FP7-ICT Strategic Targeted Research Project TrendMiner (No. 287863)

Large-scale, Cross-lingual Trend Mining and Summarisation of Real-time Media Streams

D8.2.2 Market Watch - v2

Francesca Spagnoli (Editor, EK), Francesco Bellini (EK), Paul Ringler (SORA), France Lasfargues (IMR), Chloé Martin (IMR)

Abstract

FP7-ICT Strategic Targeted Research Project TrendMiner (No. 287863)
Market Watch v2 (WP8)

This deliverable provides an updated version of Deliverable 8.2.1 Market Watch v1.
We include here information about the 2013 Social Media landscape which forms the background for the following market analysis, providing details on current vendors and providers for Sentiment Analysis services, including also estimations for the 2014 Market on commercial products similar to TrendMiner. We describe also a larger set of existing providers' tools based on use cases similar to Trendminer, taking into account existing Content Analytics providers for sentiment analysis.

Keyword list: Social Media, Market Analysis, Financial Use Case, Political Use Case

Nature: Report
Dissemination: PU
Contractual date of delivery: 1.05.2014
Actual date of delivery: 02.05.2014
Reviewed By: Daedalus, DFKI
Web links: http://trendminer-project.eu/

TrendMiner Consortium

This document is part of the TrendMiner research project (No. 287863), partially funded by the FP7-ICT Programme.

DFKI GmbH
Language Technology Lab
Stuhlsatzenhausweg 3
D-66123 Saarbrücken,Germany
Contact person: Thierry Declerck
E-mail: declerck@dfki.de

University of Southampton
Southampton SO17 1BJ, UK
Contact person: Mahensan Niranjan
E-mail: mn@ecs.soton.ac.uk

Internet Memory Research
45 ter rue de la Rvolution
F-93100 Montreuil, France
Contact person: France Lafarges
E-mail: contact@internetmemory.org

Eurokleis S.R.L.
Via Giorgio Baglivi, 3
Roma RM 0016, Italia
Contact person: Francesco Bellini
E-mail: info@eurokleis.com

University of Sheffield
Department of Computer Science
Regent Court, 211 Portobello St.
Sheffield S1 4DP, UK
Contact person: Kalina Bontcheva
E-mail: K.Bontcheva@dcs.shef.ac.uk

Ontotext AD
Polygraphia Office Center fl.4,
47A Tsarigradsko Shosse,
Sofia 1504, Bulgaria
Contact person: Atanas Kiryakov
E-mail: naso@sirma.bg

Sora Ogris and Hofinger GmbH
Bennogasse 8/2/16
A-1080 Wien, Austria
Contact person: Christoph Hofinger
E-mail: ch@sora.at

Hardik Fintrade Pvt Ltd.
227, Shree Ram Cloth Market,
Opposite Manilal Mansion,
Revdi Bazar, Ahmedabad 380002, India
Contact person: Suresh Aswani
E-mail: m.aswani@hardikgroup.com

DAEDALUS - DATA, DECISIONS AND LANGUAGE, S. A.
C/ López de Hoyos 15, 3º, 28006 Madrid,
Spain
Contact person: José Luis Martínez Fernández
Email: jmartinez@daedalus.es

Institute of Computer Science Polish Academy of Sciences
5 Jana Kazimierza Str.,Warsaw, Poland
Contact person: Maciej Ogrodniczuk
E-mail: Maciej.Ogrodniczuk@ipipan.waw.pl

Universidad Carlos III de Madrid
Av. Universidad, 30, 28911, Madrid, Spain
Contact person: Paloma Martínez Fernández
E-Mail: pmf@inf.uc3m.es

Research Institute for Linguistics of the Hungarian Academy of Sciences
Benczúr u. 33., H-1068 Budapest, Hungary
Contact person: Tamás Váradi
Email: varadi.tamas@nytud.mta.hu

Executive Summary

The Market Watch Analysis provided in this deliverable examines the different business opportunities and current market scenarios for the services similar to TrendMiner use cases. The document focuses on providing a clear scenario based on updated information of Market Watch v1 (Deliverable 8.2.1). The analysis of the potential competitors of the financial and political use cases constitutes a fundamental step for the definition of the potential business plans to be developed by the project itself and by the commercial partners of TrendMiner (this analysis is given in Deliverable 8.3.2 Business Plan v2). Specifying the targeted markets, exploring the competition and identifying potential commercial opportunities is relevant also for the definition of the final exploitation plans (to be included in Deliverable 8.1.3 Dissemination and exploitation Report v2).

In Chapter 1 we provide a general description of the TrendMiner project.
In Chapter 2 we describe the Social Media evolution and we provide updated information on the Social Media Landscape 2013. We analyse also the Social Media Ecosystem taking into account the Hype Cycle for Media and Entertainment provided by Gartner. This analysis is essential in order to identify the stage of maturity of Sentiment Analysis services within the overall Social Media framework. We provide here also an updated description of twitter users' typologies and statistics about twitter usage in 2013.
In Chapter 3 we describe in detail the market for the services based on TrendMiner by analysing the 2013 Hype Cycle for Content and Social Analytics provided by Gartner. This analysis shows that Sentiment Analysis tools are currently entering the phase of "Plateau of Productivity", together with Web and Predictive Analytics. The Content Analytics market is in the "Adolescent" phase and the market penetration is from 5% to 20% of the target audience. This implies that when the TrendMiner project ends, market enlargement will be in process, making commercial introduction feasible and economical and financially sustainable. In this document we take into account also the predictions for the period 2013-2017 related to the expected market size. Indeed, according to an International Data Corporation, the worldwide CAGR for the market is expected to be 9.7%. However, the growth of individual segments of the market varies from 6.5% for production planning applications to 11.8% for CRM analytic applications. We provide here also information about the big vendors market share, comparing the results achieved in the different business segments.
In Chapter 4 we link to an updated description of the TrendMiner software and services developed, relevant for the development of Business Plan in Deliverable 8.3.2.
In Chapter 5 we analyse the competitors in the field of Data Analysis in order to provide a complete the description of the market, taking into account also the services provided similar to the ones developed by the Internet Memory partner of TrendMiner.
In Chapter 6 we analyse in detail the Content Analytics Providers, taking into account particularly the 15 most influent providers for financial sentiment analysis. For each provider we describe also in detail the activities developed, its pricing strategies, source of data extraction, strengths and weaknesses of the product/service.

Contents

Figures

Tables

Tables of Abbreviations

API Application Programming Interface

B2B Business-to-Business

BI Business Intelligence

CAGR Compound Annual Growth Rate

CPM Corporate Performance Management

CRM Client Relations Management

C-SPARQL Continuous SPARQL

DBMS Data Base Managment System

DSMS Data Stream Managment Systems

GATE General Architecture for Text Engineering

GEMET General Multilingual Environmental Thesaurus

HLT Human Language Technology

IE Information Extraction

IGGSA Interest Group on German Sentiment Analysis

LD Linked Data

LOD Linked Open Data

MCA Media Content Analysis

NLP Natural Language Processing

OBIE Ontology-Based Information Extraction

OSN Online Social Networks

OSS Open Source Software

POS Part of Speech

PR Public Relations

RDF Resource Description Framework

R&D Research & Development

REST Representational State Transfer

RSS Really Simple Syndication

SEC Security and Exchange Commission

SPARQL Protocol and RDF Query Language

TRNA Thomson Reuters News Analytics

URI Uniform Resource Identifier

WP Workpackage

1 Relevance to TrendMiner

TrendMiner is a Research and Development (R&D) project, combining 6 academic and research institutions and 6 SMEs interested in including innovative technologies in their product and services. In the course of TrendMiner, results are validated in high-profile case studies: Financial decision support, with analysts, traders, regulators, and economists as possible target group, and political analysis and monitoring, with politicians, economists, and political journalists. In a recent extension of the consortium of the project, we are dealin gnow also with eHealth and psychologycall topics. It is very essential in those fields to have a real time access to streaming media. Since TrendMiner will not propose a purely research platform, but also economically viable solutions, this market watch is a pre-requisite in order to position the use cases of TrendMiner. While TrendMiner is not exclusively dealing with Social Media, but with different types of streaming data, the deliverable is focusing updating the information on this topic.

2 Background

2.1 Social Media and its evolution

Social media is a family of online applications which enables users personifying different roles (customers, employees, consumers, companies, institutions, etc.) to share, co-create, discuss and modify user-generated content. This can happen on a variety of platforms, most of them being web based, but in recent times also based more and more on mobile devices. In other words, the interesting novelty in Social Media services is the fact the user is at the same time the data provider and no longer only an expert (writer) in a specific field. In this context, we see often the term ”user generated content”. The relevant Wikipedia article states here that “The advent of user-generated content marked a shift among media organizations from creating online content to providing facilities for amateurs to publish their own content”[1].

2.1.1 A short History of Social Media

These Social Media technologies were developed few years back – the oldest being blogs during 1998-1999, LinkedIn, Myspace 2003, Facebook in 2004 and Twitter in 2006. Late 2009 and 2010 however, were the years when Social Media gained true popularity and respect among consumers, brands and institutions. Twitter became a platform for breaking news and keeping the social world updated anywhere and at anytime. The number of Facebook users grew rapidly and so did for LinkedIn, Instagram, YouTube and etc. The consensus among many experts nowadays is that Social Media cannot be ignored, when considering matters of sentiment or communication within a wider public.

Today Facebook has more than 1 billion users and this growth reaches beyond internet-enabled computers, as smartphone proliferation creates hundreds of millions of mobile consumers that make social networks their on-the-go digital portal. The

[1] http://en.wikipedia.org/wiki/User-generated_content (last access 29.10.2012)

shift from desktop access to mobile has not only increased the flow of information, but made it more real-time based.

2.1.2 Types of Social Media[2]

The Social Media technologies encompass Internet forums, blogs, wikis, micro blogging, social networking sites (Twitter, Facebook, LinkedIn to name the most visited) and platform for photos and videos (Flickr, YouTube). The usage of Social Media is not restricted to personal activities and networking, but also includes professional information and connections. Companies and institutions may connect with their current and potential employees, Journalists, Politicians, and engaged citizens may exchange information, professional traders may use them to gain insights into possible market moving events.

Figure 1 Type of Social MediaTypes of Social Media[3]

[2] http://www.mediassociaux.fr/2011/02/06/description-des-differents-types-de-medias-sociaux/ (last access: 29.10.2012)

[3] http://www.mediassociaux.fr/2011/02/06/description-des-differents-types-de-medias-sociaux/ (last access: 29.10.2012)

In more detail, we can distinguish several types of Social Media:

- **Forum:** A public discussion space where messages are displayed in chronological order.
- **Blog:** A simplified publishing tool where items are displayed in chronological order and sorted into categories. Readers may submit comments that are moderated afterwards. The RSS feed allows easy export of content to aggregators and readers.
- **Wiki:** An online knowledge base where users themselves write and edit articles on a any number of subjects.
- **Microblog**: Service publishing, sharing and discussion based on very short messages. Accessible on the web, mobile devices, or through applications. E.g.: Twitter, Google Buzz, WhatsApp, etc.
- **Social Network:** Site with restricted access where each user has a profile. Members are linked bilaterally or through groups. Some networks also offer more sophisticated features (messaging, publication and sharing content) and the ability to host third-party applications (platform). E.g.: Facebook, Xing, LinkedIn, Orkut, etc.
- **P2P Sharing:** Online service where users can post pictures, videos, links and etc. Each element is linked to a published member and can be commented and rated. The notorious “Pirate Bay” service is among these services, but also other sites like YouTube, flickr, etc.
- **Aggregator:** Online service to bring together all the publications of a user of Social Media (social stream). E.g.: FriendFeed, Hootsuite, etc.
- **Social gaming:** Online games based on a social platform operator member profiles to offer different social interactions between players. E.g. FarmVille
- **Geolocation service:** Applications to publish share and discuss over mobile devices. Articles or photos published are attached to a place to give them a geographic context. E.g. 4Square

2.1.3 Characteristics of Social media

Social media have specific characteristics which differentiates it from traditional media (Mislove et al 2007):

- **User Based:** The norm before the Social Media was that the content of the web pages was uploaded by a single entity (editor, company, etc.) and was read by other users on the Web. The flow of information was taken to be unidirectional, much like classical print media. By contrast, content on Social Media is created by the users, with minimal intervention from a moderator. The information can come from anyone who wants to participate. This combination of speed and variety makes Social Media attractive and useful for the users.
- **Interactive:** Many Social Media are essentially applications that enable users to interact with each other. This is quickly becoming a pastime that more people are choosing over television – because it's more than just entertainment, it's a way to connect and have fun with friends.
- **Community-Driven:** Social networks are built and thrive from community concepts. They provide virtual groups for people who share common beliefs or hobbies and allow them to share their views and interact with people having similar views.

- **Relationships/Connections:** Unlike websites of the past, social networks thrive on relationships. The more relationships that you have within the network, the more established you are toward the center of that network.
- **Emotion over content:** While traditional websites where focused on providing content, Social Media provides people with emotional support since they have their friends within easy reach.
- **Unorganized content:** Due to the fact that the information of the Social Media is added by any user, this content is highly unorganized and unstructured as compared to any website of the traditional web. Information is often incomplete, sentences and words often truncated, and posted out of context.
- **Addictive:** The major reason for the success of the Social Media is its addictiveness due to some of the above mentioned characteristics. According to estimates each user spends on an average 7 hours a month on Facebook. Compared to traditional media, this characteristic brings about many repeats, redirections, which distort the concept of weak/strong signals.

