct, Data And Digital Marketing Practice (2008) 9, 251–259..

Burrus; 2010; Social Networks In The Workplace: The Risk And Opportunity Of Business 2.0; Strategy & Leadership Vol. 38 No. 4 2010, Pp. 50-53,.

Burt; 2000; The Network Structure Of Social Capital; Research In Organizatinnal Behaviour, Volume 22, Pages 345-423..

Carley; 2003; Dynamic Network Analysis; National Science Foundation.

Chenhall, Kallunki, Sivola; 2008; Exploring The Relationships Between Strategy, Innovation And Management Control Systems: The Roles Of Social Networking, Organic Innovative Culture And Formal Controls; Journal Of Management Accounting Research.

Chenhall; 2003; Management Control Systems Design Within Its Organizational Context: Findings From Contingency-Based Research And Directions For The Future; Accounting, Organizations And Society (2003) 127–168.

Chiu, Ip, Silverman; 2012; Understanding Social media In China; Mckinsey Quarterly.

Clauset, Moore, Newman; 2008; Hierarchical Structure And The Prediction Of Missing Links In Networks; IEEE.

Clauset, Newman, Moore; 2004;Finding community structure in very large networks; IEEE.

Cook; 2010; How Social Software Will Change The Future Of Work; Emerald Group Publishing Limited.

Daly, Haahr; 2009; Social Network Analysis For Information Flow In Disconnected Delay-Tolerant Manets; Ieee Transactions On Mobile Computing, Vol. 8, No. 5, May 2009.

DöRflinger; 2011; How To Legitimate The Effort In Online Communication; Journalism And Corporate Communication.

Dron; 2006; Social Software And The Emergence Of Control; The 6th Ieee International Conference On Advanced Learning Technologies (Icalt 2006), 5-7 July 2006, Kerkrade, Netherlands..

Dutta; 2010; What's Your Personal Social media Strategy?; Harvard Business Review.

Eisenhardt, Graebner; 2007; Theory building from cases: opportunities and challenges;

Academy of Management Journal, 2007, vol. 50, n. 1, pp. 25- 32.

El-Sayed, Westrup; 2011; Adopting Enterprise Web 2.0 Collaborative Technologies In Business: The Implications For Management Accountants; Research Executive Summary Series, Chartered Institute Of Management Accountants, Manchester Business School.

Feinerer, Hornik, Meyer; 2008; Text Mining Infrastructure In R; Journal Of Statistical Software, March 2008, Volume 25, Issue 5.

Ferreira, Otley; 2009; The Design And Use Of Performance Management Systems: An Extended Framework For Analysis; Management Accounting Research 20 (2009) pp. 263–282.

Flamholtz; 1996; Effective Organizational Control. A Framework, Applications, And Implications; European Management Journal Vo114 No 6.

Foster, Mcafee; 2006; Enterprise 2.0: The Dawn Of Emergent Collaboration; Mit Sloan Management Review.

French; 2009; Action Research For Practising Managers; Journal Of Management Development Vol. 28 No. 3.

Gestwicki, Jayaraman; 2005; Methodology And Architecture Of Jive; Association For Computing Machinery, Inc..

Golbeck, Hendler; 2004; Accuracy Of Metrics For Inferring Trust And Reputation In Semantic Web-Based Social Networks; Ekaw 2004, Lnai 3257, Pp. 116–131, 2004..

Granovetter; 1973; The Strength Of Weak Ties; American Journal Of Sociology 78(6):1360-1380..

Gross, Acquisti; 2005; Information Revelation And Privacy In Online Social Networks; Wpes'05, November 7, 2005.

Gyarmati, Trinh; 2010; Measuring User Behavior in Online Social Networks; IEEE.

Haefliger, Monteiro, Foray, Von Krogh; 2011; Social Software And Strategy; Long Range Planning, Vol 44.

Hanna, Rohm, Crittenden; 2011; We're All Connected: The Power Of The Social media Ecosystem; Elley School Of Business, Indiana University, Elsevier.

Hung, Wu, Wen, Wu; 2008; Competitive Advantages Of Managing An Effective Social Network Structure To Stimulate Innovation From A Knowledge Management Perspective; International Journal Of Technology Management.

Jackson; 2011; The Value Of Weak Connections; Journal Of Business Strategy Vol. 32 No. 5 2011, Pp. 51-53.

Joshi, Bartolini, Graupner; 2012; T@Gz : Intuitive And Effortless Categorization And Sharing Of Email Conversations; Www Conference 2012.

Kaplan, Haenlein; 2010; Users Of The World, Unite! The Challenges And Opportunities Of Social media; Business Horizons (2010) 53, 59—68.

Keenan, Shiri; 2009; Sociability And Social Interaction
On Social Networking Website; Library Review Vol. 58 No. 6, 2009 Pp. 438-45, Emerald Group Publishing Limited.

Kloot; 1997; Organizational learning and management control systems: responding to environmental change; Management Accounting Research Volume 8, Issue 1, March 1997, Pages 47‑73.

Kumar, Lease; 2011; Learning To Rank From A Noisy Crowd; SIGIR'11.

Lee, Kwak; 2012; An Open Government Maturity Model For Social media-Based Public Engagement; Government Information Quarterly, Elsevier Inc.

Lee; 2012; Unsupervised And Supervised Learning To Evaluate Event Relatedness Based On Content Mining From Social-Media Streams, Expert Systems With Applications, Elsevier Inc.

Malmia, Brown; 2008; Management Control Systems As A Package-Opportunities, Challenges And Research Directions; Management Accounting Research L9 (2008) 287-300.

Mislove, Viswanath, Gummadi, Druschel; 2010;You Are Who You Know: Inferring User Profiles in Online Social Networks; WSDM'10, February 4–6.

Murdough; 2009; Social media Measurement: It's Not Impossible; Journal Of Interactive Advertising, Vol 10 No 1 (Fall 2009), Pp. 94-99..

Narayanan, Shmatikov; 2010;De-Anonymizing Social Networks; University Of Austin, Texas.

Nasukawa, Yi; 2003; Sentiment Analysis: Capturing Favorability Using Natural Language Processing; K-Cap'03, October 23-25.

Nazir, Raza, Chuah; 2008; Unveiling Facebook: A Measurement Study Of Social Network Based Applications; Imc'08, October 20–22, 2008,.

O'reilly; 2007; What Is Web 2.0: Design Patterns And Business Models For The Next Generation Of Software; Communications & Strategies, No. 65, 1 Quarter 2007, p. 17.

Otte, Rousseau; 2002; Social Network Analysis: A Powerful Strategy, Also For The Information Sciences; Journal Of Information Science 2002 28: 441.

Peng, Park; 2011; Generate Adjective Sentiment Dictionary For Social media Sentiment Analysis Using Constrained Nonnegative Matrix Factorization; Association For The Advancement Of Artificial Intelligence.

Prabowo, Thelwall; 2009; Sentiment Analysis: A Combined Approach; Journal Of Informetrics Volume 3, Issue 2, April 2009, Pages 143–157.

