FP7-ICT Strategic Targeted Research Project TrendMiner (No. 287863)
Large-scale, Cross-lingual Trend Mining and Summarisation of Real-time Media Streams

6.3 Application final results

Francesco Bellini (Editor),
Nicola Fiore, Loretta Latronico, Luca Satolli - Eurokleis s.r.l.

Abstract

FP7-ICT Strategic Targeted Research Project TrendMiner (No. 287863)
Deliverable D6.3 (WP 6)

The aim of this deliverable is to summarise the activities done in Work Package 6. A web-based prototype summarising the media stream in terms of its likely impact on a selected financial asset from economic and political-economic perspectives has been iteratively developed. The application is able to connect and visualise the events occurring in the media stream (WP3) and a tailored interactive summary (WP4) of these data with price movements of a given stock or index. The results of the prototype have been evaluated and summarised in this report collecting the results from the focus group participants. The prototype has been fully integrated in the TrendMiner platform developed in WP5.

Keyword list: financial markets, news, sentiment, behavioural finance, polarity, forecast, politics

Nature: Prototype
Dissemination: PU
Contractual date of delivery: 31/10/2014
Actual date of delivery: 03/11/2014
Reviewed By: Thierry Declerck (DFKI)
Web links: N/A

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 Révolution
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 aim of this deliverable is to summarise the activities done in Work Package 6 and in particular in task 6.3 “Application final results”. A web-based prototype summarising the media stream in terms of its likely impact on a selected financial asset from economic and political-economic perspectives has been iteratively developed. The application is able to connect and visualise the events occurring in the media stream (WP3) and a tailored interactive summary of these data with price movements of a given stock or index. The results of the prototype have been evaluated and summarised in this report collecting the results from the focus group participants.
According to the TrendMiner platform features (WP5), we developed the following process and activities that resulted in a number of software components and available data sets:

- Data Collection
- Annotation
- Computation of polarity
- Prediction of the market behaviour
- User interface development

Three examples are used as a proof-of-concepts for validating the prototype outcomes against the known market behaviours and existing literature. The TrendMiner financial use case prototype shows the ability to play as another decision support tool beside the consolidated market forecast techniques such as technical and fundamental analysis.

Contents

1 Introduction

1.1 General Background and context

The objective of work package 6 is to turn the project technologies towards practical applications in the financial domain. The work package is developing a real-time financial decision tool primarily aimed to support investors, analysts, advisors, stock brokers, but also to financial regulators, journalists and citizens. In particular, this deliverable aims to describe the activities carried out to date in task 6.3 for the development of a web-based prototype summarising the media stream in terms of its likely impact on a selected financial asset from economic and political-economic perspectives. As provided in the DoW the activities have been conducted in collaboration with the political use case workpackage (WP7) and the technical partners (WP2 and WP3) in order to agree about the data and documents upon which the core technologies will be customised and validated for the financial decision use case.

1.2 Structure of the document

The document is structured in three chapters. After the introduction, the second chapter describes first the main features of the TrendMiner platform and then provides details about the specific activities performed for collecting the data and implementing the financial use case. The last paragraph of chapter 2 describes how the user interface works at the present stage of development.
The last chapter provides an outlook of the follow up activities foreseen to optimize the prototype for the financial use case.

1.3 Relevance to TrendMiner

Given the objectives of WP6, this document summarizes the activities done during the project lifetime. This document provides a description of the activities needed to reach the TrendMiner goals.

As from the DoW, the nature of this deliverable was planned to be as a Prototype (P) and the prototype for the financial use case can be accessed at http://ek-uc.ontotext.com/TrendMiner/tracks.html.

The target of the deliverable is the wider public.

1.3.1 Relevance to project objectives

This deliverable provides the final documentation and evaluation of the TrendMiner financial uses case and the requirements for the implementation in a commercial phase.

1.3.2 Relation to other workpackages

This deliverable and the activities done in WP6 benefit of the work carried out in all the technical (WP2 and WP3) and applicative workpackages (WP7). The WP6 use case has been integrated within the TrendMiner platform in the context of WP5.