2.2 The Social Media Ecosystem

2.2.1 The 2013 scenario

In order to identify the future of the Social Media Ecosystem it is required to identify the current situation of Social Media Adoption. In the following figure provided by Gartner it is shown the 2013 scenario and future landscape for Media and Entertainment sector.

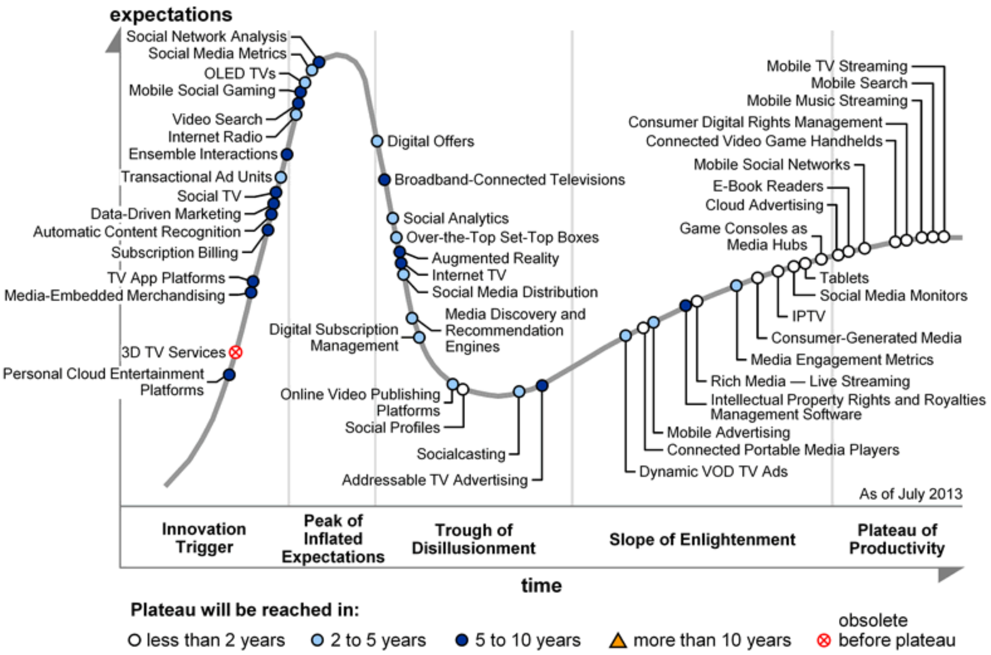


Figure 2 Hype Cycle for Media and Entertainment 2013[4]

[4] McGuire M., Hype Cycle for Media and Entertainment, Gartner 2013, available at http://www.gartner.com/document/2561116?ref=QuickSearch&sthkw=hype%20cycle%202013%20for%20media%20and%20entertainment%202013

As shown by the figure 2, Social Media and more specifically Mobile technologies for Social Media contents are now in the Plateau of Productivity and in less than 2 years will reach a mainstream adoption by producing high benefits for the society. Social Media Metrics are now in the peak of the infleated expectations and Social Analytics in the Trough of disillusionment, it will take from 2 to 5 years to mainstream adoption of these services. Social analytics will have an higher impact than Social Media Metrics. The future during the next 10 years will be dominated by Automatic Content Recognition, Data Driven Marketing and Internet TV that will constitute the most transformational technologies within the Media and Entertainment sector. Also Social Network Analysis, Social TV, Augmented Reality, Mobile Social Gaming will produce relevant benefits for the society in the next 10 years.

The past years have seen a rapid growth of Social Media offers on the Internet. Still, a visible structure has emerged. Cavazza[5] confirms that three social networks dominated the Social Media landscape also in 2013: Facebook, Google+ and Twitter. The large number of users they have attracted and the nature of services and applications they offer, suggests that this triad could remain stable for the next future, as described by Cavazza: *"I don't believe one can eat the two others, since each one have a distinct orientation: Twitter for content discovery, Google+ to manage your online identity and Facebook to interact with your friends".*
With respect to the image provided in 2012, Cavazza does not combine the applications/social networks to the devices used. There is not still a clear distinction among the activities developed through these applications, which cover more than one objective and scope. Hence, the Social Media presented in the figure are often cross-cutting. The cycle is based only on 4 activities: publishing, sharing, networking and discussing (in the figure from 2012 there were 2 main activities: conversations and interactions, and 6 secondary activities: publishing, sharing, playing, networking, buying and localization). It is interesting to note that a diverse and dense ecosystem of applications and social networks has been included into these three main networks. Several Social Networks disappeared from the cycle (such as: zynga, foursquare, kobojo, dailymotion, path, hunch, scribd, bosket, yelp, etc …), new others were included in the cycle (such as fancy, xing, github, kik, disqus, vine, overblog, mahalo, etc …) and finally others has been changed their branding strategies, changing their logo as twitter, myspace, Spotify.

[5] Cavazza F., Social Media Landscape 2013
http://www.fredcavazza.net/2013/04/17/social-media-landscape-2013/ (access 03.03.2014)

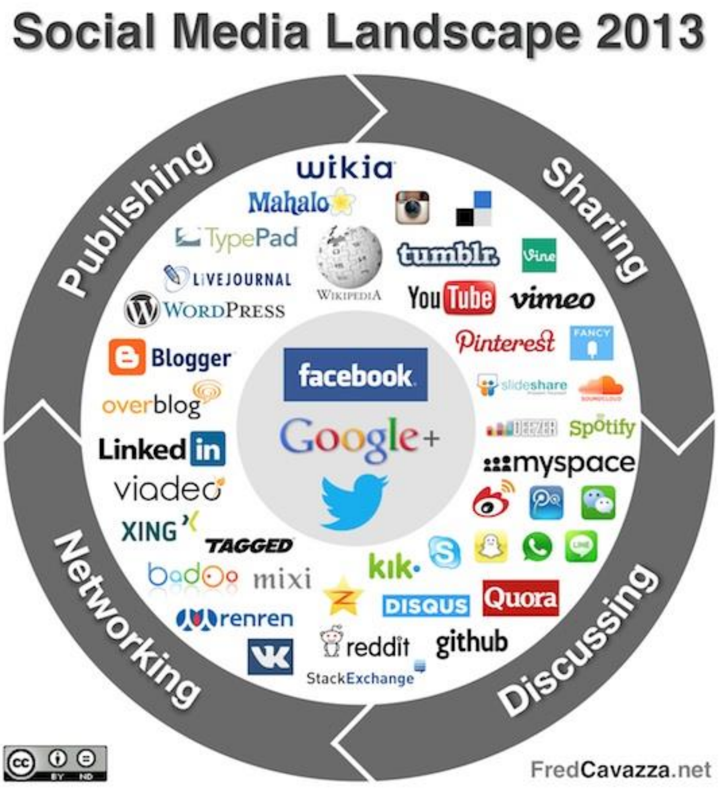


Figure 3 Social Media Landscape in 2013[6]

In addition to social networks that fill uncovered niches, another aspect of this ecosystem is the development of applications that thrive off the substantial amount of information generated by these networks, forming relationships with social networks that might be described as symbiotic.

For example, Twitter makes individual tweets and associated meta-information available via a specialised API. According to Twitter Director of Platform, Ryan Sarver[7], as of 2011, there were about 750,000 registered applications accessing the data streams provided by Twitter. Development hotspots include:

- **Publisher tools:** Companies such as SocialFlow help publishers optimize how they use Twitter, leading to increased user engagement and the production of the right tweet at the right time.
- **Curation:** Mass Relevance and Sulia provide services for large media brands to select, display, and stream the most interesting and relevant tweets for a breaking news story, topic or event.
- **Real-time data signals:** Hundreds of companies use real-time Twitter data as an input into ranking, ad targeting, or other aspects of enhancing their own core products. Klout is an example of a company which has taken this to the next level by using Twitter data to generate reputation scores for individuals. Similarly, Gnip syndicates Twitter data for licensing by third parties who want to use our real-time corpus for numerous applications (everything from hedge

[6] Image taken from: http://www.fredcavazza.net/2013/04/17/social-media-landscape-2013/

[7] https://dev.twitter.com/blog/changes-coming-to-twitter-api (last access 29.10.2012)

funds to ranking scores). Gnip has been recently acquired by Twitter[8]. In a recent development, Twitter is cooperating closely with Thomson Reuters on sentiment analysis in financial markets.

- **Value-added content and vertical experiences:** Emerging services like Formspring, Foursquare, Instagram and Quora have built into Twitter by allowing users to share unique and valuable content to their followers, while, in exchange, the services get broader reach, user acquisition, and traffic. In this regard, twitter conducted, but ultimately shut down its own experiment with #Music, a platform to recommend and discover music through other Twitter users.
- **Social CRM, entreprise clients, and brand insights**: Companies such as HootSuite, CoTweet, Radian (that is now part of Salesforce exact target marketing cloud), Seesmic, and Crimson Hexagon help brands, enterprises, and media companies tap into the zeitgeist about their brands on Twitter, and manage relationships with their consumers using Twitter as a medium for interaction.

According to this classification, TrendMiner is located in the Social CRM or B part of this ecosystem, and therefore well located in an area, where Twitter is actively encouraging development.

2.3 TrendMiner and Social Media

Online media and user-authored content (e.g. weblogs, Facebook and of course, Twitter) are nowadays a major platform for the exchange of information, which is not necessarily only of personal nature. We see that many public personalities, like politicians, also use Twitter for getting messages out to their clientele, and journalists or financial analysts also increasingly rely on alternative means of communication than the “classical” online platforms of email or websites.

The form and the increasing volume of such social and user-authored content has led to challenges of how to access and interpret these strongly multilingual data, in a timely, efficient and affordable manner. In the case of Twitter, this fact gave rise to many of the above mentioned applications. Due to the fact that authors of such TrendMiner content rarely spend a lot of time composing messages, the level completeness and correctness of communications via Social Media is often low. Messages are short and noisy and their interpretation often require references to other equally short and noisy messages, which might have been posted by the same person or by any other source. Language technology and semantic analysis are confronted with serious data quality issues here. TrendMiner is addressing this type of scientific challenge and aims at delivering innovative, portable, real-time methods for cross-lingual mining and summarisation of large-scale stream media. Summarisation is important since one cannot expect to have tools analysing every messages, but rather to first have a clustering or classification of topics extracted in an efficient and robust way from a large amount of streaming data. TrendMiner aims at achieving its goals through an inter-disciplinary approach. It combines deep linguistic methods of text processing, to be applied on summaries, knowledge-based reasoning from web science, machine learning, economics and political science. The last two points are

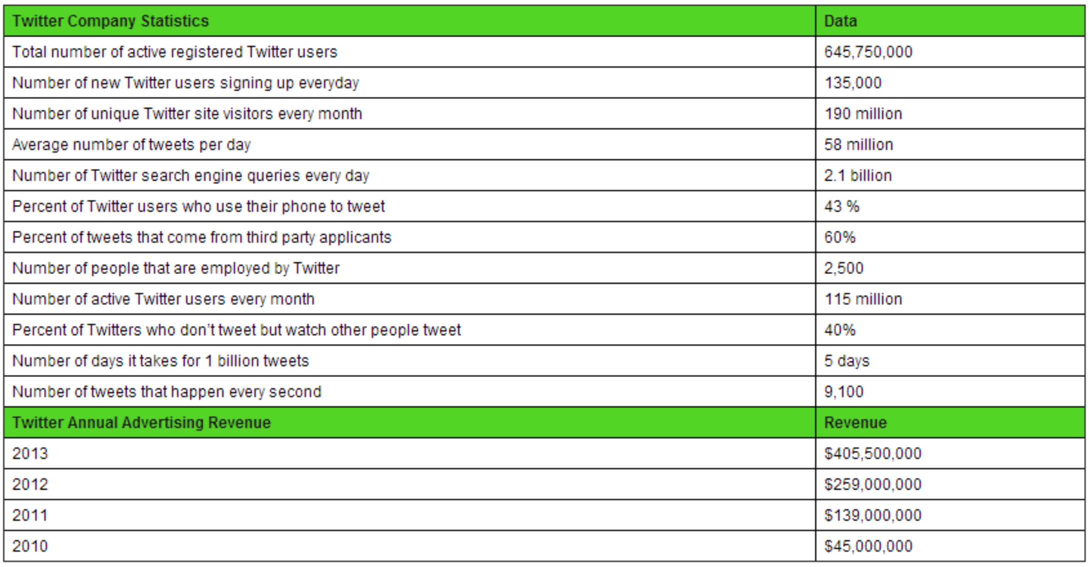
[8] http://techcrunch.com/2014/04/15/twitter-acquires-longtime-partner-and-social-data-analytics-provider-gnip/

related to the use cases, financial decision support (with analysts, traders, regulators, and economists), and political analysis and monitoring (with politicians, economists, and political journalists), that are validating TrendMiner results.

2.4 Classification of Twitter Users

Since TrendMiner is dedicated particularly to the detection of opinions, sentiments and trends developed through social networking and microblogging services offered by Twitter, we take a short look at its users.

The number of Twitter's users is continuously growing, as shown in the table below, on the 1st January 2014 the total number of active registered users were 645.750.000 and the annual advertising revenue for year 2013 were $405.500.000, almost doubled from 2012. The research detected that everyday Twitter's users send 58 million of tweets and 9,100 tweets every second[9].