Prell, Hubacek, Quinn, Reed; 2008;'Who's In The Network?' When Stakeholders Influence Data Analysis; Ó Springer Science+Business Media, LLC 2008.

Prell, Hubacek, Reed; 2009; Stakeholder Analysis And Social Network Analysis In

Natural Resource Management; Society & Natural Resources.

Qu, Dumay; 2011; The qualitative research interview; Qualitative Research in Accounting & Management, Vol. 8 Iss: 3 pp. 238 – 264.

Reiss, Reiss; 2005; Demonstration Of Jive And Jove:Java As It Happens; Icse'05,May 15-21, 2005,.

Richardson; 2012; Social Bpm; Future Strategies Inc..

Rooksby; 2009; Social Networking And The Workplace; Strategic Direction Vol. 25 No. 8 2009, Pp. 20-23, .

Roxane Divol, David Edelman, And Hugo Sarrazin; 2012; Demystifying Social media; Mckinsey Quarterly, Marketing & Sales Practice.

Sandt; 2012; Management Control Systems As A Package: The Relationship Of Financial Measurement Systems And Six Sigma; Iubh Working Paper No.1.

Serrat; 2009; Social Network Analysis; International Publications; Cornell University ILR School.

Siggelkow; 2007; Persuasion With Case Studies; Academy Of Management Journal 2007, Vol. 50, No. 1, 20-24.

Sobkowicz, Kaschesky, Bouchard; 2012; Opinion Mining In Social media: Modeling, Simulating, And Forecasting Political Opinions In The Web; Government Information Quarterly, Elsevier Inc.

Spina; 2008; La Gestione Dell'impresa: Organizzazione, Processi Decisionali, Marketing, Acquisti E Supply Chain.

Sundaresan, Fischoo, Dusho, Rubenstein; 2006; Network Metrics Reveal Differences In Social Organization Between Two Fission–Fusion Species, Grevy's Zebra And Onager; Springer-Verlag 2006.

Svore, Volkovs, Burges; 2011; Learning to Rank with Multiple Objective Functions; WWW

Sykes, Venkatesh; 2009; Model Of Acceptance With Peer Support: A Social Network Perspective To Understand Employees' System Use; Mis Quarterly Vol. 33 No. 2, Pp. 371-393/June 2009.

Tichy, Tushman, Fombrun; 1979; Social Network Analysis For Organizations; The Academy Of Management Review, Vol. 4, No. 4 (Oct., 1979), Pp. 507-519.

Ugander, Kareer, Backstrom, Marlow; 2011; The Anatomy Of The Facebook Social Graph; .

Van Der Aalst, Song; 2004; Mining Social Networks: Uncovering Interaction Patterns In Business Processes; Springer-Verlag Berlin Heidelberg 2004.

Wang; 2009; Integrating Organizational, Social, And Individual Perspectives In Web 2.0-Based Workplace E-Learning; Springer Science + Business Media, Llc 2009.

Wasserman, Faust; 1994; Social Network Analysis: Methods And Application; Cambridge University Press.

Wen, Lin; 2010; On The Quality Of Inferring Interests From Social Neighbors, KDD'10.

Willett, Heer, Agrawala; 2012; Strategies For Crowdsourcing Social Data Analysis; CHI.

Wright; 1996; Exploring the in-depth interview as a qualitative research technique with American and Japanese firms; Marketing Intelligence & Planning 14/6 [1996] 59–64.

Xie, Li, Xia; 2010;Towards Efficient Content Management In Online Social Communities – A Study Of User Interest And Context; IEEE.

Yin, R.K.; 2002, Case Study Research: Design and Methods, Thousand Oaks, CA: Sage.

Zach; 2000; Researching Organizational Systems Using Social Network Analysis; IEEE.

Practitioners contributions

Amati; 2011; Sentiment Analysis e il WEB 2.0; Fondazione Ugo Bordoni.

Annau; 2010; Online Sentiment Analysis using R; tm.plugin.sentiment.

Aswani; 2012; Updated Sentiment Analysis and a Word Cloud for Netflix – The R Way!; http://www.r-bloggers.com/.

Belleghem, Peteghem; 2012;The social dynamics model; Insites consulting.

Bodnar, Cohen; 2012; Social media B2B: Exploring the impact of Social media on B2B companies; http://socialmediab2b.com/

Breen; 2011; R By Example: Mining Twitter For Consumer Attitudes Towards Airlines; Boston Predictive Analytics Meetup.

Business Link; 2012; Keep Your Business Fighting Fit; European Union Investing In Your Future.

Court, Divol, Edelman; 2012; Making sense of social media; Mckinsey Quarterly.

Curini, Iacus, Porro; 2012; Voices From The Web; http://voicesfromtheblogs.com.

Davis; 2012; 5 Essential & Easy Social media Metrics You Should Be Measuring Right Now; http://blog.kissmetrics.com/.

Dellamaggiore, Smith; 2012; Linkedin: A Professional Social Network Built With Java Technologies And Agile Practices; Java Conference.

Digital method initiative; https://www.digitalmethods.net/Digitalmethods/WebHome.

Etlinger; 2011; A Framework For Social Analytics; Altimeter.

Francis; 2011; Measuring The Future: New Metrics For New Media; Web 2.0 SF

Kaushik; 2011; Best Social media Metrics: Conversation, Amplification, Applause, Economic Value; http://www.kaushik.net/avinash/.

Kistner, O'Neill; 2012; How The Big Brand Do Facebook Marketing; Webtrends.

Laird, Schipul; 2012; Social media For Business; The Web Marketing Company.

Latham; 2008; Measuring Value Form Social media; Spur Digital.

Little; 2010; The Value Of Social media: Getting Down To Business; NMincite.

Mofitt; 2008; Monitoring And Measuring Conversation; Agent Wildfire Inc.

Musante; 2012; Under Development: A Social media Measurement Model; Social media Today.

Noci; 2012; Social Network e B2B: quali opportunità? ; MIP.

Owyang, Jones; 2012; Social media Management: An Industry Perspective; Altimeter.

Owyang; 2012; a strategy for managing social media proliferation; Altimeter.

Paruchi; 2012; Tracking US Sentiments Over Time In Wikileaks; http://www.r-bloggers.com/.

Pentin; 2012; A New Framework For Measuring Social media Activities; IAB Social media Council.

Rozwell; 2012; Develop A Social media Strategy That Will Improve Your Business; Gartner.

Rubel; 2011; Social media Probook; Eloqua.

Tarcizio; 2012; Social media Metrics; Paperclio.

Venables, Smith; 1990-2; An Introduction to R; http://cran.r-project.org/.

Yongfook; 2012; Social media ROI: Measuring The Unmeasurable; Egg. Co.

Zhao; 2012; An example on sentiment analysis with R; http://www.r-bloggers.com/.