2 Financial Use Case development

The following paragraphs describe the system architecture and the process included in the framework of the Financial Use Case.
The process considers the following phases:

- Data Collection
- Annotation
- Computation of polarity
- Studies of the market behaviour
- User interface and examples of market investigations

2.1 System architecture and components

As described in the D5.3.1 and D5.3.2 the high-level architecture for the TrendMiner platform can be summarised in the following Figure 1:

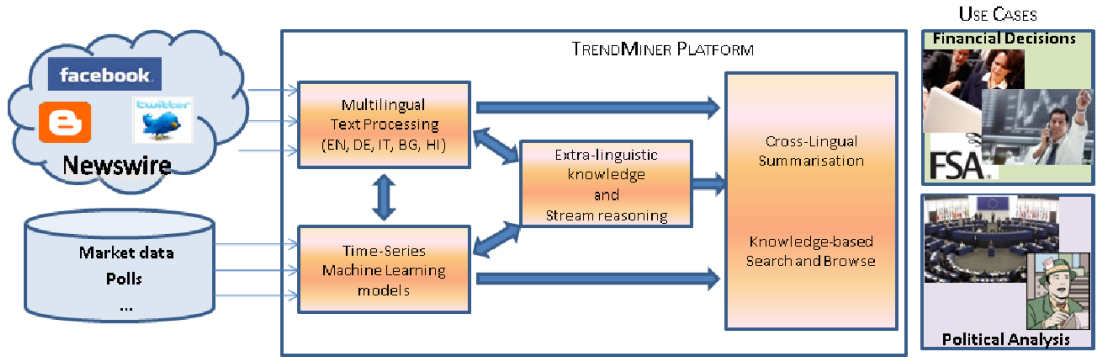


Figure 1 - TrendMiner Platform

This architecture implies a number of tasks to be carried out sharing the common challenges of the data volume and the data velocity.

2.1.1 Data Collection

This task provides large scale, real-time collection, aggregation and storage of data from social media streams. The major challenges for this task are: data volume, velocity (timely new data harvesting) and variety (heterogeneity of sources and formats)

2.1.2 Pre-processing

Various tasks from WP2 (Multilingual Ontology-Based Information Extraction and Knowledge Modelling), WP3 (Machine Learning models for Mining Trends from Streaming Media) and WP5 (Platform for Real-time Stream Media Collection, Analysis, and Storage) need to pre-process data in order to improve the quality and reduce the time of the subsequent data processing. Such pre-processing tasks include data cleaning, data de-duplication, some basic extraction and pattern matching tasks, formatting tasks.

2.1.3 Multilingual Information Extraction

The ontology based multi-lingual information extraction tasks within WP2 aims at extracting structured information from unstructured sources. Such information may include:

- entity mentions (of people, organisations, places, etc.);
- relations between two entities;
- more complex structures involving (related) entities, like events in various periods.

2.1.4 Entity Disambiguation and Linking

Entity disambiguation aims at identifying the concrete class and the unique entity corresponding to a text label in the input. Depending on the particular approach employed – string similarity, structural similarity or context similarity techniques (or even a combination of the above), the disambiguation process may become computationally expensive.

2.1.5 Sentiment Extraction and Trend Detection

Sentiment extraction and trend detection in WP2 and WP3 aim at extracting sentiments from real-time streams and detecting trends based on various statistical approaches (regression and clustering models) and rule-based approaches (relying on the TrendMiner knowledge base). The Machine Learning tasks have two different phases: a *training* phase and a *prediction* phase and that the volume and velocity of the data is different for the two phases. The training phase is performed offline at certain intervals (daily, in the beginning) and is done over the batch of data for the interval (day). The prediction phase on the other hand will be performed in real time, as new data enters the system, or in relatively small batches for short time periods. This phase is not computationally expensive, but requires a distributed processing approach in order to scale to the amount of new data in real time.

2.1.6 Summarization

The summarization task within WP4 aims at identifying representative text fragments for trends and events within a time-period.

2.1.7 Stream reasoning

Stream reasoning aims at deriving new knowledge from existing knowledge. Unlike traditional reasoning approaches which aim at deriving the complete closure of a set of logical statements (via forward or backward chaining approaches), stream reasoning works by deriving only a partial closure of the statements, with the added benefit that it can scale to cope with real-time input streams which will otherwise be impractical to process with traditional approaches.