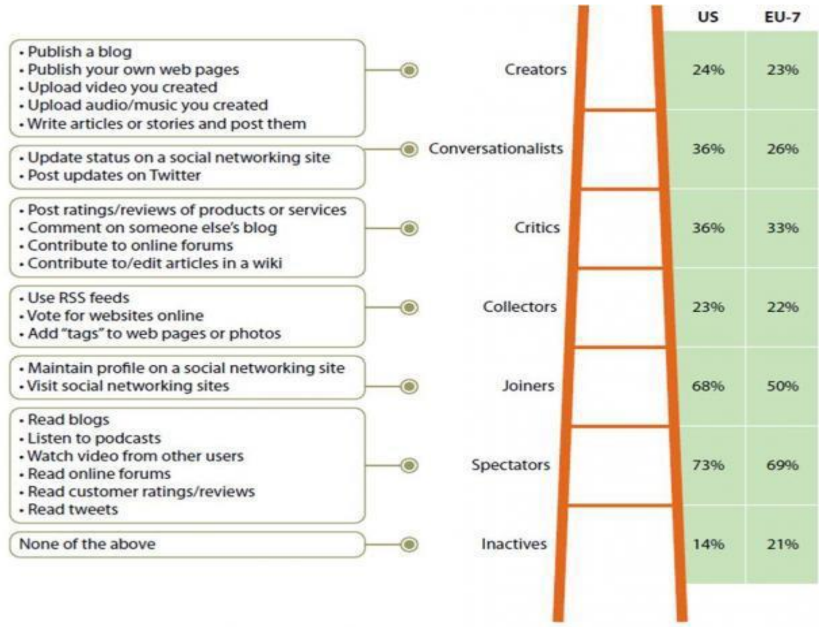
Twitter Company Statistics	Data
Total number of active registered Twitter users	645,750,000
Number of new Twitter users signing up everyday	135,000
Number of unique Twitter site visitors every month	190 million
Average number of tweets per day	58 million
Number of Twitter search engine queries every day	2.1 billion
Percent of Twitter users who use their phone to tweet	43 %
Percent of tweets that come from third party applicants	60%
Number of people that are employed by Twitter	2,500
Number of active Twitter users every month	115 million
Percent of Twitters who don't tweet but watch other people tweet	40%
Number of days it takes for 1 billion tweets	5 days
Number of tweets that happen every second	9,100
Twitter Annual Advertising Revenue	**Revenue**
2013	$405,500,000
2012	$259,000,000
2011	$139,000,000
2010	$45,000,000

Table 1 Twitter 2013 statistics

Users of Twitter (and of Social Media in general) are a heterogeneous group, ranging from occasional to very frequent, from amateur to experts, etc. We present a classification of Social Media users, proposed in the context of "Social Technographics"[10] , displaying the distribution of users in the form of a ladder.

[9] Statistic Brain, Twitter statistics, 2014, available at http://www.statisticbrain.com/twitter-statistics/ (last access 01.01.2014)

[10] http://blogs.forrester.com/gina_sverdlov/12-01-04-global_social_technographics_update_2011_us_and_eu_mature_emerging_markets_show_lots_of_activity (last access: 29.10.2012)


US
EU-7
• Publish a blog
• Publish your own web pages
• Upload video you created
• Upload audio/music you created
• Write articles or stories and post them
Creators
24%
23%
• Update status on a social networking site
• Post updates on Twitter
Conversationalists
36%
26%
• Post ratings/reviews of products or services
• Comment on someone else's blog
• Contribute to online forums
• Contribute to/edit articles in a wiki
Critics
36%
33%
• Use RSS feeds
• Vote for websites online
• Add "tags" to web pages or photos
Collectors
23%
22%
• Maintain profile on a social networking site
• Visit social networking sites
Joiners
68%
50%
• Read blogs
• Listen to podcasts
• Watch video from other users
• Read online forums
• Read customer ratings/reviews
• Read tweets
Spectators
73%
69%
None of the above
Inactives
14%
21%

Base: 57,924 US online adults (18+); 16,473 European online adults (18+)

Source: North American Technographics® Online Benchmark Survey, Q3 2011 (US, Canada); European Technographics Online Benchmark Survey, Q3 2011

Figure 4 Ladder of activity levels of users of Social Media[11]

The authors notice that even if only a small part of the society is using Twitter, it turns out that a large proportion of Twitter writers are very influential individuals. This is corroborated by our own observations, e.g. through interviews with party managers and PR professionals who deal with Twitter on a daily basis. Among others, users relevant for Trendminer range from journalists and politicians to engaged citizens, company directors and stock market traders. The high level of influence that can be found among users of Twitter is one of the reasons, why Twitter is privileged in our selection of data sources for TrendMiner.

It is interesting to show also a more recent analysis of Twitter users provided by Steve Faktor[12] in 2013, who detected 10 typologies of twitterers (the analysis is based on statistics from “beevolve” and proprietary sampling of 25 random accounts for each category):

1. **Undead:** 60% of twitters accounts are inactive. Some of them are second accounts and are not used.
2. **Protector:** are mainly locked accounts and it corresponds to the 12% of the population on twitter.
3. **Chirper:** it is the largest sample of twitter’s users. It is really difficult to identify specific characteristics in terms of age, gender, etc … within this set of users and they do not focus on specific topics. This category of users does not have a clear idea of why they use Twitter.
4. **Fan:** is the second category in terms of density of population on Twitter. They are not sending tweets, but they often retweet or send comments. They use

[11] Illustration taken from Li & Bernoff (2011)

[12] Faktor S., The 10 types of twitter users, 2013, available at http://www.ideafaktory.com/social-media/on-forbes-the-10-types-of-twitter-users-and-how-to-make-them-to-love-you-2/ (last access 03.03.2014)

Twitter mainly to follow the celebrities, but they do not discuss topics of interest.

5. **Networker:** this category is formed by people aimed to create new contacts often for improving their careers or to attract influencers in their network. It is a considerable sample and it is positioned on the axes of fame.
6. **Scouts:** it is a small sample of users that share photos, ideas, experiences. They have networks that can amplify the impact and diffusion of the information shared.
7. **Stars:** it is a very small sample and it is formed by celebrities. The top 1000 twitter account have over 3 billion of followers.
8. **E-lebrities:** this category is formed by comedians, writers, top podcasters, authors, tech gurus and columnists. They make profit from using Twitter.
9. **MediaCo:** this category is formed mainly by big media and entertainment companies.
10. **Organizations:** this category is formed by companies that often use Twitter as a tool for monitoring brand perception or as a customer service channel.

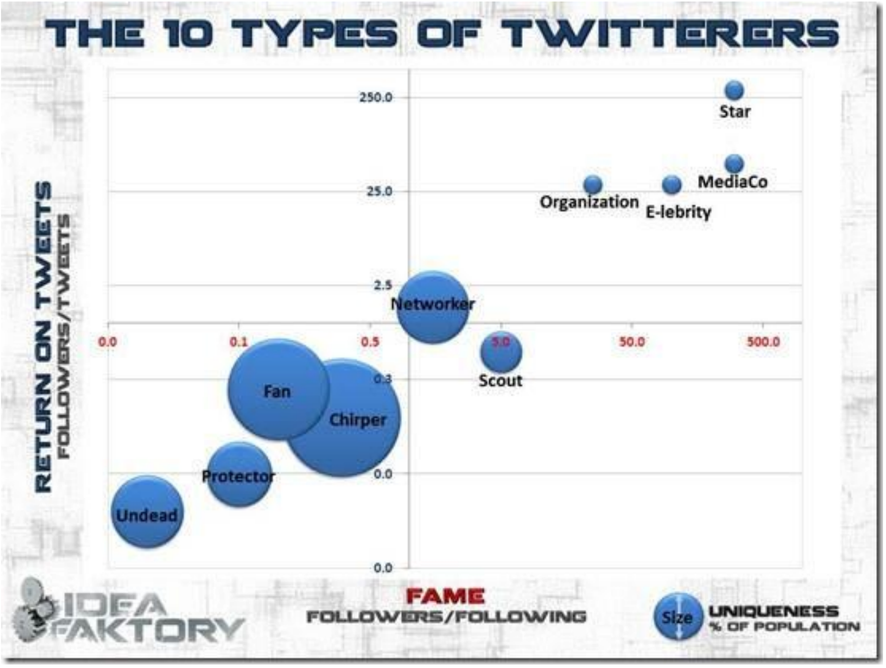


Figure 5 The 10 types of twitterers[13]

[13] Ibidem

3 The Market for Services based on TrendMiner

3.1 General Overview

The target market of services based on TrendMiner will consist of investors, traders, analysts, financial institutions, industries, politicians, policy makers, journalists, bloggers and interested citizens. Apart of this, it is clear that social scientists could also benefit from TrendMiner results, since it offers a way to collect substantial amounts of analysed data, giving insights on social interaction in Social Media, political attitudes or the communicative part of financial markets. The market analysis of TrendMiner will consider not only the use of Social Media for content and sentiment analysis, but will investigate the possibility of also covering more conventional online media streams (such as comment sections of online newspapers, etc …).The most relevant benefit of TrendMiner is the ability to develop an accurate detection of opinion and sentiments on a variety of subjects related to the financial and political sphere, also through Social Media, that may be shared by large communities.

3.2 Business Environment

Sentiment analysis can be seen as a market segment of Enterprise Performance Management and Content Analytics. Business Intelligence applications are predominant in this market. Many of these applications are Business Intelligence (BI) and Corporate Performance Management (CPM) tools serving purposes such as CRM, controlling, data warehousing, knowledge management and collaboration. Many of them use techniques also used in TrendMiner, such as text analytics, regression methods, stream reasoning, etc. Such functions are also centrepieces of many Social Media Monitoring applications. Currently, most of the services and consultancy offered in the market for these applications focus on marketing professionals from either companies that are branching out their sales channels into Social Media, or companies that base their business plan on eCommerce from the beginning.

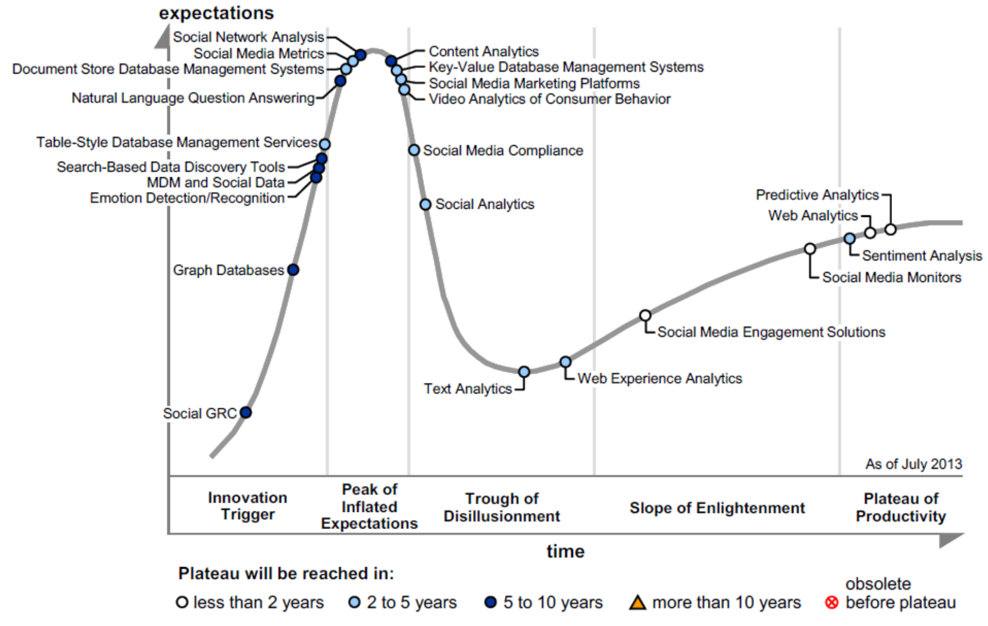
According to Gartner Inc. (Fenn et al 2009), a content analytics application can be a single function, such as keyword extraction. More often, it is a complex function, such as sentiment or trend analysis, fact extraction or reputation analysis. Web 2.0 and business intelligence (BI) is an umbrella of other technologies, it refers to a set of Web 2.0 collaboration and communication technologies such as Really Simple Syndication (RSS), Representational State Transfer (REST), Ajax, blogs and social networking. Web 2.0 and BI provides the ability to synthesize, tag and share information, and deliver analysis more easily, with a greater variety of delivery methods for a greater variety of use cases and deliver a richer, more collaborative experience.

3.3 Market Size

According to Rozwell et Sallam[14], by 2015, more than 30% of analytics projects will deliver insights based on structured and unstructured data. This is even more relevant by considering that by 2015, 85% of Fortune 500 top organizations will require support for big data analysis. The ability to exploit big data is becoming essential for

[14] Rozwell C., Sallam R., Hype Cycle for Content and Social Analytics, 2013, Gartner, 23.07.2013, http://www.gartner.com/document/2556115?ref=QuickSearch&sthkw=hype%20cycle%20for%20social%20analytics (last access 04.04.2014)

companies aimed at implementing the following strategies: risk management, customer engagement, brand management, marketing effectiveness, channel effectiveness, sales effectiveness, organizational effectiveness. In order to identify the best tools for developing Content Analytics or Social Media analysis it is useful to describe the Hype Cycle for Content and Social Analytics 2013 provided by Gartner.