APPENDIX

Case studies

Questioner for interviews:

POLITECNICO
DI MILANO

Company:
Department:
Interviewee:
Mode:
Date:
Duration:

Questions for the Interview:

1. Do you use Social Media for/ in your company?
2. Which Social Media do you use?
3. How often do you do updates in your Social Media profile?
4. Do you collect and analyze information from Social Media?
5. What are the problems associated with the process of data collection and analysis from Social Media?
6. Which types of tools do you use to collect and analyze the information?
7. Do you use any specific metrics to monitor activities on Social Media?
8. Do you monitor only your Social Media entry points or of your competitors?
9. Which are the most important Social Media for your business?
10. How the Social Media helps in marketing?
11. Have you ever thought about using Social Media information for the management of the company?
12. Have you ever used the Social Media for the benchmarking with the competitors?
13. Have you ever used the Social Media for human resources purposes?
14. What are the main benefits associated with the usage of information from the Social Media?
15. What are the main disadvantages?
16. Do you generate any reports for the information from Social Media? How do you represent information from Social Media?
17. Who is using these reports?
18. How frequently you are doing these reports?

POLITECNICO
DI MILANO

Company:
Department:
Interviewee:
Mode:
Date:
Duration:

Questions for the Interview:

1. How often the consulting for the Social Media is requested right now on the market?
2. For which kind of firms Social Media could be used?
3. Pleas describe what client expects from usage of the Social Media and from the information from Social Media?
4. Do you foresee that clients would switch from classical method of collection information to Social Media or only implementing it as additional source of information?
5. In which areas of business clients are willing to implement Social Media? (LinkedIn for Human Resources, Facebook for Marketing, Twitter for Customer satisfaction and etc.)?
6. Which of the Social Media resources are more popular?
7. What kind of reports do you provide to clients and how frequently?
8. What are the main benefits and drawbacks you see in the Social Media and usage of information from Social Media for company's operation?
9. Which types of tools do you use to collect and analyze the information?
10. Do you use any specific metrics to monitor activities on Social Media?

Proposal for consultancy company.

TESI DI LAUREA

PROPOSTA DI PROGETTO

INTRODUZIONE

Una delle più grandi innovazioni venuta alla ribalta negli ultimi anni è il nuovo universo dei *social network*. Nati come semplici mezzi di aggregazione sociale, si sono trasformati in uno strumento molto efficace per la gestione aziendale (marketing, customer experience, innovazione, importanza del marchio, efficienza operativa, generazione di bilancio), ed hanno così aperto nuove opportunità per le aziende che possono sfruttarne le potenzialità. Nonostante un crescente interesse da parte delle imprese, le applicazioni specifiche, ma soprattutto l'utilizzo dei dati generati dai *social network* in ambito aziendale è ancora limitato. Questo lavoro di tesi mira a contribuire a quest'area focalizzandosi su tema specifico: l'utilizzo dei social network per i processi di ricerca e selezione del personale.

Uno delle problematiche critiche per una compagnia è garantire un alto e variegato livello di competenze, soprattutto quando esiste un elevato ricambio di personale. Attualmente la scelta di nuovi consulenti avviene tramite un lungo iter di selezione volto a filtrare e soprattutto a scegliere pretendenti che rispettino le qualità cercate. Uno strumento che si è imposto sul mercato in questi ultimi anni e che permette di integrare e ottenere un quadro completo dei candidati è il social network Linkedin. Inoltre risulta essere uno strumento utile per diffondere una veritiera immagine dell'azienda per i partner commerciali e per chi cerca opportunità lavorative.

OBIETTIVI

A fronte di questo contesto, la tesi di laurea mira a sviluppare un modello che assista e supporti il processo di selezione del personale, ottenendo preziose informazioni preliminari che aiuteranno a filtrare le possibili figure assumibili, soprattutto quando si tratta di personale specializzato. Nello specifico l'applicazione selezionata per sviluppare il modello è **Linkedin**, che con le sue potenzialità, assisterà il recruiter nel processo di selezione e permetterà di controllare il personale sotto diverse prospettive, motivarlo e anche introdurre nuovi processi di apprendimento.

FASI E SCADENZE

Qui verranno sinteticamente spiegati i passi per lo sviluppo del progetto con le relative scadenze:

- Iniziale raccolta dati e definizione di un percorso con l'azienda (30 aprile-15 maggio)
- Preparazione dell'applicazione (16- maggio- 29 giugno)
- Raccolta dati (30 giugno - 31 luglio)
- Analisi dei dati (16 agosto - 29 settembre)
- Validazione del modello (30 settembre - 30 ottobre)
- Spedizione del rapporto finale (15 novembre)

OUTPUT FINALE

L'output finale del progetto sarà:

1. Creazione e mantenimento di un profilo **LinkedIn** che sia in grado di generare traffico rilevante.
2. Creazione di un modello che permetta di integrare le varie sorgenti di informazione.

3. Set di indicatori creati "ad hoc" direttamente dalle informazioni ottenute.

RISORSE

Il progetto è previsto all'interno della tesi di Jacopo Radaelli e Yulia Sidorova (in allegato i curricula), che stanno terminando il percorso di laurea magistrale in Ingegneria Gestionale. Il progetto sarà supervisionato della Prof.ssa Michela Arnaboldi.

Nessun contributo finanziario è richiesto alla vostra azienda.

Proposal for Career service:

TESI DI LAUREA

PROPOSTA DI PROGETTO

INTRODUZIONE

Una delle più grandi innovazioni venuta alla ribalta negli ultimi anni è il nuovo universo dei *social network*. Nati come semplici mezzi di aggregazione sociale, si sono trasformati in uno strumento molto efficace per la gestione aziendale (marketing, customer experience, innovazione, importanza del marchio, efficienza operativa, generazione di bilancio), ed hanno così aperto nuove opportunità per le aziende che possono sfruttarne le potenzialità. Nonostante un crescente interesse da parte delle imprese, le applicazioni specifiche, ma soprattutto l'utilizzo dei dati generati dai *social network* in ambito aziendale è ancora limitato. Questo lavoro di tesi mira a contribuire a quest'area focalizzandosi su tema specifico: l'utilizzo dei social network per i processi di ricerca e selezione del personale.

Una delle peculiarità del servizio career service è proporre un'ottima descrizione per le aziende degli studenti laureandi e appena laureati provenienti dal Politecnico avendoli seguiti per tutta la loro carriera accademica. Questo profilo molto dettagliato e puntuale rimane però come un'istantanea al conseguimento della laurea degli studenti poiché non viene aggiornato con le loro successive esperienze lavorative e quindi non può gestire richieste di aziende che richiedono personale più esperto. Questo programma aiuterebbe nell'integrazione delle informazioni e delle esperienze lavorative degli ex studenti, connettendo ulteriormente il Politecnico all'ambito lavorativo.

OBIETTIVI

A fronte di questo contesto, la tesi di laurea mira a è un modello che possa affiancare il career service per il completamento delle informazioni in modo da avere sempre profili completi indipendentemente dal tipo di personale cercato (laureando, neo-laureato o laureato). Questo modello verrà supportato da **LinkedIn**, consentendo di tener traccia dei laureati del Politecnico ottenendo tutte le informazioni necessarie per un profiling accurato anche dopo anni dal conseguimento della laurea. L'associazione a questo gruppo e l'inserimento del proprio curriculum nel programma connesso al career service, garantirebbero agli occhi delle aziende la veridicità dell'informazione e la preparazione di alto livello tipica del Politecnico.