2.1.8 Querying over annotated streams

The data generated by all the processing tasks in the TrendMiner pipeline is persistently stored (by WP2) for further access, aggregation and visualisation (by

components within WP4). The storage and query answering tasks are constrained by the data volume and velocity, since the storage and query engine must be able to handle efficiently storage and indexing of huge volumes of data, as well as efficient querying over it.

2.2 Data collection

Data collection has been strongly improved, especially with regards to Twitter accounts. The activities described in the following paragraphs have been carried out in order to isolate a number of meaningful resources needed for the use case implementation. Annex 1 includes a sample of the dataset built for this delivery.

2.2.1 Tweets Selection and Collection

In order to develop the last version of the prototype, we have completely revised and strongly improved the strategy of Tweets collection. In the previous version (D6.2) we drove the tweets collection through specific keywords, also because the crawler developed by the IMR partner was not able (for legal reasons) to crawl Twitter accounts. The keywords were manually extracted from news that generated "rumours" during an observed period. Keywords allow building a detailed sample of Tweets which isolates the comments on observed companies, persons and products.

In this final version we reverse the approach: instead of collecting the tweets that match certain keywords, we collect all the tweets originated from selected financial sources and in a second stage we carry out the search for specific keywords. In this way we do not restrict the collection only to the topics related to some rumours, but we collect whatever is generating interest in the financial domain and we build a dataset to be further investigated afterwards.

The identification of sources was a time consuming task and was carried out following the approach below:

- worldwide search of Twitter accounts accessed by people who shares ideas on investments and global economy;
- identification, within each account, of lists of users focused on specific issues of economics and finance. A preference was given for the Twitters acquisition. Our search was aimed to expand as much as possible the number of users who share investment ideas, and for this objective we visited each account individually;
- in the financial domain English is the language primarily accessed worldwide, so the majority of the selected lists use this language. However, some lists in German and in Italian language were selected too.

By using the above described approach, more than 3000 users were identified; of which more than 2000 are individual users and the remaining 1000 are users who operate within news providers. We started to collect the Tweets selected as above since August 2014 and we synchronized our collection with the Data Warehouse of the TrendMiner platform (see D5.3.1 and D5.3.2). We believe that this mix of users generated a good data set as a starting point for our investigations. A key point for future development will be its maintenance and improvement.

In order to perform the collection of Tweets we used the APIs provided by Twitter that allows to search within the timeline according to the keywords provided.
A python routine[1] was implemented by using these APIs. The routine creates the connection, performs the authentication and run the search on the timeline. The routine was scheduled twice a day for the same period of collection described During this period, the keyword list has been daily refined according to the trending topics in order to refine the selection of Tweets that will concur to the sentiment definition.

2.2.2 News Collection and Analysis

The news documents are collected from selected sources[2] that are supposed to provide "price sensitive" news. The information has been crawled within WP5 using the Internet Memory tool according to the sources identified by WP6 researchers and to the following crawling strategy:

a. Companies

- Official Web Page:
  - Crawl URL
  - Crawl 1 hop of links in the same URL domain

- RSS feeds:
  - Crawl RSS feed URL
  - Crawl item URLs published by the feed
  - Do not crawl links in item pages

- Official reports:
  - Crawl URL
  - Crawl 1 hop of links in the same URL domain

- News reports
  - Crawl URL
  - Do not crawl links in pages
  - Crawl RSS feed of headline news
  - Crawl item URLs published by the feed
  - Do not crawl links in item pages

- Historical prices
  - Crawl URL
  - Do not crawl links in pages

[1] https://wiki.internetmemory.org/download/attachments/23891397/TweetSearchGrabber.py?version=1&modificationDate=1385929869821&api=v2
https://wiki.internetmemory.org/download/attachments/23891397/TwitterProcess.sh?version=1&modificationDate=1385929918025&api=v2

[2] https://wiki.internetmemory.org/download/attachments/23891397/sources_data_collection_EK_01122013.xlsx?version=1&modificationDate=1385915316139&api=v2

b. News providers

- URL
  - Crawl URL
  - Crawl 1 hop of links in the same URL domain

- RSS feeds:
  - Crawl RSS feed URL
  - Crawl item URLs published by the feed
  - Do not crawl links in item pages

The results of the collection can be visualised and accessed through the IMR dashboard[3]. The following figures show a summary of the data collection results performed in the period 29/10/2013-29/08/2014.