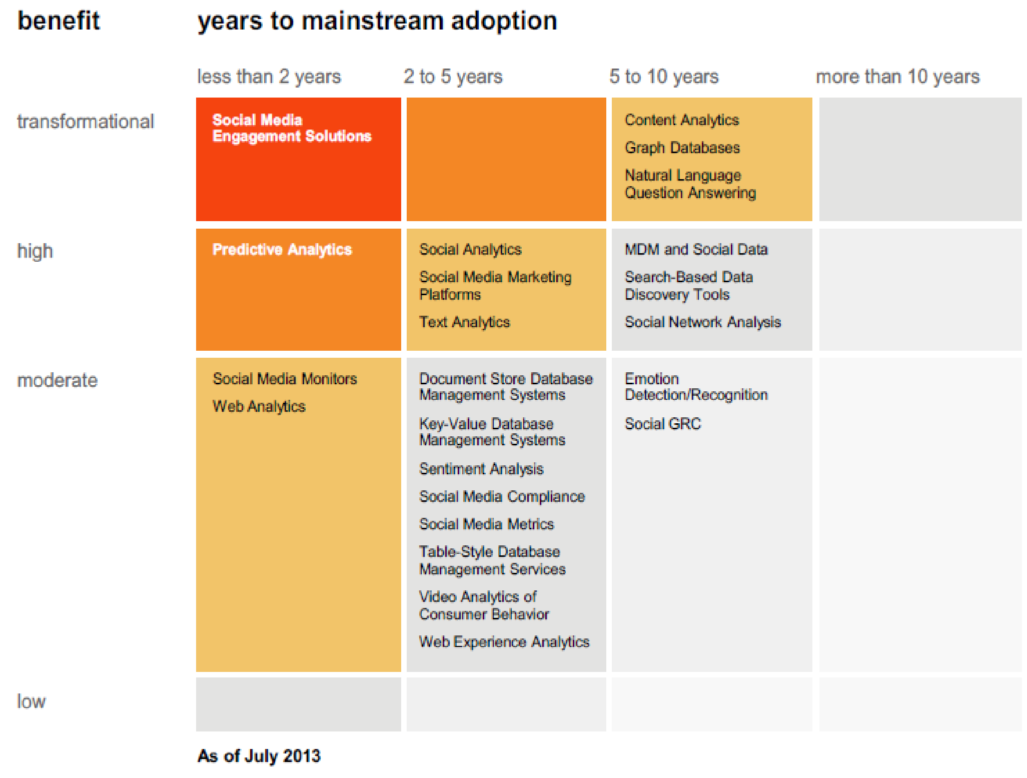


Source: Gartner (July 2013)

Figure 6 Hype Cycle for Content and Social Analytics 2013[15]

The figure above shows that in less than 2 years Social Media Engagement solutions, such as social data analysis and discovery tools, will provide transformational benefits to the companies and society as a whole. More in detail, by taking into account the Priority Matrix, Predictive Analysis will also have an high impact and Social Media Monitoring and Web Analytics will have a moderate impact. From 2 to 5 years no transformational solutions are expected, but Social Analytics, Social Media Marketing Platforms and Text Analytics will provide the highest benefits. It is relevant to identify that in this category are included Sentiment Analysis tools, currently entering the phase of “Plateau of Productivity”, together with Web and Predictive Analytics. More transformational solutions in the next 5 to 10 years will be constituted by Content Analytics, Graph Databases and Natural Language Question Answering, as shown in the figure below. According to Gartner the Content Analytics market is in the “Adolescent” phase and the market penetration is from 5% to 20% of the target audience. This implies that when the TrendMiner project ends, market enlargement will be in process, making commercial introduction feasible and economical and financially sustainable.

[15] Rozwell C., Sallam R. L., Hype Cycle for Content and Social Analytics 2013, Gartner, 2013, available at http://www.gartner.com/document/2556115?ref=QuickSearch&sthkw=hype%20cycle%20for%20content%20and%20social%20analytics%202013



benefit	years to mainstream adoption			
	less than 2 years	2 to 5 years	5 to 10 years	more than 10 years
transformational	Social Media Engagement Solutions		Content Analytics Graph Databases Natural Language Question Answering	
high	Predictive Analytics	Social Analytics Social Media Marketing Platforms Text Analytics	MDM and Social Data Search-Based Data Discovery Tools Social Network Analysis	
moderate	Social Media Monitors Web Analytics	Document Store Database Management Systems Key-Value Database Management Systems Sentiment Analysis Social Media Compliance Social Media Metrics Table-Style Database Management Services Video Analytics of Consumer Behavior Web Experience Analytics	Emotion Detection/Recognition Social GRC	
low				

As of July 2013

Source: Gartner (July 2013)

Figure 7: Content and Social Analytics Priority Matrix 2013[16]

While there is no specific hard data on the market for tools that provide Content Analytics or Social Media monitoring, it is possible to analyse the Business Analytics Software market as a whole, which is formed by three main categories of tools: Business Intelligence, Performance Management and Analytics Applications, Data Warehousing Platform Software.

According to an IDC report[17], in 2012 the worldwide business analytics software market grew 8.7% to reach $34.9 billion. The market is now forecast to grow at a 9.7% compound annual growth rate (CAGR) through 2017. Compared to 15% of 2011, the value is significantly low. By analyzing the detail of the three segments of the Business Analytics Software market, in 2012 the data warehousing platform software segment is the fastest to grow with a value of 10.8%, followed by the business intelligence and analytic tools and performance management and analytic applications segments — both of which grew at 7.7% each. Workforce analytic applications and content analytics markets grew the fastest with 15.8% and 14.9%, respectively.

With respect to the previous analysis provided in Deliverable 8.2.1, also SAS is included in the market as large vendor and in fact the market is dominated by five large vendors (Oracle, SAP, IBM, Microsoft and SAS). The values related to the percentage of growth rate are all decreased except for Microsoft that achieved a

[16] Ibidem

[17] IDC, (2013), Worldwide Business Analytics Software 2013–2017 Forecast and 2012 Vendor Shares, http://idcdocserv.com/241689e_sas (last access 04.04.2013)

growth of 15,3% during 2011-2012 (it was 11.4 % during 2010-2011). These data reflect also the fact that Microsoft is the only company to have produced an increased percentage of market share in 2012 (7.8%), with respect to the other four vendors. Oracle achieved the highest value in terms of revenues within the Worldwide Business Analytics Software market (6,484.6 Million of Dollars).

With reference to the detail of the Worldwide Advanced Analytics Software Market, it is shown that no single solution provider dominates the market for tools providing Content Analytics, or Social Media monitoring. Players range from the largest enterprise software vendors (SAS, IBM, Microsoft) to a stream of new entrants (KXEN, Pitney Bowes Software and other small vendors), both commercializing research technologies and bringing solutions to new markets. This leaves room for applications based on TrendMiner.

IDC’s report provides also predictions on the market for the period 2013-2017. The worldwide CAGR for the market is expected to be 9.7%. However, the growth of individual segments of the market varies from 6.5% for production planning applications to 11.8% for CRM analytic applications.

4 The TrendMiner Infrastructure

4.1 Possible Users of TrendMiner

Three rather distinct groups of potential users of TrendMiner can be anticipated, each with distinct sets of interests and requirements regarding possible use features.

4.1.1 The general Public

This heterogeneous group of users would most likely approach a TrendMiner tool from the same perspective as other, similar tools that are out there on the web[18]. Their main motivation for use would be casual interest. They might take a look at certain topics of interests or explore trends and sentiments visualised by TrendMiner on a daily basis, or just once because they were sent a link by a friend over Twitter and wanted to take a look.

4.1.2 PR or Marketing Professionals

Among this group we might find users such as PR executives, journalists, politicians or party managers etc. Some of its members may be highly active in their professional use of Social Media. This group of users would use a TrendMiner tool as a quick and easy way of exploring trends and hot topics on Twitter or other Social Media because

[18] Examples of such applications are (only to name a few):
https://tool.opiniontracker.net
http://visualization.geblogs.com/visualization/cancerconversation
http://twendz.waggneredstrom.com
http://twittratr.com
http://election.twitter.com
http://twittersentiment.appspot.com/
http://socialmention.com/

it is professionally relevant for them. The TrendMiner consortium is very interested in this group because they constitute a very important segment of both users and potential customers.

4.1.3 Scientific Researchers

This is a group of users who need a TrendMiner tool to be not only a powerful and reliable way of visualising streaming media, but also to be a scientific research tool, such as social researchers at SORA or financial researchers at EK or HFPL. Their use-mode would require high levels of customisation and comprehensive recording functions.

4.2 Basic Functions

In the course of TrendMiner, a set of tools have been developed to help validate the results of the technological development. Some basic functions which areavailable are[19]:

- Keyword search
- Graphical timeline of topics and sentiment
- Other graphical analysis features, e.g. for number of mentions, level of emotionality, associated topics
- Drill-down to raw data
- Options to filter search results by Time and Date, Places and Languages

4.3 Technological Background

During the development process of TrendMiner, a number of technologies are involved in the processing and delivering of data and results. Components which are currently being worked on are described briefly in this section (but more details are available in the Deliverables of WP2, WP3, WP4 and WP5). All components are integrated in a platform (see Deliverable D5.3.1: Architecture for distributed text annotation - v2, Real-Stream Media Processing Platform - v1), which will cover all the phases from the (social) stream processing lifecycle: large scale data collection, multilingual information extraction and entity linking, sentiment extraction, trend detection, summarization and visualisation. In figure 8 below we display the new view on the TrendMiner infrastructure, including the new use cases that we added after the extension of the consortium of the project.

We describe briefly the main components that are integrated now in the platform.

[19] More advanced functionalities, like reasoning and queries for a knowledge base are under implementation, but those will probably made accessible only to domain experts.

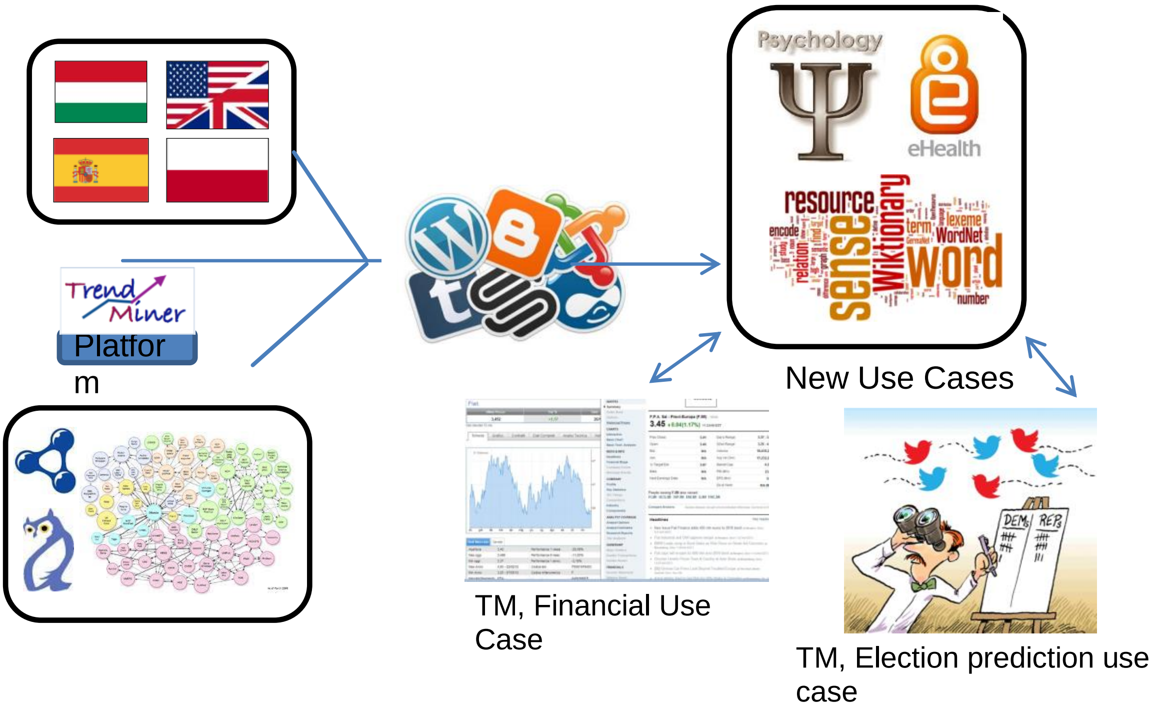


Figure 8 TrendMiner Infrastructure

4.3.1 Opinion and Sentiment Mining

Detection of sentiments and opinions in the two fields covered by the use cases is a central aspect of the project, and TrendMiner is developing and using novel weakly supervised machine learning algorithms for automatic discovery of correlations between textual data and factual information, supporting the detection of trends. Scalability and affordability are addressed through a cloud-based infrastructure for real-time text mining from stream media.

In the field of Human Language Technology (HLT), opinion and sentiment analysis are playing a steadily growing role. This is partly due to the fact, that the detection of such phenomena in an automated way can show that the natural language processing is getting closer to the task of language understanding, adding the capability to detect high-level contextual semantics that is inherent to language, and not directly dependent from the kind of semantics one can access in knowledge repositories. This increasing interest in opinion, sentiment and/or subjectivity detection in natural language expressions is demonstrated by a high number of conferences, workshops, shared tasks and evaluation campaigns, dealing with opinion and sentiment detection and analysis, resulting in a huge quantity of available data sets for training and testing systems[20].

[20] e.g. International AAAI Conference on Weblogs and Social Media (icwsm.org/) provides data sets for training and testing

One feature making opinion and sentiment analysis highly interesting for HLT is the fact that opinion and sentiment detecting is often applied to Social Media, where many different kinds of textual data so far untouched by HLT projects can be found.