FASI E SCADENZE

Qui verranno sinteticamente spiegati i passi per lo sviluppo del progetto con le relative scadenze:

- Iniziale raccolta dati e definizione di un percorso con l'azienda (30 aprile-15 maggio)
- Preparazione dell'applicazione (16 maggio - 29 giugno)
- Raccolta dati (30 giugno - 31 luglio)
- Analisi dei dati (16 agosto - 29 settembre)
- Validazione del modello (30 settembre - 30 ottobre)
- Spedizione del rapporto finale (15 novembre)

OUTPUT FINALE

L'output finale del progetto sarà:

1. Creazione e mantenimento di un profilo **LinkedIn** che sia in grado di generare traffico rilevante.

2. Creazione di un modello che permetta di integrare le varie sorgenti di informazione.
3. Set di indicatori creati "ad hoc" direttamente dalle informazioni ottenute.

RISORSE

Il progetto è previsto all'interno della tesi di Jacopo Radaelli e Yulia Sidorova (in allegato i curricula), che stanno terminando il percorso di laurea magistrale in Ingegneria Gestionale. Il progetto sarà supervisionato della Prof.ssa Michela Arnaboldi.

Nessun contributo finanziario è richiesto alla vostra azienda.

Proposal for HR department of a company:

TESI DI LAUREA

PROPOSTA DI PROGETTO

INTRODUZIONE

Una delle più grandi innovazioni venuta alla ribalta negli ultimi anni è il nuovo universo dei *social network*. Nati come semplici mezzi di aggregazione sociale, si sono trasformati in uno strumento molto efficace per la gestione aziendale (marketing, customer experience, innovazione, importanza del marchio, efficienza operativa, generazione di bilancio), ed hanno così aperto nuove opportunità per le aziende che possono sfruttarne le potenzialità. Nonostante un crescente interesse da parte delle imprese, le applicazioni specifiche, ma soprattutto l'utilizzo dei dati generati dai *social network* in ambito aziendale è ancora limitato. Questo lavoro di tesi mira a contribuire a quest'area focalizzandosi su tema specifico: l'utilizzo dei social network per i processi di ricerca e selezione del personale.

Uno dei problemi che infatti emerge prepotentemente nel fornire alle aziende personale aderente alle loro necessità e l'impossibilità ad identificare un insieme ampio di persone qualificate e in cerca di un impiego. Con la diffusione dei *social network* e l'avvento delle notizie in tempo reale è possibile migliorare i propri strumenti di ricerca e selezione riducendo notevolmente i costi e ampliando le possibilità di visibilità, sia per le offerte lavorative, sia per l'offerta di competenze.

OBIETTIVI

A fronte di questo contesto, la tesi di laurea mira a sviluppare un modello che assista e supporti il processo di selezione del personale, ottenendo preziose informazioni preliminari che aiuteranno a filtrare le possibili figure assumibili, soprattutto quando si tratta di personale specializzato. Nello specifico l'applicazione selezionata per sviluppare il modello è **Linkedin**, che con le sue potenzialità, assisterà il recruiter nel processo di selezione. In questo modo il recruiter potrà fornire all'azienda delle proposte di assunzione di alta qualità riuscendo ad avvicinarsi sempre meglio alle sue richieste.

FASI E SCADENZE

Qui verranno sinteticamente spiegati i passi per lo sviluppo del progetto con le relative scadenze:

- Iniziale raccolta dati e definizione di un percorso con l'azienda (30 aprile-15 maggio)
- Preparazione dell'applicazione (16- maggio- 29 giugno)
- Raccolta dati (30 giugno - 31 luglio)
- Analisi dei dati (16 agosto - 29 settembre)
- Validazione del modello (30 settembre - 30 ottobre)
- Spedizione del rapporto finale (15 novembre)

OUTPUT FINALE

L'output finale del progetto sarà:

1. Creazione e mantenimento di un profilo **LinkedIn** che sia in grado di generare traffico rilevante.
2. Creazione di un modello che permetta di integrare le varie sorgenti di informazione.
3. Set di indicatori creati "ad hoc" direttamente dalle informazioni ottenute.

RISORSE

Il progetto è previsto all'interno della tesi di Jacopo Radaelli e Yulia Sidorova (in allegato i curricula), che stanno terminando il percorso di laurea magistrale in Ingegneria Gestionale. Il progetto sarà supervisionato della Prof.ssa Michela Arnaboldi.

Nessun contributo finanziario è richiesto alla vostra azienda.

Technology

Please note that these codes were developed for the specific web-sites, including the codes for crawling, text mining and sentiment, as discussed in the instruments' part of our thesis. Hereby represented the extracts of the codes in sequence of the work proceed together with their notes.