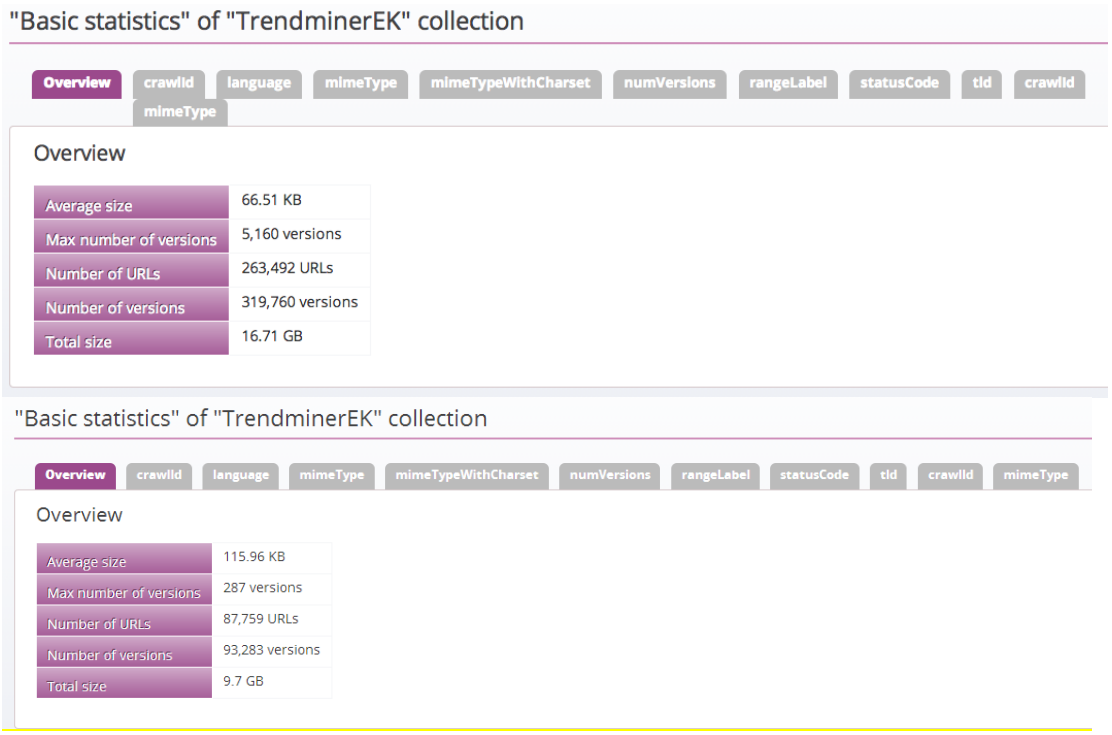


Figure 2 - Data collection basic statistics

[3] http://dashboard.mignify.com/