4.3.2 Data Collection and Crawler Service

Working on TrendMiner enables project partner IMR to improve their web-scale crawler, the result of which would be of great interest for several institutions who require real time big data (News, RSS), including Social Media (blogs and forums). IMR would adapt its current crawler to capture real time and social data and its current architecture to use and store this data. The challenge is to do this at scale, to enable real trend mining based on a large number of sources (millions of RSS items, forums and blogs with many comments) while using limited amount of resources and storage to make this a viable service platform.

The adaptation includes developing algorithms for identifying relevant media information sources in multiple languages, prioritization, ranking and filtering of data sources (WP5). These features on top of existing features of spam detection, data cleaning and data de-duplication make IMR’s crawler and platform efficient in handling relevant real time and social data and produce high quality repository of data for the consumer to work on. For example concerning blogs and forums, it is important to be able to identify the structure of this kind of website, in order to capture all comments.

4.3.3 Ontology-based Information Extraction

Information Extraction (IE), a form of natural language analysis, is becoming a central technology for bridging the gap between unstructured text and actionable knowledge. Ontology-Based IE (OBIE) is IE which is adapted specifically to the challenge of annotating unstructured content with respect to a formal ontology (a formal conceptualisation). In particular, TrendMiner is addressing the challenge of using large-scale Linked Open Data resources, which contain millions of instances, as well as formal classes and relationships between them.

The automatic semantic annotation of Social Media streams enables the semantic-based summarisation, search, browsing, and visual analytics techniques also developed in TrendMiner as part of WP4. Such knowledge is also needed for building semantic models of the user, their social network, and online behaviour. TrendMiner’s LOD-based IE technology is thus relevant in many application contexts, going beyond the use cases pursued in the project, e.g., knowledge management, competitor intelligence, customer relation management, eBusiness, eScience, eHealth, and eGovernment.

4.3.4 Reasoning over Streams

Engaging actively with high-value, high-volume, and dynamic Social Media streams has now become a daily challenge for both organisations and ordinary people. TrendMiner is developing intelligent Social Media summarisation methods that can automate, at least partially, this process. State-of-the-art automatic text summarisation algorithms have been developed primarily on news articles and other carefully written, long documents. In contrast, user generated content tends to be very different: often short, strongly grounded in context, temporal, noisy, and full of slang.

To make use of large collections of data, a key requirement is the ability to query the data and sometimes to reason over the data. Though often presented separately, in a very broad sense both querying and reasoning can be thought of as a filtering activity over data and a combination of this filtered data to extract or recombine to produce novel data. The structure of such filter/construction activities are defined in a query language in the querying context or as rules in a reasoning context. Many scientific and commercial activities can produce high throughput streams of structured data. Such activities may include sensor network activity monitoring, Social Media streams, telecom call recording, financial transactions etc. The data produced by these activities is often structured in semantic web data technologies such as RDF. It is often desirable for such data to be queried and reasoned over in real-time, however doing so presents many interesting challenges. Firstly, treating the stream as a standard static Data Base Managment System (DBMS) can prove difficult. One strategy might be to fill a traditional static DBMS with time-separated windows of the stream, though in doing so questions arise as to what happens to the results of a query or entailments of reasoning rules when the stream (by definition) generates more data over time. It is clear that Data Stream Managment Systems (DSMS) represent a separate problem space and over the last decade DSMSes have generated a great deal of interesting research activities. Southampton is developing a framework which leverages the well-studied Rete reasoning algorithm and implements it over a distributable streaming processing framework called Storm. This ReteStorm provides a principled means by which to compile both reasoning rules (e.g. Jena Rules orother rule languages) and queries (e.g. SPARQL, C-SPARQL) as distributed reasoning networks which can be applied to streams. Our implementation supports distributable filters and joins which means our stream reasoning techniques is built to scale with extremely high yield streams, a common problem in many modern streaming data sources. Further, we use a “sliding window” technique wherein joins forget data after certain time period, after some memory limit is reached or some other criteria. This allows ReteStorm to efficiently match queries and produce entailments over a potentially infinite stream.

4.3.5 An Open Source, Social Media Text Processor

With the rise of large online social networks (OSN) there has been an equal rise in the desire to analyse, process and understand activities across these networks. This can be a quite challenging task especially considering the unique issues presented by OSN data including: short messages with little structure; inconsistent spelling, grammar and capitalisation; threaded discussions with many participants across large graphs and so on. Standard text processing tools have been shown to have great difficulty in dealing with such data and therefore many authors have explored algorithms and techniques which purport to handle the unique challenges set by such data. However, these approaches have often been adhoc, implemented in a piecemeal, closed-source fashion usually to address an individual problem and, amongst their other issues; do not address the scale of the data available in modern OSN. To address this gap, Southampton and Sheffield have developed the TrendMiner preprocessing Tool. The tool contains implementations of various components expected in a low level text pre-processing pipeline specifically geared to deal with data from OSN. The techniques provided include tokenisation, language detection, stemming, Pointwise mutual information and location detection with future plans to provide open implementations of sentiment analysis, POS tagging and Name Entity Recognition. Furthermore the tool provides support for multiple social network formats through support of USMF1, filtering functionality and varying output formats. The tool is engineered in a modular manner, allowing for easy extension and addition of novel techniques. Furthermore, the pipeline of the tool is specifically designed to work either as a library, a single machine command line tool or even to analyse massive datasets with a MapReduce enabled implementation of the tool. DFKI is developing tools for the automatic extraction of polarity lexicons out of newswires, and grammars for computing the polarity of phrasal and sentential units. Addtional to this, DFKI is implementing domain ontologies, a biography ontology and an opinion/sentiment ontology. The combination of those ontologies allows storing consolidated information about opinionated entities, resulting from information extraction, stream reasoning or summarization.

5 Existing Content Providers/Brokers

In the framework of the massive growth due to volume and velocity of online media and online social network (OSN), TrendMiner project focuses on improving the way to access and interpret this strongly multilingual content, in a timely efficient, and affordable manner.

Indeed, as the size of collaborative and/or conversational platforms is exploding, this content is attracting a growing interest for many industries, companies and SMEs. All OSN or Web content in general are not organized on the same business model and that implies different strategies in terms of accessing its content.

So before the content analytics market watch (part 6), a market watch focused on companies that are specialized in aggregating and providing online media and content from OSN brings significant information.

In terms of accessing OSN content such as Twitter, Tumblr or Facebook few players operate. In fact, it is Twitter, as one of the leaders of OSN that paves the way towards an organized value chain. By putting in place its Firehose, Twitter decided to control the access to their user-generated content and take the most to build a business model on it. These OSN are difficult if not impossible to crawl and collect (robots.txt, blacklist, technological barriers) and require using APIs, which imply strong limits in terms of requests and volume of content. This Firehose is directly sold to Content brokers such as DataSift or Gnip, that will resell it partially or totally to other technological and/or analytic companies.

DataSift or GNIP offers (Figures bellow) are an indicator of how the sector is starting to be structured. These major OSN content resellers put in place agreements and a revenue share model for the majority with data sources. Financial terms are not public, but are quite similar to the relationship built with Twitter (License + $/1000 interactions). Via this agreement, and after a repackaging of raw streams provided by these sources, DataSift and GNIP customer’s access to posts, comments, status and all user information (author’s name, number of followers, age, gender, locality...).

It is also interesting to note that Twitter has just bought GNIP[21]. It will certainly bring a new role of packaging and selling data and it remains unknown how this GNIP deal will affect Twitter’s relationship with other Twitter data provider.

DataSift Offer	Interactions (/day)	Users	Real Time (/interaction)	Historics	Cost (/1000)
Twitter	500+ million	200+ million	200ms	3+ Years of data	$0.10
Facebook	50+ million	1.2 billion monthly	N.P.	From May 2012	Free
Sina Weibo*	100+ million posts	500+ million	N.P.	N.P.	N.P.
WordPress[22]*	2-5 million	400+ million visitors /month	200ms	From Dec. 2013	N.P.
Intense Debate*	250+ thousand	N.P.	200ms	From Dec. 2013	N.P.

[21] The Verge, 15/04/2014, *Twitter buys social data specialist Gnip after years of working together*

[22] DataSift, 5/12/2013, *DataSift Extends Market Leadership With Automattic/WordPress Global Partnership*

Tumblr[23]*	150+ million	300+ million	200ms	Since August 2013	$0.20
LexisNexis*	400,000+ items	N.P.	N.P.	Coming soon	N.P.
Google +	N.P.	300+ million monthly	N.P.	N.P.	N.P.
YouTube	2+ million	N.P.	N.P.	N.P.	$4
bitly*	N.P.	N.P.	200ms	From Nov. 2012	N.P.
Instagram	N.P.	N.P.	N.P.	N.P.	N.P.
NewsCred*	100+ thousand	N.P.	N.P.	N.P.	N.P.
Reddit	200+ thousand	120'000+	N.P.	N.P.	Free
Wikipedia	120+ thousand	N.P.	N.P.	N.P.	Free
DailyMotion	30+ thousand	N.P.	N.P.	N.P.	$4
Topix	80+ thousand	N.P.	N.P.	N.P.	$4
IMDb	20+ thousand	N.P.	N.P.	N.P.	$4
Videos	200+ thousand	N.P.	N.P.	N.P.	$4
Blogs	120+ million	N.P.	N.P.	N.P.	$4
MessageBoards	30+ million	N.P.	N.P.	N.P.	$4
2channel	N.P.	N.P.	N.P.	N.P.	$4

* DataSift has extended in the last few months its services by offering access to these channels.

Table 2 DataSift offer

GNIP Offer	Type of Access
Twitter	Complete
Tumblr	Complete
foursquare	Complete
WordPress	Complete
Disqus	Complete
StockTwits	Complete
intensedebate	Complete
estimize	Complete
sitrion	Complete
bitly	API Access
dailymotion	API Access
del.icio.us	API Access
facebook	API Access
flickr	API Access
google+	API Access
Instagram	API Access
metacafe	API Access
panoramio	API Access
photobucket	API Access
plurk	API Access

[23] TechCrunch, 16/09/2013, *Tumblr Inks Firehose Deal With DataSift To Tumble Further Into The World Of*

reddit	API Access
stackoverflow	API Access
vimeo	API Access
VK	API Access
YouTube	API Access

Table 3 GNIP offer

MediaSift/DataSift[24]

DataSift is operated by MediaSift Limited, registered in England and Wales (United Kingdom). DataSift is one of the leading social data platforms, enabling companies to aggregate, filter and extract insights from the billions of public social conversations on Twitter, leading social networks and millions of other sources. DataSift provides access to both real-time and historical social data to uncover insights and trends that relate to brands, businesses, financial markets, news and public opinion. Delivered as a cloud platform, DataSift does the heavy lifting for companies creating social media monitoring, social CRM, business intelligence, financial trading and news monitoring applications. DataSift is a certified Twitter data reseller partner.

MediaSift / DataSift	
Pricing strategies	monthly fee + cost per use
Sources	See Figure Data sources provided by DataSift
Strengths of the products/services	The platform provides a lot of sources and access to historical data. DataSift has created its own programming language that can be used to create complex data filters and to augment the data. DataSift also offers advanced social data enrichments and utilizes machine learning, which allows social data to be automatically categorized based on meaning and context
Weaknesses of the products/services	The use is complex and pipes are long to put in place. The current uncertainty of the future services based on Twitter is a real weakness

Table 4 MediaSift / DataSift

GNIP[25]

Gnip is headquartered in Boulder, Colo., and has offices in San Francisco, New York and Washington D.C.
Gnip is one of the world's largest provider of social data. Gnip was the first to partner with Twitter to make their social data available, and since then have been the first to work with Tumblr, Foursquare, WordPress, Disqus, StockTwits and more. In April 2014, Gnip has been acquired by Twitter.

[24] http://datasift.com
[25] http://gnip.com

Gnip delivers social data to customers in more than 40 countries, and Gnip's customers deliver social media analytics to more than 95% of the Fortune 500.

Gnip	
Pricing strategies	monthly fee + cost per use
Sources	See Figure Data sources provided by GNIP
Strengths of the products/services	The platform provides many sources and some of them are exclusive. Access to historical data. Gnip offers Data Collector, a turn-key solution that provides users a way to collect social data from up to six different social APIs using a single connection (the Gnip API). Data Collector also removes duplicate data and normalizes the format across all APIs. Moreover the acquisition by Twitter gives a more long-term vision to the access to Twitter source
Weaknesses of the products/services	Current uncertainty on future sources other than Twitter. Pricing are not displayed on the website, so customers need to contact them and it delay the first use. No online free trial

Table 5 Gnip

Other companies are also specialized in aggregating and providing various types of sources, such as online news, corporate webpages, or other websites.

Trendiction/Talkwalker[26]

Talkwalker is developed by Trendiction, a privately held company founded in 2009 and based in Luxembourg.

Talkwalker is a social media monitoring and analytics tool. It delivers insights in a user-friendly dashboard recommended by agencies and brands worldwide. With a focus on big data crawling the Talkwalker search index is one of the largest, covering over 150 million sources in 187 languages and 247 countries. It crawls millions of social media but also online news sources. With 12 millions of documents indexed daily, this Luxembourg company provides three main services: for media monitoring companies, for PR agencies and another for market research companies.

Talkwalker was rated one of the top 5 best social media monitoring tools and online reputation management tools globally in 2013.