RUSSIAN PROJECT EXAMPLE OF CODE:

```
#punti da cambiare per altri forum:

#popola:

#urlstart con le url del search

#nella grep tengo la root del forum che mi interessa per filtrare dei link

#nel getNodeSet devo mettere gli identificativi del tag con dentro i post

library(XML)

#funzione che serve per estrarre i link da una pagina web

getLinks = function() {

 links = character()

 list(a = function(node, ...) {

  links <<- c(links, xmlGetAttr(node, "href"))

  node
```

```
  },

      links = function()links)

}

#in questo vettore metto tutte le pagine che risultano dal search sul forum

#di solito sono pagine con nell'url un progressivo tipo pagina=1,pagina=2 etc..

#dovrei provare con l'rss feed

urlliststart <- c(

  "http://forum360.ru/search.php?searchid=145051")

#urlliststart1=paste(urlliststart[2],2:10,sep="")

#urlliststart=c(urlliststart[1],urlliststart1)

#rm(urlliststart1)

#in questa parte di codice mi creo il vettore urllist con l'elenco delle url con

#dentro post recuperati dal search

urllist=c()

for (i in c(1:length(urlliststart)) ){

  urlst <- urlliststart[i]

  ##estraggo tutti i link
```

```
  h1 = getLinks()

  htmlTreeParse(urlst, handlers = h1)

  url1 <- h1$links()

url1[grep('showthread',url1)]

  urllist1=url1[grep('showthread',url1)]

  urllist=c(urllist,urllist1)

}

rm(i)

urllistcrawl=urllist[!duplicated(x=urllist)]

#generazione lista url per crawl

#urllistcrawl=substring(urllist, first = 2, last = length(urllist))

#per ogni url con dentro i post vado a cercare gli articoli e li estraggo dalla

#pagina, tutti gli articoli finiscono nel vettore mydata.vector

v2=c()

v3=c()

v4=c()
```

```
v5=c()

vdata=c()

vtitolo=c()

vbody=c()

vautore=c()

v=c()

for (i in c(1:length(urllist)) ){

  url=paste("http://forum360.ru/",urllist[i],sep="")

  #enconding permette di stampare giusti i caratteri speciali

  doc <- htmlTreeParse(url, useInternalNodes=T, encoding="UTF-8")

  #cercare il tag corrispondente al contenuto che voglio

  #body

  nset <- getNodeSet(doc, "//blockquote[starts-with(@class, 'postcontent restore ')]" )

  v2 <- sapply(nset, xmlValue)

  v2 = v2[nchar(v2)>0]

  vbody=c(vbody,v2)

  #titolo
```

```
  nset <- getNodeSet(doc, "//span[starts-with(@class, 'threadtitle')]" )

  v3 <- sapply(nset, xmlValue)

  v3 = v3[nchar(v2)>0]

  vtitolo=c(vtitolo,v3)

  #data autore

  nset <- getNodeSet(doc, "//div[starts-with(@class, 'userinfo')]" )

  v4 <- sapply(nset, xmlValue)

  v4 = v4[nchar(v2)>0]

  vautore=c(vautore,v4)

  #data data

  nset <- getNodeSet(doc, "//span[starts-with(@class, 'date')]" )

  v5 <- sapply(nset, xmlValue)

  v5 = v5[nchar(v2)>0]

  vdata=c(vdata,v5)

}

vtitolo=gsub(pattern="[\t\n\r\f]",replacement="",x=vtitolo)
```

```
vdata=gsub(pattern="[\t\n\r\f]",replacement="",x=vdata)

vbody=gsub(pattern="[\t\n\r\f]",replacement="",x=vbody)

vautore=gsub(pattern="[\t\n\r\f]",replacement="",x=vautore)

v=cbind(vtitolo,vdata,vautore,vbody)

rm(i,v2,v3,v4,v5)

rm(vtitolo,vdata,vautore,vbody)

html2txt = function(str) {

  onestring = xpathApply(htmlParse(str, asText = TRUE),

                         "//body//text()",

                         xmlValue)[[1]]

  liststring = onestring

  return(liststring[[1]])

}

mydata.vector=c()

for (i in c(1:length(v)) ){

  mydata.vector[i] = html2txt(v[i])

}
```

```
rm(i)

write.csv(x=v,file="/Users/–Æ–ª—è/Desktop/chugun.csv")

#ora dobbiamo pulire i testi dagli accenti sbagliati come fatto nel programma

#sas sul server prova1--->excel??/corretto con encoding ora escono giusti i caratteri speciali
out <- file("/Users/werfgh/Desktop/textminingITA/russian/360_metall.txt", "r", encoding="utf-8")
v=read.table(file=out,header=T,sep="\n")
close(out)
library(koRpus)
library(tm)
#filter comments
comments=subset(v,!duplicated(v))
head(comments)
#creare df con lemming and non lemming
nlemmato= as.character(comments[,1])
#LEMMING/tokenize
#set.kRp.env(TT.cmd="/Users/werfgh/Desktop/textminingITA/treetag/cmd/tree-tagger-italian", lang="it")
#setto il file da usare, si pu√≤ usare soltanto da file importante!!!
lemmato= as.character(comments[,1])
infile <- file("/Users/werfgh/Desktop/lemmato.txt", "w", encoding="utf-8")
write.table(x=lemmato, file=infile,fileEncoding="utf-8", row.names=F, col.names=F,eol="")
close(infile)
# taggare testo
#tagged.text = treetag("/Users/werfgh/Desktop/lemmato.txt",treetagger="kRp.env",encoding="utf-8", debug=F, lang="it")
tagged.text <- treetag("/Users/werfgh/Desktop/lemmato.txt", treetagger="manual",
```

```
lang="ru",debug=T,
                TT.options=list(path="/Users/werfgh/Desktop/textminingITA/treetag/",
preset="ru"))
a=data.frame(tagged.text)
#head(a[a$tag=="NBC",])
summary(a$tag)
head(tagged.text,50)
a=a[a$lemma!='<unknown>',]
#filtri per analisi parole
mydata=a[grep(pattern="^[NA][^p].*",a$tag,ignore.case=F, value=F),]
#voglio creare una cosa che mi analizza da frase e non da singola parola
#CREAZIONE TABELLA CORPUS
mydata.corpus.lemmato = Corpus(VectorSource(mydata$lemma, encoding="UTF-8"),
readerControl=list(language="ru"))
mydata.corpus.nlemmato = Corpus(VectorSource(nlemmato, encoding="UTF-8"),
readerControl=list(language="ru"))
#per vedere cosa creo
#inspect(mydata.corpus[1])
#PROCESSO DI CLEANING
mydata.corpus.nlemmato = tm_map(mydata.corpus.nlemmato, FUN = tolower)
mydata.corpus.nlemmato = tm_map(mydata.corpus.nlemmato, FUN =
removePunctuation)
mydata.corpus.nlemmato = tm_map(mydata.corpus.nlemmato, FUN = stripWhitespace)
#ELIMINAZIONE STOPWORDS
out <- file("/Users/werfgh/Desktop/stopwords_russo.txt", "r", encoding="utf-8")
v=read.table(file=out, quote="", header=F, sep="\n")
v=as.character(v[,1])
close(out)
head(v)
my_stopwords = c(stopwords('russian'), v, "<unknown>")
my_stopwords=subset(my_stopwords,!duplicated(my_stopwords))
mydata.corpus.lemmato = tm_map(mydata.corpus.lemmato, FUN = removeWords,
my_stopwords)
```

```
mydata.corpus.nlemmato = tm_map(mydata.corpus.nlemmato, FUN = removeWords, my_stopwords)
#ANALISI DATI
mydata.dtm.lemmato = TermDocumentMatrix(mydata.corpus.lemmato)