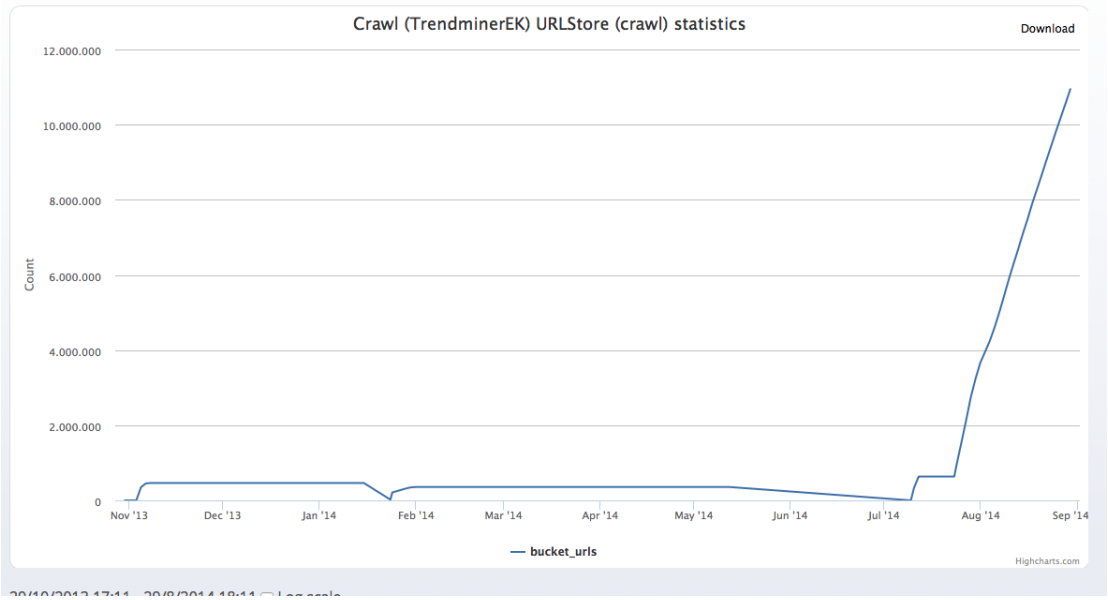


Figure 3 - Bucket URLs

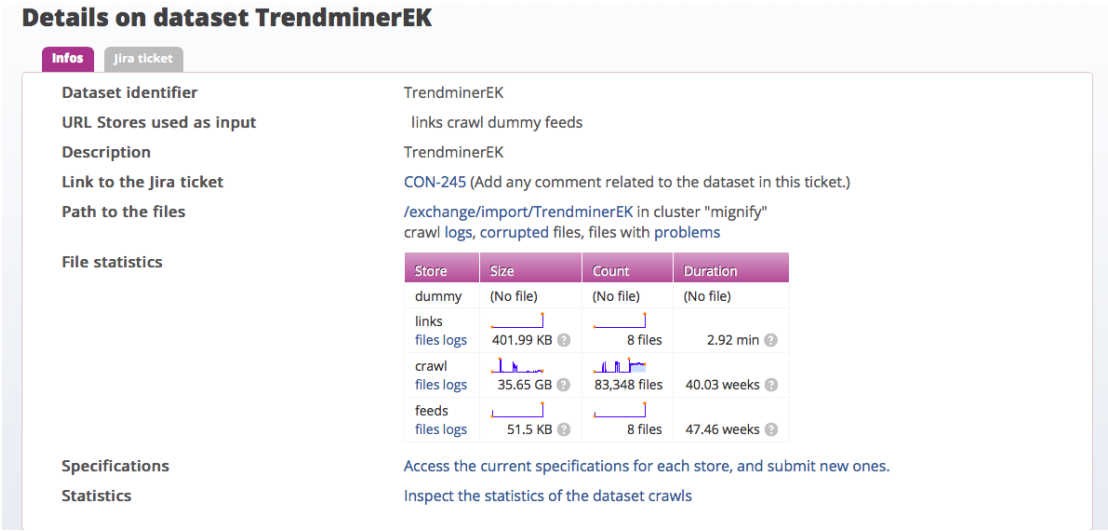