Trendiction / Talkwalker	
Pricing strategies	monthly fees
Sources	Social media and News
Strengths of the products/services	Frequent updates of all current data sources. Data

[26] http://www.trendiction.com, http://www.talkwalker.com

	provides more European sources than American providers
Weaknesses of the products/services	The scope of sources is quite static, without any regular discovery of new sources. The price, which is high for start-ups

Table 6 Trendiction / Talkwalker

Moreover Technologies[27]

Moreover Technologies, Inc., an US company founded in 1998. Through US (Chicago, Dayton, Fort Lauderdale, New York, and San Francisco) and UK (London) offices, the firm offers corporate customers worldwide direct access to comprehensive, yet targeted, real-time business and consumer information from the Web’s most read and respected sources. Daily, Moreover Technologies offers unified portal access to 2.5 million news articles and social media posts from 3.0 million editorially vetted sources across 100+ countries, 75+ languages and 800+ searchable industry categories. It crawls and works with publishers and licensing organizations to provide various services of alerts on RSS news feeds, customized sources addition and incorporation of internal and licensed content for various sector as marketing and financial.

Moreover Technologies	
Pricing strategies	monthly fees
Sources	Onilne News from RSS, sources from publishers (with licensing model), customer sources
Strengths of the products/services	Services on the top of sources delivery. Good coverage on international and American sources
Weaknesses of the products/services	Weak coverage of European sources

Table 7 Moreover Technologies

[27] http://moreover.com

6 Existing Content Analytics Providers

6.1 General Introduction

Content analytics and Social Media monitoring are growing branches of software services, there is a large variety of providers. All of them have in common, that they tap the huge mass of content and data available on the web through sources like Social Media, online newspapers, stock exchanges, online forums and the like. Another parallel is their general purpose: to aggregate, structure and visualise the constant data flows. Using different ways of analysis and summarisation, including natural language technologies, they produce information on a scale that a human user can manage. They differ however, in the exact use to which they are put, in the details of the sources they access and the analytical options they offer.

We now present selected examples of important providers of financial and political content analytics and their products. In the tables we identify also pricing strategies (where available online), data sources, strengths and weaknesses of the products/services with respect to TrendMiner.

6.2 Financial providers

RavenPack[28]

RavenPack is the leading provider of real-time news analysis services. The clients include some of the best performing quantitative and algorithmic trading firms in the world. RavenPack’s products are aimed at analysing news through linguistic methodologies. RavenPack's analysis is based on hundreds of thousands of stories per day coming from several sources in different formats in milliseconds, RavenPack's news analysis is also able to identify short-term trends.

The RavenPack’s product for Content analysis is named News Analytics that automatically monitors and analyses information on over 100 countries and governments, more than 140,000 key geographical locations and 30,000 companies. RavenPack Analysis also collect more than 1,000 types of events related to corporate actions, terrorist threats and natural disasters. The news related to events are categorised and the sentiment analysis is based on different metadata divided in terms of place, organisation, company, currency or commodity. RavenPack News Analysis service is in partnership with Dowjones.

Dow Jones

Dow Jones & Company is a big company of journalism and smart technology. Dow Jones owns several newspapers, newswires, websites, apps, newsletters, magazines, proprietary databases, conferences and more. Dow Jones’ premier brands include: The Wall Street Journal, Dow Jones Newswires, Factiva, Barron's, MarketWatch, SmartMoney and All Things D. The Dow Jones Local Media Group publishes community newspapers, websites and other products in six U.S. states.

[28] http://www.ravenpack.com/products/ravenpack-news-analytics/

Dow Jones' products are mainly focused on testing short- and longer-term algorithmic trading models by analysing news and data archives. Dow Jones News & Archives For Algorithmic applications: this is one of the two Dow Jones' products for news analysis and it is aimed at mining news and data from Dow Jones' 30 years archive through algorithmic trading models. This product covers also the sentiment analysis. Dow Jones News Analytics: this product analyses real-time and historical news and data for sentiment through trading models and it is able to analyse the relevance, volume, novelty and other market signals via a range of technology options. As mentioned above, the Dow Jones applications are apparently based on technology developed by RavenPack, and are therefore most likely very similar in makeup and capabilities. Both Providers seem to be focused on analysing classical news streams rather than data streams from Social Media, making them significantly different from TrendMiner in terms of data source.

RavenPack & DowJones	
Pricing strategies	-
Sources	Online news
Strengths of the products/services	It is used to enhance returns or improve efficiency by quantitative & algorithmic traders, portfolio managers and surveillance analysts RavenPack detects news and produces analytics data on over 31,000 listed stocks from the world's equity markets DowJones provides breaking and exclusive news and context to fill the void between the nationals and narrowly focused specialist trade press
Weaknessess of the products/services	Both Providers seem to be focused on analysing classical news streams rather than data streams from Social Media

Table 8 RavenPack & DowJones

Thomson Reuters[29]

"Thomson Reuters is the world's leading source of intelligent information for businesses and professionals. It combines industry expertise with innovative technology to deliver critical information to leading decision makers in the financial, legal, tax and accounting, healthcare, science and media markets, powered by the world's most trusted news organization". The Sentiment Analysis product by Thomson Reuters is named News Analytics (TRNA) for Internet News and Social Media and is a tool aimed at analysing millions of public and premium sources of internet content, tag and filter. The tool then turns the data into actionable analytics in real time to support trading, investment and risk management decisions. The TRNA

[29] http://thomsonreuters.com/products/financial-risk/01_255/News_Analytics_-_Product_Brochure-_Oct_2010_1_.pdf

engine is mainly deployed by trading firms to analyse Thomson Reuters' News and a host of professional news wire services. TRNA for Internet News and Social Media product aggregates content from more than four million Social Media channels and 50,000 Internet news sites. The TRNA engine is also able to analyse not only sentiment, but also relevance and novelty. Thomson Reuters and its News Analytics product seems to be the most similar provider to the TrendMiner objectives. It analyses, in real-time, more than 50000 aggregated news and more of 4 million Social Media channels. The main difference is in the business model behind the Thomson Reuters services and the aim of the TrendMiner project. One difference lies on the outputs of these services: Thomson Reuter generates an output of quantifiable data points across a number of dimensions, but it does not offer directly forecast on the market through proprietary tools; TrendMiner aims to provide sentiment and, at the same time, a forecast on the markets through a financial model ad hoc developed. Another difference regards the provision of the main functionalities of TrendMiner for free to a public audience, and while Thomson Reuters is an expensive service and its completely provided in order to make profits. In a recent development, Thomson Reuters and Twitter are cooperating on sentiment analysis in financial markets30. Thomson Reuters developed also OpenCalais[31] a toolkit to automatically incorporate state-of-the-art semantic functionalities within Social Media, as blogs, content management systems, websites or application. It does not provide sentiment analysis functionalities, but aggregates, makes connection and create new context of data with online sources. This is relevant for the social media landscape.

Thomson Reuters	
Pricing strategies	-
Sources	Internet news, social media
Strengths of the products/services	The tool turns the data into actionable analytics in real time to support trading, investment and risk management decisions Thomson Reuters news analytics (TRNA) scores news items on over 25000 equities and nearly 40 commodities and energy topics The TRNA engine is also able to analyse not only sentiment, but also relevance and novelty of more than 50000 aggregated news. Close cooperation with Twitter
Weaknessess of the products/services	It does not offer directly forecast on the market through proprietary tools It is an expensive service and its completely provided in order to make profits

Table 9 Thomson Reuters

[30] http://techcrunch.com/2014/02/03/twitter-raises-its-enterprise-cred-with-thomson-reuters-sentiment-analysis-deal/

[31] http://www.opencalais.com/

FINIF[32]

FINIF examines various sources of real time data (e.g., SEC filings, news headlines, and tweets) and use textual analysis to measure the sentiment of the text or flag certain sensitive phrases sensitive phrases in order to measure the information environment for a firm. The algorithms developed by FINIF make sense of the thicket of SEC filings, news articles and thousands of tweets every minute using textual analysis to identify investor sentiment.

The product evaluates the number of positive and negative news of a specific company compared to the stock price. The FINIF product is mainly based on the analysis of Tweets and information flows of a specific company. FINIF product creates also a sentiment score by aggregating the sentiment measures across a sample of the most recent Tweets.

FINIF	
Pricing strategies	-
Sources	Twitter, online news
Strengths of the products/services	FINIF product creates a sentiment score by aggregating the sentiment measures across a sample of the most recent Tweets It shows the stocks that have the most positive or negative sentiment The product evaluates the number of positive and negative news of a specific company compared to the stock price FINIF financial sentiment analysis scan for new 10-K and 10-Q20 filings and then apply the sentiment algorithms to create investment's reports
Weaknessess of the products/services	FINIF does not use many social media It lacks multilingual capabilities and it is not suited to provide outputs that would facilitate forecasting based on available data, as instead TrendMiner does

Table 10 FINIF

SNTMNT[33]

SNTMNT is a financial startup based in Amsterdam and focused on sentiment analysis. It was founded in 2011. SNTMNT work is mainly based on academic researches. SNTMNT helps investors to evaluate and monitor their portfolio, as well

[32] http://www.finif.com/
[33] http://www.sntmnt.com/

as to develop profitable trade ideas. It provides solutions to financial institutions like brokers, banks and asset managers. SNTMNT captures news from Twitter and Facebook, then it processes these messages. The sentiment index is analysed using their own financial sentiment algorithms that has been trained to recognize specific languages by market professionals. Their trading API indicator is an off-the-shelf data API providing a data feed consisting of the social sentiment for individual stocks. SNTMNT processes big data, analyses messages using financial sentiment algorithms and provides to customers insights about the sentiment.

SNTMNT	
Pricing strategies	-
Sources	Facebook, Twitter
Strengths of the products/services	They calculate a sentiment index, an average financial sentiment and a financial sentiment momentum, every 24 hours, 6 hours, 3 hours and 1 hour SNTMNT provides 14 sentiment parameters SNTMNT provides a web-based platform It is available for a large selection of U.S. stocks Their trading indicator API provides hourly updates
Weaknessess of the products/services	SNTMNT does not develop an app for mobile devices SNTMNT does not use financial news websites

Table 11 SNTMNT

FINSENTS[34]

FINSENTS statistical and semantic engine scans thousands of financial sources to capture the mood of the market and to pull back the content for the investors' needs. Their application is based on web analysis and on other unstructured data, it provides instantaneous insights on stocks, Forex and commodities. FINTSENTS indexes web or private content in a way similar to Google, Bing or Yahoo. FINSENTS provides four types of subscriptions: one for free and the others based on fees. The free subscription does not allow to use the sentiment index, but provides only buzz and a sentiment overview. If the underwriters would like to use a sentiment index for a wider list of equities the cost of subscriptions is expensive (for more information please look at the pricing strategies presented in the table below). FINSENTS has developed also an app for apple devices.

[34] http://www.finsents.com/

FINSENTS	
Pricing strategies	"Free" - this type of subscription provides only buzz and a sentiment overview "Investor" 50$/m – this type of subscriptions presents a sentiment index for major global equities "ProTrader" 200$/m - this type of subscriptions gives a sentiment index for equities, forex and commodities "Corporate" - this subscriptions gives plug-in of private sources
Sources	Thousands of Financial sources
Strengths of the products/services	Sentiment data are correlated to asset price movements to ensure consistency It provides instantaneous insights on stocks, FX, commodities. They have an app for apple devices. FINSENTS has developed also an app for apple devices
Weaknessess of the products/services	If the underwriters would like to receive a sentiment index for more equities the costs of subscriptions are expensive

Table 12 FINSENTS

H2O consulting[35]

H2O Consulting is a risk analytics firm operating in Lugano (Switzerland) and specialised in Quantitative Finance, Risk Management, Behavioral Finance and Sentiment Analysis solutions. It processes millions of social media post messages every day, it monitors social media channels and analyses the overall sentiment through the use of their own financial algorithms. It classifies post messages as strong buy, buy, hold, sell and strong sell on the basis of their sentiment tag. H2O Consulting has created an app for mobile devices, while the purpose of TrendMiner is also to provide a web-based platform. H2O Sentiment App provides financial sentiment analysis for investors to discover, react and respond to market opinions. The app is built on the 30 Dow Jones Index (DJIA) stocks. The free app provides the financial sentiment analysis of one stock (Microsoft).

H2OConsulting	
Pricing strategies	Free app for mobile device - it gives a sentiment index for one stock of Dow Jones

[35] http://www.h2oconsulting.ch/

	1,79 € app for mobile device - it gives a sentiment index for all the stocks of Dow Jones
Sources	Social Media
Strengths of the products/services	The sentiment index and real time charts are updated every 15 minutes Their research has demonstrated that the trading strategies based on sentiment analysis is better than investment strategies (like Buy and Hold)
Weaknessess of the products/services	H2O consulting has not a web-based platform H2O consulting does not use financial news websites

Table 13 H2O Consulting

SENTDEX[36]

SENTDEX is a company focused on big data analytics, natural language processing and sentiment analysis. It pulls news from over 20 major news websites (stocks, politics, general news) and from smaller news websites. It also processes several million tweets a day from Twitter and multiple terabytes of text every day. It calculates one sentiment index for market stocks and one for political topics. They provides a sentiment index for UK stocks and for forex. About political topics, the SENTDEX sentiment analysis program first crawls the internet for any information on political topics. Once a political article is located, it is read and then assessed for extracting the political sentiment. Sentiment here means the general mood or feeling expressed within a body of text. SENTDEX provides several types of subscriptions for sentiment API which are not expensive.