mydata.dtm.nlemmato = TermDocumentMatrix(mydata.corpus.nlemmato)
frequenza=findFreqTerms(mydata.dtm.lemmato, 20)
#frequenza assoluta
#library(plyr)
#frequenzaassoluta1=count(mydata$lemma)
#frequenzaassoluta2=frequenzaassoluta1[order(-frequenzaassoluta1$freq),]
#frequenzaassoluta2=frequenzaassoluta2[frequenzaassoluta2$x!="<unknown>",]
#head(frequenzaassoluta2)
m = as.matrix(mydata.dtm.lemmato)
v = sort(rowSums(m), decreasing = TRUE)
myNames = names(v)
frequenzaassoluta = data.frame(word = myNames, freq = v)
rm(m,v, myNames)
write.table(frequenzaassoluta,file="/Users/werfgh/Desktop/temporaneo.txt",fileEncoding="utf-8")
library(wordcloud)
wordcloud(frequenzaassoluta$word, frequenzaassoluta$freq, min.freq = 6, colors = rainbow(nrow(frequenzaassoluta[frequenzaassoluta$freq >=6, ])))
#fare dictionary con parole chiave
out <- file("/Users/werfgh/Desktop/russian_tassonomy.txt", "r", encoding="utf-8")
v=read.table(file=out, quote="", header=F, sep="\t")
v=as.character(v[,2])
close(out)
diz <- Dictionary(v)
parolechiave=TermDocumentMatrix(mydata.corpus.lemmato, list(dictionary = diz))
inspect(parolechiave)
findFreqTerms(parolechiave,5)
m = as.matrix(parolechiave)
v = sort(rowSums(m), decreasing = TRUE)
```

```
myNames = names(v)
frequenzaassoluta1 = data.frame(word = myNames, freq = v)
rm(m,v, myNames)
head(frequenzaassoluta1)
write.table(frequenzaassoluta1,file="/Users/werfgh/Desktop/temporaneo.txt",fileEncodi
ng="utf-8")
#da aggiungere somma per tema
#clustering
# remove sparse terms to simplify the cluster plot
# Note: tweak the sparse parameter to determine the number of words.
# About 10-30 words is good.
mydata.dtm2 <- removeSparseTerms(mydata.dtm.lemmato, sparse=0.995)
# convert the sparse term-document matrix to a standard data frame
mydata.df <- as.data.frame(inspect(mydata.dtm2))
# inspect dimensions of the data frame
nrow(mydata.df)
ncol(mydata.df)
mydata.df.scale <- scale(mydata.df)
d <- dist(mydata.df.scale, method = "euclidean") # distance matrix
fit <- hclust(d, method="ward")
plot(fit) # display dendogram?
groups <- cutree(fit, k=5) # cut tree into 5 clusters
# draw dendogram with red borders around the 5 clusters
rect.hclust(fit, k=5, border="red")
#aggiungere associazione con parole chiavi
findAssocs(mydatasen.dtm.lemmato, term = 'kovka', corlimit = 0.95)

#creare df per frasi sentiment
 # inizializzo il vettore
 sentence = rep(1, length(a$tag))
 #analizzo il primo token
{
   if (a$tag[1] == "SENT") { #fullstop se non va#da sostituire con a$desc[] ==
```

```
\Sentence ending punctuation\""
    sentence[1] = 1
    sentencelast = 0
  } else {
    sentence[1] = 1
    sentencelast = 1
  }
}

#ciclo su tutte le successive parole per definire le sentences
if(length(a$tag)>1){
  for (j in 2:length(a$tag)) {
    if (a$tag[j - 1] == "SENT") {
      sentence[j] = sentencelast + 1
    } else {
      sentence[j] = sentencelast
    }
    sentencelast = sentence[j]
  }
}
#definisco l'id della frase
a$id_sentence = sentence

#definisco l'id della parola
#ciclo sulle frasi
id_parola = c()

for (j in 1:max(sentence)) {
  id_parola = c(id_parola, 1:length(a$id_sentence[a$id_sentence==j]))
}

a$id_parola = id_parola
```

```
  if (is.null(dim(df)) ) {
    df = a
  } else {
    df = rbind(df, a)
  }

}

# metto tutto in un dataframe
df_sent <- list()
df_tmp <- list()

  for (sent in 1:max(df$id_sentence)){
    tmp_sent = df[df$id_sentence==sent,]
    frase_orig = paste(tmp_sent$token, collapse= ' ')
    frase_lemma = paste(tmp_sent$lemma, collapse= ' ')
    df_tmp = data.frame (sent,frase_orig,frase_lemma)
    df_sent = rbind(df_sent, df_tmp)
  }
summary(df_sent)
head(df_sent)
#caricosentimentlist, decidere che set usare secondo me meglio usare #wtot
pwtot=read.table(file="/Users/werfgh/Desktop/pnwords/pwru.txt",sep="")
pwtot=as.character(pwtot$x)
nwtot=read.table(file="/Users/werfgh/Desktop/pnwords/nwru.txt",sep="")
nwtot=as.character(nwtot$x)

#prima possibilit√ †
library(tm)
library(tm.plugin.sentiment)
mydatasen.corpus.lemmato = Corpus(VectorSource(df_sent$frase_lemma,
encoding="UTF-8"), readerControl=list(language="ru"))
mydatasen.corpus.lemmato=score(mydatasen.corpus.lemmato)
```

```
mydatasen.dtm.lemmato = TermDocumentMatrix(mydatasen.corpus.lemmato)
sent1lem=polarity(mydatasen.dtm.lemmato, pwtot, nwtot)
write.table(sent1lem,file="/Users/werfgh/Desktop/temporaneo.txt")
summary(sent1lem)

#neg1lem=neg_refs_per_ref(mydatasen.dtm.lemmato, nwtot)
#pos1lem=pos_refs_per_ref(mydatasen.dtm.lemmato,pwtot)
#sentdiff1lem=senti_diffs_per_ref(mydatasen.dtm.lemmato, pwtot, nwtot)



#altra funzione
 library(plyr)
 library(stringr)
 sent2lem=score.sentiment(df_sent$frase_lemma, pwtot, nwtot)
 sent2lem$score
 score.sentiment = function(sentences, pos.words, neg.words, .progress='none')
 {
  require(plyr)
  require(stringr)

  # we got a vector of sentences. plyr will handle a list
  # or a vector as an "l" for us
  # we want a simple array ("a") of scores back, so we use
  # "l" + "a" + "ply" = "laply":
  scores = laply(sentences, function(sentence, pos.words, neg.words) {

    # clean up sentences with R's regex-driven global substitute, gsub():
    sentence = gsub('[[:punct:]]', '', sentence)
    sentence = gsub('[[:cntrl:]]', '', sentence)
    sentence = gsub('\\d+', '', sentence)
    # and convert to lower case:
    sentence = tolower(sentence)
```

```
  # split into words. str_split is in the stringr package
  word.list = str_split(sentence, '\\s+')
  # sometimes a list() is one level of hierarchy too much
  words = unlist(word.list)

  # compare our words to the dictionaries of positive & negative terms
  pos.matches = match(words, pos.words)
  neg.matches = match(words, neg.words)

  # match() returns the position of the matched term or NA
  # we just want a TRUE/FALSE:
  pos.matches = !is.na(pos.matches)
  neg.matches = !is.na(neg.matches)

  # and conveniently enough, TRUE/FALSE will be treated as 1/0 by sum():
  score = sum(pos.matches) - sum(neg.matches)

  return(score)
 }, pos.words, neg.words, .progress=.progress )

 scores.df = data.frame(score=scores, text=sentences)
 return(scores.df)
}
```