Figure 4 – Financial news dataset details

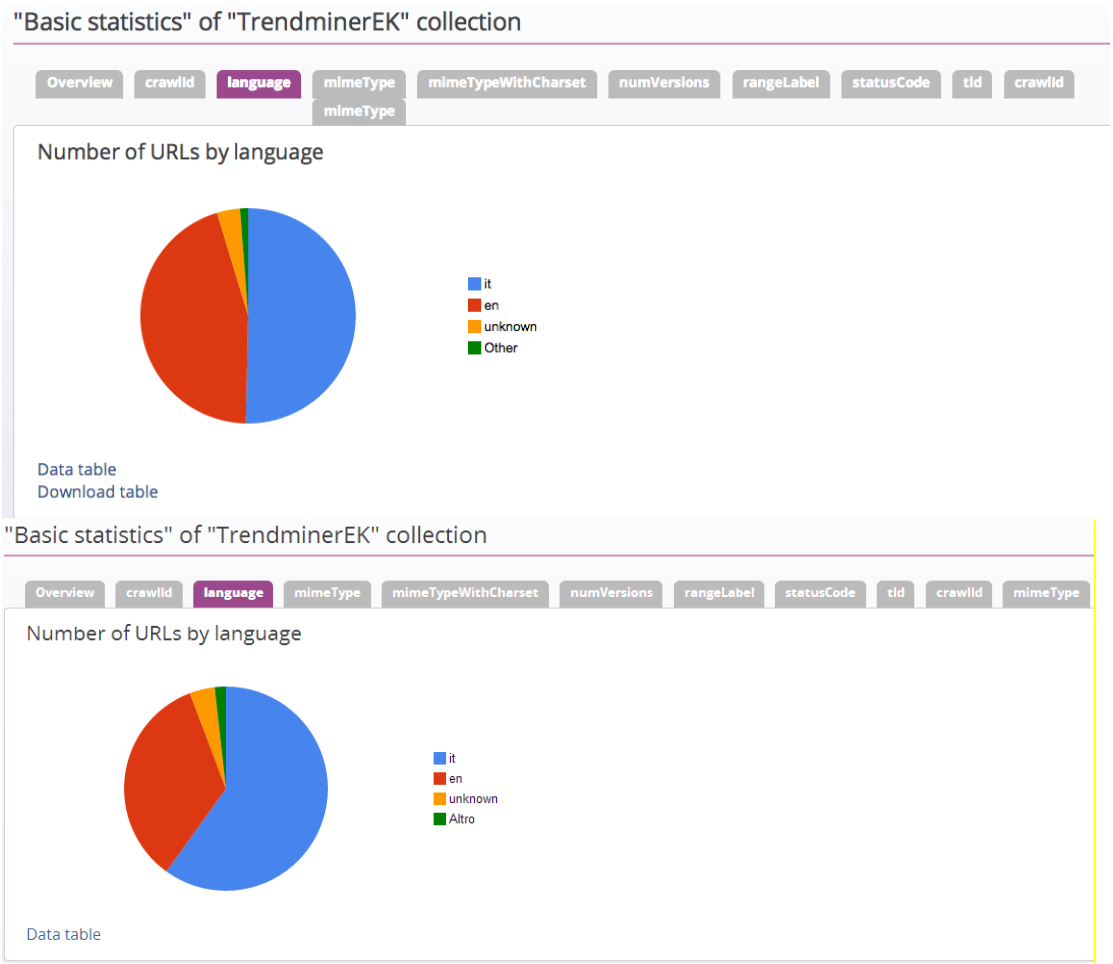
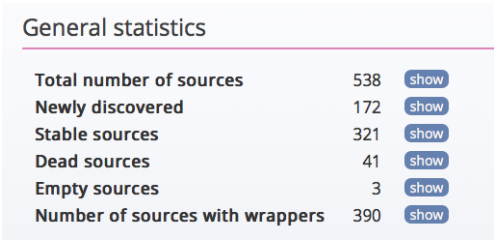