SENTDEX	
Pricing strategies	1-250 API calls/day (up to ~125 typical articles a day and 7,500 calls a month) = $9.99/mo 251-500 API calls/day (up to ~250 typical articles a day and 15,000 calls a month) = $19.99/mo 501-1,000 API calls/day (up to ~500 typical articles a day and ~30,000 calls a month) = $29.99/mo 1,001-5,000 API calls/day (up to ~2,500 typical articles a day and ~150,000 calls a month) = $39.99/mo 5,000-10,000 API calls/day (up to ~5,000 typical

[36] http://sentdex.com/

	articles a day and ~300,000 calls a month) = $49.99/mo
Sources	From twitter and from over 20 major news websites
Strengths of the products/services	It calculates one sentiment index for market stocks and one for political topics It provides a sentiment index for UK stocks and for forex
Weaknessess of the products/services	Sentdex has not developed an app for mobile devices

Table 14 SENTDEX

Ability Factors[37]

Ability Factors Pte. Ltd. is a Singapore-based company founded in 2012 providing consulting services for small and medium financial firms. They manage outsourcing partners for major financial institutions, as well as provide strategic outsourcing services for global, regional and local firms. They primarily conduct business in Japan, Singapore, Indonesia and Malaysia. They have developed the Ability Factors Sentiment Index (based on tweets). The sentiment index is calculated on two steps. First of all, they estimates the strength of positive and negative sentiment for each word of every twitter message. The sentiment score is then calculated and stored. To generate live trading signals they track the index and create the following 3 derived indicators:

1) Mean – rolling average of the index
2) Upper Band – the mean indicator plus a dynamic number of standard deviations.
3) Lower Band – the mean indicator minus a dynamic number of standard deviations.

The index generates buy & sell signals daily for global currencies, as well as commodities and equity indices. These predictions take away extra noise in asset pricing and provides investors an additional trading indicator on top of fundamental analysis and/or technical analysis. TrendMiner uses a larger number of social media.

Ability Factors	
Pricing strategies	-
Sources	Twitter
Strengths of the products/services	To generate live trading signals they track the index and create the following 3 derived

[37] http://www.abilityfactors.com/

	indicators (mean, upper band and lower band) The index generates buy & sell signals daily for global currencies, as well as commodities and equity indices
Weaknessess of the products/services	Ability Factors uses only twitter as social media

Table 15 Ability Factors

SEMLAB[38]

SemLab was founded in September 2000. With a strong background in knowledge management and business development, they have created a European based company with developers of semantic software applications. Their main expertise is in Natural Language Processing, Computational Linguistics and Artificial Intelligence technologies. NewsSentiment is provided as a free beta showcase website connected to ViewerPro, a semantic processing platform. It is connected to the public web data of 57 leading finance sources and uses their default finance ontologies to capture the semantics of the breaking news. The website further displays a list of the events, the timing at which they occurred and a graph with the average market sentiment over the last few hours. The resulting finance events are ranked by Semlab according to their implicit sentiment on a scale from -5 (negative) to 5 (positive). TrendMiner also collects news from Social Media.

SEMLAB	
Pricing strategies	1,79 € app for iPad
Sources	57 leading finance sources
Strengths of the products/services	NewSentiment provides a graphical display, intended to provide direct insight of the current sentiment of major European equities
Weaknessess of the products/services	Semlab does not use social media

Table 16 SEMLAB

6.3 Non-financial social media monitoring providers

TAME[39]

Tame is operated by Tazaldoo, a Berlin based company that provides a freemium-model "context search engine" to help users find relevant information thus "taming"

[38] http://www.semlab.nl/

[39] http://tame.it/

the information overload on Twitter. Its product is an online tool that requires the user to log on with their Twitter account. It then provides ranked overviews of the mentioned content of a user´s timeline and the lists they follow, divided into "Links", "#Topics" (i.e. hashtags) and "People" (i.e. Twitter user accounts). A global search tool that reaches beyond individual accounts is also available, but shows only limited results in the freemium version. Additional functions available for all accounts include relevant hashtag suggestions for the user´s own tweets, regular digests and the possibility to share the results of an analysis. Paid accounts can connect multiple Twitter accounts, receive follower analyes, and can store saved search data. The results are very nearly real-time. The timescale for which data can be displayed ranges from 1h in the past to up to 7 days in the past, with fixed intervals in between. In practice, the timescale is limited to what is available from the Twitter API.

TAME	
Pricing strategies	Free Timeline analysis, Daily/Weekly digests Pro Features as above. Additionally included: Lists analysis, Global search, Multiple Twitter accounts (49€ /mo) Business Features as above. Additionally included: Follower analysis (99€ /mo) Premium Features as above. Additionally included: Tame API, Saved Searches, Website Widgets (pricing available upon request)
Sources	Twitter
Strengths of the products/services	Real-time information on relevant content (news articles, hashtags, users)
Weaknessess of the products/services	Limited to Twitter only No entity recognition No graphical display of past frequencies of entities No automated summaries No automated assistance in spotting separate trends

Table 17 TAME

In the course of our research on existing provides of content analysis for social media, the consortium has noted an absence of large-scale dedicated providers of content analysis and monitoring appliations for the political domain. Among the above mentioned providers, Datenwerk, Tame, Crimson Hexagon, Brandwatch and Attensity all provide the possibility of collecting and analysing political content, yet none list this as a specific focus of their software platform. On the other hand, close attention to the developing landscape of social media analysis during last two years has revealed a tendency for small websites to appear during political election phases. Below we list some noteworthy examples.

CRIMSON HEXAGON[40]

Crimson Hexagon is a Boston-based company, operating offices in the US, UK, Canada, France, Japan and Australia. It provides a "Social Media Intelligence" service through its ForSight^{TM} plattform. It provides Sentiment Analysis, Buzz detection, and detection of conversiation Topics on Social Media plattforms (including Twitter, Facebook, Youtube, Sina Weibo, an unspecified variety of Blogs, News, Comments, etc) from 230 counties in 82 languages. Possible data visualisation are varied and allow for the retrospective analysis of trends over time. Possible use cases cover nearly every aspect of public relations and marketing and the application is aimed at clients from nearly every large industry, including political applications. In terms of sentiment analysis, the proprietary Brightview^{TM} algorithm supposedly even detected sarcasm in the case of the Australian election campaign in 2013.

CRIMSON HEXAGON	
Pricing strategies	Pricing available upon request
Sources	Twitter (full firehose), Facebook, Sina Weibo, Salesforce Ideas, Disqus, Bazaarvoice, Blogs, Forums, Consumer reviews, YouTube, News, Online comments
Strengths of the products/services	Full real-time service suite, including trend detection and sentiment analysis Versatile, customised to individual client´s needs Access to a large spectrum of data over many countries and languages
Weaknessess of the products/services	No access to financial markets data Broad focus may mean that performance in the domain of politics does reach its full potential

Table 18 CRIMSON HEXAGON

BRANDWATCH[41]

[40] http://www.crimsonhexagon.com/

[41] http://www.brandwatch.com/

Brandwatch is based in the UK, the US and Germany and offers 2 kinds of services: Brandwatch Analytics is geared towards in-depth analytical functions, while Brandwatch Vizia is a data visualisation for KPI dashboards. Analysis and visualisation functions are described as extremely customisable. Additional functions include human verified sentiment analysis, topic detection, email alerts, data export and tools to support engagement within online discourse.

Brandwatch	
Pricing strategies	Brandwatch/Pro: Unlimited users, projects and dashboards, Automation, including data tagging, Workflows & sharing, 27 Languages, Initial training, Support Portal access, Dedicated Account Manager via eMail, 1 month Historical data, 12 months Data storage (500 £/mo) Enterprise/M: Features as above. Additionally included: dedicated account manager via Telephone, and longer periods for historical data analysis and data storage (2000 £/mo) Enterprise/Q: Features as above, Additionally included: Large scale service for Enterprises with high volumes of queries (2000 £/mo)
Sources	Over 70 million sources (including Twitter firehose, Facebook, Blogs, Video pages, News, Reviews,...) in 25 languages
Strengths of the products/services	Versatile, customisable analytics and visualsation Features human verified sentiment detection Many data sources
Weaknessess of the products/services	No access to financial markets data No automated entity detection Broad focus may mean that performance in the domain of politics does reach its full potential

Table 19 BRANDWATCH

ATTENSITY[42]

Attensity is a Business Intelligence Tool that is developed by a company of the same name, based in the US, Germany, UK and Belgium. Their tool is focused at providing an overview of the Discussions relevant to a company in different channels

[42] http://www.attensity.com/home/

(including, but not limited to, Social Media), generating analyses and facilitating engangement. Their "Attensity Pipeline" collects real-time data from more than 150 million social media and online sources including the full Twitter Firehose, public Facebook and Google Plus posts, YouTube, Reddit, blogs, forums, and video and review sites. Attensity describes the high added-value of their data (it includes annotation of entities, relationships, events and categories as well as sentiment analysis, data augmentation with geographical, demographic and influencer metrics as well as spam filtering) as one of their key products. Attensity "Analyze" delivers real-time analysis and report generation, based on the data provided by "Pipeline". Attensity´s third product "Attensity Respond", then gives the client access to services geared toward CRM activities and client engagement.

Attensity	
Pricing strategies	Pricing available upon request
Sources	150 million social media and online sources including the full Twitter Firehose, public Facebook and Google Plus posts, YouTube, Reddit, blogs, forums, and video and review sites in 35 languages
Strengths of the products/services	Full real-time service, with different products that cover both API requirements for third-party products, as well as individual analysts or PR experts Access to a large spectrum of data over many countries and languages
Weaknessess of the products/services	Geared mainly toward market research, no apparent focus on financial data or political discourse

Table 20 Attensity

Topsy Labs Inc.[43]

Maintained by "Topsy", a San Francisco based company, partnered with Twitter, the "Twitter Political Index" was a dedicated Twitter application, used during the US Presidential Elections of 2012. It incorporated a basic sentiment index, and aggregation functions for the most frequently or hotly discussed topics during the election campaign, such as "Gun Control".

Topsy itself maintains a large index of Tweets, dating back to Twitter´s inception in 2006 and offers free access to search for Tweets, Links, Photos, Videos and Influencers as well basic trend detection. A paid "Pro" version, intended for Government use to "facilitate disaster response, quantify political issues, detect disease outbreak and monitor global anomalies"[44] is available. As of April 2014, no

[43] https://election.twitter.com/

[44] http://www.marketwired.com/press-release/topsy-introduces-topsy-pro-analytics-for-the-public-sector-1699405.htm

additional information on features and pricing could be gleaned from Topsy´s website, as all respective links were redirected to the Topsy search bar homepage. In 2013, Topsy was acquired by Apple[45], which indicates.

Topsy Labs Inc.	
Pricing strategies	Not available
Sources	Twitter, Google+
Strengths of the products/services	Full real-time service, API, Full access to Twitter data Has a clear focus on politics
Weaknessess of the products/services	Bought by Apple, future of the service is unclear Basic aggregation functions only, indications of sentiment analysis or semantic abilities Very little information available at the moment

Table 21 Topsy Labs Inc.

Tweetminster[46]/Electionista/

Tweetminster is a UK-based company that provides Twitter analytics. Originally focused on UK politics and tracking british MPs upon launch in 2008, they have expanded to cover other topics and areas in addition to this, in 50 languages. They also cover financial market news in their service. Their dedicated web application "electionista" covers politicians, governments, embassies, parliaments and media related to politics on Twitter, collecting their tweets and the links they are sharing most, providing information from over 100 countries worldwide.
All-in-all, from their dual focus on politics and finances, Tweetminster appears to most closely resemble TrendMiner in scope and target audience, while not posessing advanced analytical functions such as automatic entity detection, summaries or clustering.

Tweetminster/Elektionista	
Pricing strategies	Available on request
Sources	Twitter only, 55 different languages
Strengths of the products/services	Full real-time service, API, Visualisations Has a clear focus on politics and market news
Weaknessess of the products/services	Only basic analysis functions, like filtering

[45] http://www.forbes.com/sites/connieguglielmo/2013/12/02/apple-not-known-for-being-socially-minded-buys-social-media-analytics-firm-topsy
[46] http://tweetminster.co.uk/, http://electionista.com

	countries,retweet counts, trending topics, numbers of mentions, shared media and links, timeline filtering

Table 22 Tweetminster/Elektionista

Buzzrank (Twitterbarometer)[47]

Buzzrank is a German provider of Social media analysis. Buzzrank accesses Facebook pages, Twitter, Google+ and an undislosed number of discussion boards and blogs. It offers basic metrics for social media analysis, monitoring of keywords, sentiment analysis and report generation, via its product "Buzzrank Monitor" An additional service is "Buzzrank Connector", which serves as a CRM tool for engagement and communication with customers via Social Media channels.

Its showcase application is "Twitterbarometer", originally developed for the German general elections in 2013 and shortly after adapted to the Austrian Elections, which took place around the same time by the Austrian PR agency Mindworker, led by former politican Rudi Fussi. Twitterbarometer is a dedicated sentiment aggregation plattform for the political domain. It is based on hashtags of political party names that have either a "+" or a "-" sign suffixed to them, signifying either positive or negative sentiment, respectively (e.g. "#spoe+"). Mentions are visualised on a timeline and serve as a sentiment index. Additionally, latest Tweets are displayed in a separate frame on the website.

Buzzrank	
Pricing strategies	Free (Twitterbarometer), available upon request for "Buzzrank Monitor" and "Buzzrank Connector"
Sources	Facebook pages, Twitter, Google+ and an undislosed number of discussion boards and blogs
Strengths of the products/services	Metrics like keywords, top users, numbers of mentions and tweets, basic sentiment analysis available. Visualisations available.Data export and report generation possible Dedicated focus on the political domain
Weaknessess of the products/services	No multilingual features. No advanced analytical features No semantic entity detection

Table 23 Buzzrank

Lexalytics[48]

[47] http://www.twitterbarometer.at/(defunct) or http://www.twitterbarometer.de/ (website for germany, still running)

[48] http://www.lexalytics.com/

Lexalytics is an American company that provides a software for turning unstructured data into structured data. The software is able to detect the topic of discussion, the subject (who) and the sentiment (if it is positive or negative). The softare also detects trends, sends alerts and provides an high level of cutsomisation of services to its customers. The platform can integrate market research, social media monitoring, survey analysis/voice of customers, enterprise search and public policies.