ITALIAN PROJECT EXAMPLE OF CODES:

```
#punti da cambiare per altri forum:
#popola:
#urlstart con le url del search
#nella grep tengo la root del forum che mi interessa per filtrare dei link
#nel getNodeSet devo mettere gli identificativi del tag con dentro i post
library(XML)

#funzione che serve per estrarre i link da una pagina web
getLinks = function() {
  links = character()
  list(a = function(node, ...) {
    links <<- c(links, xmlGetAttr(node, "href"))
    node
  },
       links = function()links)
}

#in questo vettore metto tutte le pagine che risultano dal search sul forum
#di solito sono pagine con nell'url un progressivo tipo pagina=1,pagina=2 etc..
#creare con comando paste pagine progressive
urlliststart <- c(

"http://wwwferroearteit.globalfreeforum.com/search.php?st=0&sk=t&sd=d&sr=posts&
keywords=INOX+acciaio"
  )

#in questa parte di codice mi creo il vettore urllist con l'elenco delle url con
#dentro post recuperati dal search

urllist=c()
for (i in c(1:length(urlliststart)) ){
```

```
  urlst <- urlliststart[i]
  ##estraggo tutti i link
  h1 = getLinks()
  htmlTreeParse(urlst, handlers = h1)
  url1 <- h1$links()
url1[grep('viewtopic',url1)]
  urllist1=url1[grep('viewtopic',url1)]
  urllist=c(urllist,urllist1)
}
rm(i)
#generazione lista url per crawl
urllistcrawl=gsub(pattern="./",
replacement="http://wwwferroearteit.globalfreeforum.com/", urllist)

#per ogni url con dentro i post vado a cercare gli articoli e li estraggo dalla
#pagina, tutti gli articoli finiscono nel vettore mydata.vector
v2=c()
v3=c()
v4=c()
vdata=c()
vtitolo=c()
vbody=c()
v=c()

for (i in c(1:length(urllistcrawl)) ){
  url= urllistcrawl[i]
  #enconding permette di stampare giusti i caratteri speciali
  doc <- htmlTreeParse(url, useInternalNodes=T, encoding="UTF-8")
  #cercare il tag corrispondente al contenuto che voglio
  #body
  nset <- getNodeSet(doc, "//div[starts-with(@class, 'content')]" )
  v2 <- sapply(nset, xmlValue)
  v2 = v2[nchar(v2)>0]
```

```
  vbody=c(vbody,v2)
  #titolo
  nset <- getNodeSet(doc, "//h3[starts-with(@class, 'first')]" )
  v3 <- sapply(nset, xmlValue)
  v3 = v3[nchar(v2)>0]
  vtitolo=c(vtitolo,v3)
  #data autore e data
  nset <- getNodeSet(doc, "//p[starts-with(@class, 'author')]" )
  v4 <- sapply(nset, xmlValue)
  v4 = v4[nchar(v2)>0]
  vdata=c(vdata,v4)

}
#prima pulizia per vettori
vtitolo=gsub(pattern="[\t\n\r\f]",replacement="",x=vtitolo)
vdata=gsub(pattern="[\t\n\r\f]",replacement="",x=vdata)
vbody=gsub(pattern="[\t\n\r\f]",replacement="",x=vbody)
#creazione dataframe
v=cbind(vtitolo,vdata,vbody)
rm(i,v2,v3,v4)
rm(vtitolo,vdata,vbody)

#html2txt = function(str) {
#  onestring = xpathApply(htmlParse(str, asText = TRUE),
 #                    "//body//text()",
  #                   xmlValue)[[1]]
  #liststring = onestring
  #return(liststring[[1]])
#}
#mydata.vector=c()
#for (i in c(1:length(v)) ){
 # mydata.vector[i] = html2txt(v[i])
#}
```

```
rm(i)

#cambiare l'encoding a seconda del sistema operativo
#MacRoman=MAC
out <- file("/Users/werfgh/Desktop/prova.csv", "w", encoding="MacRoman")
write.csv(x=v, file=out)
close(out)

#ora dobbiamo pulire i testi dagli accenti sbagliati come fatto nel programma
#sas sul server prova1--->excel??/corretto con encoding ora escono giusti i caratteri speciali
out <- file("/Users/werfgh/Desktop/textminingITA/italiano/fabbridarte/ferrobattuto_2.csv", "r", encoding="MacRoman")
v=read.csv(file=out, header=T, sep=",")
close(out)
library(koRpus)
library(tm)
#filter comments
comments=subset(v,!duplicated(v))
head(comments)
comments=comments[grep(pattern=".*ferro.*battuto.*",comments$vbody,ignore.case=F, value=F),]
#creare df con lemming and non lemming
nlemmato= as.character(comments$vbody)
#LEMMING/tokenize
#set.kRp.env(TT.cmd="/Users/werfgh/Desktop/textminingITA/treetag/cmd/tree-tagger-italian", lang="it")
#setto il file da usare, si pu√≤ usare soltanto da file importante!!!
lemmato= as.character(comments$vbody)
infile <- file("/Users/werfgh/Desktop/lemmato.txt", "w", encoding="utf-8")
write.table(x=lemmato, file=infile,fileEncoding="utf-8", row.names=F, col.names=F,eol="")
```

```
close(infile)
# taggare testo
#tagged.text = treetag("/Users/werfgh/Desktop/lemmato.txt",treetagger="kRp.env",encoding="utf-8", debug=F, lang="it")
tagged.text <- treetag("/Users/werfgh/Desktop/lemmato.txt", treetagger="manual", lang="it",debug=T,
                       TT.options=list(path="/Users/werfgh/Desktop/textminingITA/treetag/", preset="it"))
a=data.frame(tagged.text)
head(a[a$tag=="LS",])
head(tagged.text,50)
#filtri per analisi parole
filtro=c("NOM","ADJ","NPR")
mydata=a[a$tag==filtro,]
#voglio creare una cosa che mi analizza da frase e non da singola parola
#CREAZIONE TABELLA CORPUS
mydata.corpus.lemmato = Corpus(VectorSource(mydata$lemma, encoding="UTF-8"), readerControl=list(language="ita"))
mydata.corpus.nlemmato = Corpus(VectorSource(comments$vbody, encoding="UTF-8"), readerControl=list(language="ita"))
#per vedere cosa creo
#inspect(mydata.corpus[1])
#PROCESSO DI CLEANING
mydata.corpus.nlemmato = tm_map(mydata.corpus.nlemmato, FUN = tolower)
mydata.corpus.nlemmato = tm_map(mydata.corpus.nlemmato, FUN = removePunctuation)
mydata.corpus.nlemmato = tm_map(mydata.corpus.nlemmato, FUN = stripWhitespace)
#ELIMINAZIONE STOPWORDS
out <- file("/Users/werfgh/Desktop/stopwordsITAWIP.txt", "r", encoding="utf-8")
v=read.table(file=out, quote="", header=F, sep="\n")
v=as.character(v[,1])
close(out)
my_stopwords = c(stopwords('italian'), v, "<unknown>")
```

```
my_stopwords=subset(my_stopwords,!duplicated(my_stopwords))
mydata.corpus.lemmato = tm_map(mydata.corpus.lemmato, FUN = removeWords, my_stopwords)
mydata.corpus.nlemmato = tm_map(mydata.corpus.nlemmato, FUN = removeWords, my_stopwords)
#ANALISI DATI
mydata.dtm.lemmato = TermDocumentMatrix(mydata.corpus.lemmato)
mydata.dtm.nlemmato = TermDocumentMatrix(mydata.corpus.nlemmato)
frequenza=findFreqTerms(mydata.dtm.lemmato, 20)
#frequenza assoluta
#library(plyr)