Lexalytics	
Pricing strategies	-
Sources	Structured online data, social media
Strengths of the products/services	It provides sentiment analysis and predictive analysis. Moreover it integrate this analysis with market researches, extracts contexts and provide summarization. It allows to detect market intelligence and provides social media monitoring. It provides multilinguality support
Weaknessess of the products/services	No specific topics of analysis, is generic No real-time analysis

Table 24 Lexalytics

Temis[49]

Temis is a French company based in Paris providing a platform called Luxid that helps organizations to structure, manage and leverage their unstructured information assets. The platform is based on a natural language processing technology providing scalability. Luxid extracts structured information from unstructured documents by recognizing the key topics, entities and relations mentioned in text. IT supports 20 languages and analyses specific use cases. Temis within the Luxid platform offers to its customers four tools: Knowledge Editor, Skill Cartridge Builder, Category Workbench and Annotation Workbench. Luxid also provides a biopharmaceutical use case.

Temis	
Pricing strategies	-
Sources	Structured online data, social media
Strengths of the products/services	Luxid® Content Enrichment Platform can be deployed in the ECM of several companies to build value-added applications that leverage semantic metadata. It provides contextual alerts, intelligent archival, analysis and visualization dashboards. Temis has

[49] http://www.temis.com/home

	developed a robust and scalable natural language processing pipeline that supports fast and reliable annotation services
Weaknessess of the products/services	No specific topic focus. No real-time analysis

Table 25 Temis

AlchemyAPI[50]

Alchemyapi is an American company providing a text mining platform for sentiment analysis, used by more than 40.000 developers in six continents and supporting 8 languages. AlchemyAPI provides a popular natural language processing service via a SaaS API. AlchemyAPI offers two services: on-premise solutions for customers with specific latency, data-security or regulatory needs or a customizable service. AlchemyAPI provides entity and keywords extractions, concept tagging, relation, text and author extraction, taxonomy and language/feed detection. It allows also microformat parsing and analysis of linked data.

AlchemyAPI	
Pricing strategies	Starter: Free (1000 transactions/day, All NPL functions and 5 Concurrent requests) Small business: 250$/m (5000 transactions/day, All NPL functions and 5 Concurrent requests) Basic: 800$/m (70,000 transactions/day, All NPL functions, 15 Concurrents requests Professional: Contact sales (150,000 transactions/day, Alll NPL functions, 25 Concurrent requests Metered: Contact sales (200,000,000 +, All NPL functions, 25+concurrent requests
Sources	Structured data
Strengths of the products/services	It provides NPL through a SaaS API, microformats parsing and multilinguality. It adds high-level semantic information.
Weaknessess of the products/services	No specific topic focus. No real-time analysis and the services are expensive

Table 26 AlchemyAPI

OpenAmplify[51]

[50] http://www.alchemyapi.com

OpenAmplify is an international company, primarily based in America providing Social Media Analytics in real time. It provides also social view analytics and analysis of online conversation. OpenAmplify packs 20 years of Natural Language Processing research, 15 granted patents and 7 years of proven commercial application into a platform. OpenAmplify detects the topics and entity being discussed, domains, categories and classifications, sentiment (topic and text level), actions, intent, decisiveness, emotions and topic descriptions including sentiment scoring.

OpenAmplify	
Pricing strategies	-
Sources	Social Media
Strengths of the products/services	The platform is easy to integrate, supports many output formats including XML/JSON and can process hundreds of millions of requests per day
Weaknessess of the products/services	No specific topic focus and multilinguality

Table 27 OpenAmplify

datenwerk innovationsagentur GmbH[52]

datenwerk innovationsagentur GmbH is a private limited company headquartered in Vienna, Austria. The commercial object of the company is consultation, conception, development and sale of computer-aided methods for interaction with clients, the construction of communities and provision of services in the fields of Internet and New Media. The main target communities are businesses, politicians, institutions and organizations. Their main proprietary product is Opinion Tracker which analyses data in German and English from online news and Social Media. Both single and enterprise level subscriptions are available on a monthly fee. The software enables an interactive visualisation (including a timeline), a search function, track sharing, ad hoc analysis (most important people, topics, and organisations) and a daily email status update. It is similar to TrendMiner, but the target is only the German speaking countries such as Austria, Germany and Switzerland and English.

datenwerk	
Pricing strategies	monthly fee
Sources	Social media, online news (German and English)
Strengths of the products/services	The software provides a lot of functions: interactive visualisation, a search function, track

[51] http://www.openamplify.com/
[52] http://www.datenwerk.at/

	sharing, ad hoc analysis and a daily email status update
Weaknessess of the products/services	It is focused only on the German speaking countries Austria, Germany and Switzerland and English

Table 28 datenwerk innovationsagentur GmbH

6.4 Small scale providers

Politikeronline.at (OGM/ISA)[53]

The first of two monitoring platforms focused on Austrian politics, Politikeronline.at is an aggregator platform developed by the Austrian market research company OGM and the Austrian political consultancy and research firm ISA (Institut für Strategische Analysen). The platform covers Austrian politicans on Twitter, Facebook and YouTube, starting from federal office holders and candidate lists and planning to move on to municipal levels at later dates. The platform can be accessed freely, as part of a drive to offer better access to politicians in Austria and support political discourse. Analysis features are limited, ranging from Top Tweets and ranking politicians by activity level, as well as basic topical classification. This appears to be more of a showcase and pro bono project for the two companies involved and seems to be directed at interested citizens mainly.

Politikeronline	
Pricing strategies	Free service
Sources	Twitter, Facebook, YouTube
Strengths of the products/services	Near real-time updates, basic retrospective analysis possible Basic entity recognition Dedicated focus on the political domain Aims to enhance political discourse in Austria
Weaknessess of the products/services	Only very basic search and analysis functions, no visualisations, no sentiment analysis, no advanced analytical functions, no functions for raw data extraction, no multilingual features, , no semantic entity detection Focused only on Austria

Table 29 Politikeronline

Politometer (Superfi/monopol)[54]

[53] http://www.politikeronline.at/

This Austria-focused platform is run by Austrian Advertising, PR and media agency Superfi/Monopol, two companies founded by an Austrian entrepreneur turned politician, Niko Alm. The politometer is actually a politically-oriented spin-off of the company´s "Social Media Ranking"[55]. Rather than aggegating content, both web applications access public data from Facebook, Twitter, google+ and foursquare to create scores based on range, activity and engagement metrics. The political actors featured on the website are grouped into the categories "Parties", "Politicians", "Members of Austrian Parliament", "Government", "Members of European Parliament", "NGOs", "Media" and "Journalists". Rankings are refreshed once every week. This platform seems to act as showcase for the operating company, as the results are freely available.

Politometer	
Pricing strategies	Free service
Sources	Twitter, Facebook, Google+, Foursquare
Strengths of the products/services	Basic entity recognition, based on human curation Dedicated focus on the political domain
Weaknessess of the products/services	Only basic analysis functions, No content aggegration, no visualisations, no sentiment analysis, no advanced analytical functions, no functions for raw data extraction, no real time updates, no multilingual features, no semantic entity detection Focused only on Austria

Table 30 Politometer

[54] http://www.politometer.at/

[55] http://www.socialmediaranking.at/index.php

Conclusions

The analysis of the current social media landscape and market analysis for services similar to TrendMiner detected an high level of competition. Indeed, especially in the Social Media Monitoring and in the Content Analytics frameworks we proposed a clear description of a complex framework constituted by a huge set of big and small companies.

International organisations providing sentiment analysis services for the financial scenario are offering really different services, but there is not an homogeneity of functionalities and none of them dominates the others. In terms of sentiment analysis for the political use case, the topic is still new and only few providers are currently working in this specific field. The services provided are not specific and the progress of research and services offered in this field is still at a basic level.

By analysing the overall framework, we can finally provides some considerations about the future of the TrendMiner services and products that will constitute the value propositions of TrendMiner:

- one of the main advantage of TrendMiner is the ability to provide multilinguality and this will constitute one of the core benefit of the platform. Only other 4 Content Analytics providers out of 15 offer sentiment analysis services by using different languages.
- The ability to predict the sentiment and to provide an entity specific sentiment index and alerts are the main benefits of TrendMiner.
- TrendMiner is able to offer portability of services.
- TrendMiner software provides tools for risk management by using the sentiment detected. This is really valuable especially in the case of financial analysis. Only few other Content Analytics providers take into account risk management services and offer them to their customers.
- TrendMiner provides both for the political and financial use cases cross lingual mining, the identification of the topic/leader, crawling and monitoring activities, a semantic search interface and metadata/entity extraction.
- The main benefit of TrendMiner will be constituted by the pricing strategies, as very few Content Analytics Providers offer sentiment analysis services at affordable prices and TrendMiner will be able to reduce this pricing.

The analysis provided here in the Market Watch was relevant in order to identify the previous mentioned relevant benefits of our services and to show the high potentialities of the TrendMiner software, also in terms of future developments and ability to efficiently enter the sentiment analysis market for the financial and political use cases.

Bibliography and references

Cavazza F., Social Media Landscape 2013,
http://www.fredcavazza.net/2013/04/17/social-media-landscape-2013/
(last access 03.03.2014)

Faktor S., The 10 types of twitter users, 2013, available at
http://www.ideafaktory.com/social-media/on-forbes-the-10-types-of-twitter-users-and-how-to-make-them-to-love-you-2/ (last access 03.03.2014)

Fenn J, Raskino M, Gammage B (2009): *Hype Cycle for Enterprise Information Managemen*t. Gartner

Grimes S, (2011): *Text/Content Analytics 2011: User Perspectives on Solutions and Providers.* Alta Plana

IDC, (2013), Worldwide Business Analytics Software 2013–2017 Forecast and 2012 Vendor Shares, http://idcdocserv.com/241689e_sas(last access 04.04.2013)

Maireder A, Ausserhofer J, Kittenberger A (2012): “Mapping the Austrian Political Twittersphere.” In: *Proceedings of CeDem12 Conference for E-Democracy and Open Government*. 151–154. Danube University Krems

Maireder A (2011): *Links auf Twitter - Wie verweisen deutschsprachige Tweets auf Medieninhalte?* Vienna. http://www.univie.ac.at/publizistik/twitterstudie/
(last access 29.10.2012)

McGuire M., Hype Cycle for Media and Entertainment, Gartner 2013, available at http://www.gartner.com/document/2561116?ref=QuickSearch&sthkw=hype%20cycle%202013%20for%20media%20and%20entertainment%202013

Mislove A, Marcon M,Gummadi KP, Druschel P, Bhattacharjee B (2007): Measurement and Analysis of Online Social Networks. In: *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement.* 29-42, San Diego

Rozwell C., Sallam R., Hype Cycle for Content and Social Analytics, 2013, Gartner, 23.07.2013,
http://www.gartner.com/document/2556115?ref=QuickSearch&sthkw=hype%20cycle%20for%20social%20analytics
(last access 04.04.2014)

Rozwell C., Sallam R. L., Hype Cycle for Content and Social Analytics 2013, Gartner, 2013, available at
http://www.gartner.com/document/2556115?ref=QuickSearch&sthkw=hype%20cycle%20for%20content%20and%20social%20analytics%202013

Statistic Brain, Twitter statistics, 2014, available at
http://www.statisticbrain.com/twitter-statistics/ (last access 01.01.2014)

Links

http://www.attensity.com/home/

http://www.abilityfactors.com/

http://www.alchemyapi.com

http://www.brandwatch.com/

http://blogs.forrester.com/gina_sverdlov/12-01-04global_social_technographics_update_2011_us_and_eu_mature_emerging_markets_show_lots_of_activity

http://www.crimsonhexagon.com/

http://www.datasift.com

http://www.datenwerk.at/

https://dev.twitter.com/blog/changes-coming-to-twitter-api

http://www.election.twitter.com

http://electionista.com

http://en.wikipedia.org/wiki/User-generated_content

http://www.finif.com/

http://www.finsents.com/

http://www.gnip.com

http://www.h2oconsulting.ch/

http://www.lexalytics.com/

http://www.mediassociaux.fr/2011/02/06/description-des-differents-types-de-medias-sociaux

http://www.moreover.com

http://www.openamplify.com/

http://www.opencalais.com/

http://www.ravenpack.com/products/ravenpack-news-analytics/

http://www.semlab.nl/

http://www.sentdex.com/

http://www.sntmnt.com/

http://www.socialmention.com/

http://www.talkwalker.com

http://www.tame.it/

http://techcrunch.com/2014/04/15/twitter-acquires-longtime-partner-and-social-data-analytics-provider-gnip/

http://www.temis.com/home

http://www.thomsonreuters.com/products/financial-risk/01_255/News_Analytics_-_Product_Brochure-_Oct_2010_1_.pdf

https://www.tool.opiniontracker.net

http://www.trendiction.com

http://www.twendz.waggneredstrom.com

http://www.twitterbarometer.at/(defunct) or http://www.twitterbarometer.de/ (website for germany, still running)

http://www.twittersentiment.appspot.com/

http://tweetminster.co.uk/

http://www.twittratr.com

http://www.visualization.geblogs.com/visualization/cancerconversation