#frequenzaassoluta1=count(mydata$lemma)
#frequenzaassoluta2=frequenzaassoluta1[order(-frequenzaassoluta1$freq),]
#frequenzaassoluta2=frequenzaassoluta2[frequenzaassoluta2$x!="<unknown>",]
#head(frequenzaassoluta2)
m = as.matrix(mydata.dtm.lemmato)
v = sort(rowSums(m), decreasing = TRUE)
myNames = names(v)
frequenzaassoluta = data.frame(word = myNames, freq = v)
rm(m,v, myNames)
write.table(frequenzaassoluta,file="/Users/werfgh/Desktop/temporaneo.txt")
library(wordcloud)
wordcloud(frequenzaassoluta$word, frequenzaassoluta$freq, min.freq = 10, colors = rainbow(nrow(frequenzaassoluta[frequenzaassoluta$freq >= 10, ])))
#fare dictionary con parole chiave
out <- file("/Users/werfgh/Desktop/italian_tassonomy.txt", "r", encoding="utf-8")
v=read.table(file=out, quote="", header=F, sep=";")
v=as.character(v[,2])
close(out)
diz <- Dictionary(v)
parolechiave=TermDocumentMatrix(mydata.corpus.lemmato, list(dictionary = diz))
inspect(parolechiave)
findFreqTerms(parolechiave,5)
```

```
m = as.matrix(parolechiave)
v = sort(rowSums(m), decreasing = TRUE)
myNames = names(v)
frequenzaassoluta1 = data.frame(word = myNames, freq = v)
rm(m,v, myNames)
head(frequenzaassoluta1)
write.table(frequenzaassoluta1,file="/Users/werfgh/Desktop/temporaneo.txt")
#da aggiungere somma per tema
#clustering
# remove sparse terms to simplify the cluster plot
# Note: tweak the sparse parameter to determine the number of words.
# About 10-30 words is good.
mydata.dtm2 <- removeSparseTerms(mydata.dtm.nlemmato, sparse=0.90)
# convert the sparse term-document matrix to a standard data frame
mydata.df <- as.data.frame(inspect(mydata.dtm2))
# inspect dimensions of the data frame
nrow(mydata.df)
ncol(mydata.df)
mydata.df.scale <- scale(mydata.df)
d <- dist(mydata.df.scale, method = "euclidean") # distance matrix
fit <- hclust(d, method="ward")
plot(fit) # display dendogram?
groups <- cutree(fit, k=5) # cut tree into 5 clusters
# draw dendogram with red borders around the 5 clusters
rect.hclust(fit, k=5, border="red")
#
#aggiungere associazione con parole chiavi
findAssocs(mydata.dtm.nlemmato, term = 'ferro', corlimit = 0.65)

#creare df per frasi sentiment
  # inizializzo il vettore
  sentence = rep(1, length(a$tag))
  #analizzo il primo token
```

```
{
  if (a$tag[1] == "SENT") { #fullstop se non va#da sostituire con a$desc[] == \Sentence ending punctuation\""
    sentence[1] = 1
    sentencelast = 0
  } else {
    sentence[1] = 1
    sentencelast = 1
  }
 }

 #ciclo su tutte le successive parole per definire le sentences
 if(length(a$tag)>1){
  for (j in 2:length(a$tag)) {
    if (a$tag[j - 1] == "SENT") {
      sentence[j] = sentencelast + 1
    } else {
      sentence[j] = sentencelast
    }
    sentencelast = sentence[j]
  }
 }
 #definisco l'id della frase
 a$id_sentence = sentence

 #definisco l'id della parola
 #ciclo sulle frasi
 id_parola = c()

 for (j in 1:max(sentence)) {
  id_parola = c(id_parola, 1:length(a$id_sentence[a$id_sentence==j]))
 }
```

```
  a$id_parola = id_parola

  if (is.null(dim(df)) ) {
    df = a
  } else {
    df = rbind(df, a)
  }

}

# metto tutto in un dataframe
df_sent <- list()
df_tmp <- list()

  for (sent in 1:max(df$id_sentence)){
    tmp_sent = df[df$id_sentence==sent,]
    frase_orig = paste(tmp_sent$token, collapse= ' ')
    frase_lemma = paste(tmp_sent$lemma, collapse= ' ')
    df_tmp = data.frame (sent,frase_orig,frase_lemma)
    df_sent = rbind(df_sent, df_tmp)
  }
summary(df_sent)
head(df_sent)
#caricosentimentlist, decidere che set usare secondo me meglio usare #wtot
pwtot=read.table(file="/Users/werfgh/Desktop/pnwords/pwita.txt",sep="")
pwtot=as.character(pwtot$x)
nwtot=read.table(file="/Users/werfgh/Desktop/pnwords/nwita.txt",sep="")
nwtot=as.character(nwtot$x)
#prima possibilit√ †
library(tm)
library(tm.plugin.sentiment)
mydatasen.corpus.lemmato = Corpus(VectorSource(df_sent$frase_lemma,
encoding="UTF-8"), readerControl=list(language="ita"))
```

```
mydatasen.corpus.lemmato=score(mydatasen.corpus.lemmato)
mydatasen.dtm.lemmato = TermDocumentMatrix(mydatasen.corpus.lemmato)
sent1lem=polarity(mydatasen.dtm.lemmato, pwtot, nwtot)
write.table(sent1lem,file="/Users/werfgh/Desktop/temporaneo.txt")
summary(sent1lem)

#neg1lem=neg_refs_per_ref(mydatasen.dtm.lemmato, nwtot)
#pos1lem=pos_refs_per_ref(mydatasen.dtm.lemmato,pwtot)
#sentdiff1lem=senti_diffs_per_ref(mydatasen.dtm.lemmato, pwtot, nwtot)



#altra funzione
 library(plyr)
 library(stringr)
 sent2lem=score.sentiment(df_sent$frase_lemma, pwtot, nwtot)
 sent2lem$score
 score.sentiment = function(sentences, pos.words, neg.words, .progress='none')
 {
  require(plyr)
  require(stringr)

  # we got a vector of sentences. plyr will handle a list
  # or a vector as an "l" for us
  # we want a simple array ("a") of scores back, so we use
  # "l" + "a" + "ply" = "laply":
  scores = laply(sentences, function(sentence, pos.words, neg.words) {

    # clean up sentences with R's regex-driven global substitute, gsub():
    sentence = gsub('[[:punct:]]', '', sentence)
    sentence = gsub('[[:cntrl:]]', '', sentence)
    sentence = gsub('\\d+', '', sentence)
    # and convert to lower case:
    sentence = tolower(sentence)
```

```
    # split into words. str_split is in the stringr package
    word.list = str_split(sentence, '\\s+')
    # sometimes a list() is one level of hierarchy too much
    words = unlist(word.list)

    # compare our words to the dictionaries of positive & negative terms
    pos.matches = match(words, pos.words)
    neg.matches = match(words, neg.words)

    # match() returns the position of the matched term or NA
    # we just want a TRUE/FALSE:
    pos.matches = !is.na(pos.matches)
    neg.matches = !is.na(neg.matches)

    # and conveniently enough, TRUE/FALSE will be treated as 1/0 by sum():
    score = sum(pos.matches) - sum(neg.matches)

    return(score)
  }, pos.words, neg.words, .progress=.progress )

  scores.df = data.frame(score=scores, text=sentences)
  return(scores.df)
}
```