Figure 5 - Number of URLs by language



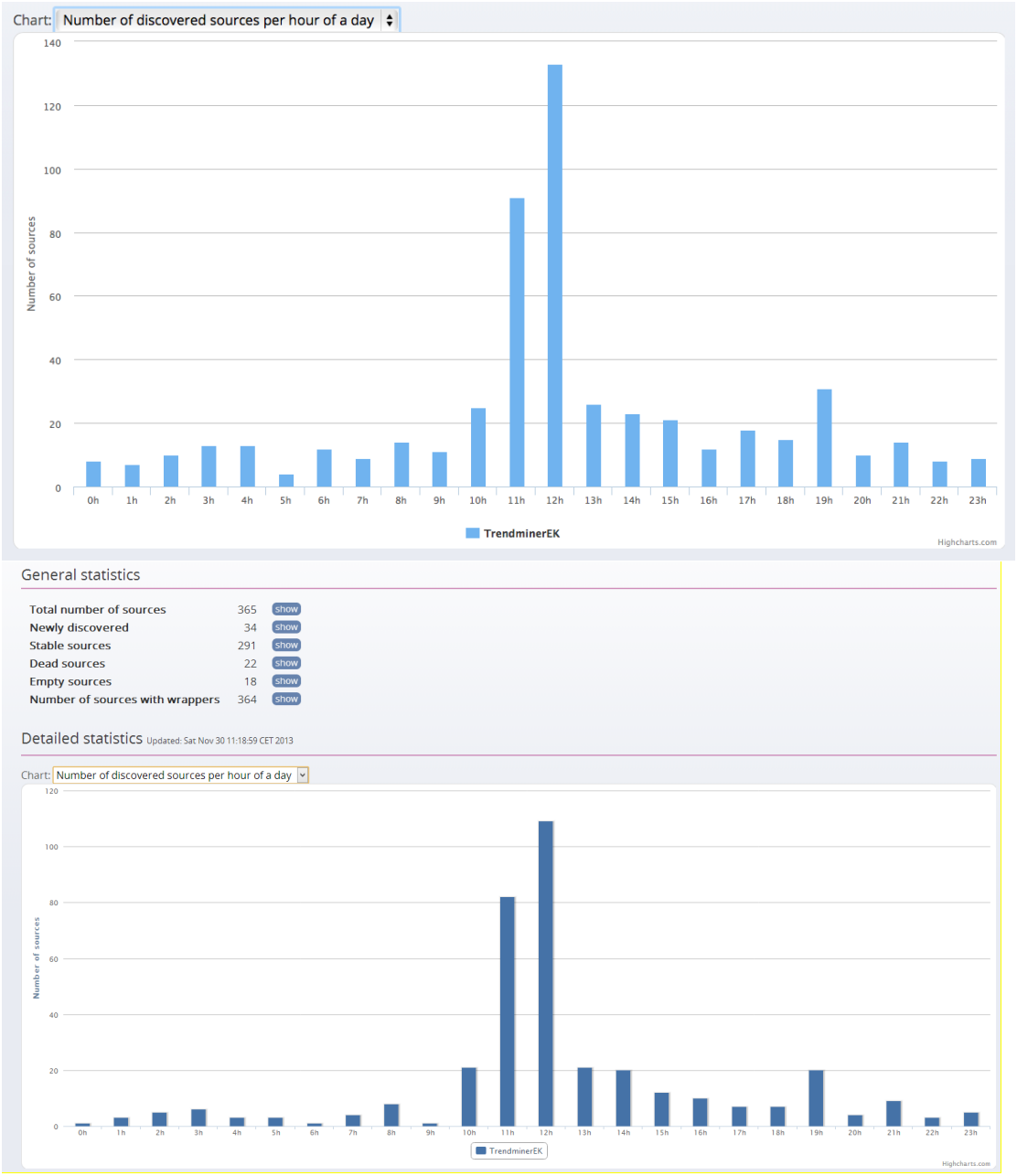


Figure 6 - Number of discovered sources per hour of a day

The last figure is interesting because it shows a peak of potentially relevant news published every day between 11 and 12 when the European markets are usually fully operational.

2.2.3 Financial instruments data collection

In order to combine sentiment and market prices we have designed and implemented a structure to collect market data with the following features:

- the sentiment can be correlated not only with equity prices but also with currencies, bonds, common funds, exchange trade funds (ETF), etc. This means that the data structure shall allow the detection of each kind of prices

that we generalized through the Financial Instrument (FI) concept. In this way we can access prices and volumes (if any) of any type of market data.

- the structure is flexible in terms of time period to allow investigating the links between sentiment and market data on monthly, weekly, daily, and infra daily basis.

The prototype focuses only on stock prices and stock indices gathered daily, however more refined investigations could be conducted in the future.

Annex 1 reports, in the section related to the Companies, the indications about the price sources for each company. The sample does not shows market indices but, thanks to the architecture, they can easily be added being the index a financial instrument as well as the company stock price.

2.3 Annotation

The output of the collection is a list of json arrays containing the Tweets. This format is processed by the text analysis tools developed in WP2 and WP3. In order to compute the sentiment the annotation marks positive and negative terms for each Tweet in a given day. We are taking here also advantage of the possibility to encode aggregated opinions in the MARL model that has been updated by DFKI in the context of the TMO ontologies (see D2.1.2 Knowledge & Provenance Modelling and Stream Reasoning - v2).

2.4 Computation of polarity

The following step is the evaluation of entity polarity[4] using the sentiment data for a specific single day. The polarity is the sentiment associated with the entity referring to the key-words list and can be positive, very positive, negative, very negative. We give a score to each word depending on the category of polarity it belongs to and increase/decrease the polarity strength when a word is preceded by a modifier i.e.: not good = -1; good = +1; very good = +2. The same thing if the Tweets contain one of positive/negative words plus one of word from the Strong Word List. We developed for this a tool that reads shorts notices associated with actual values of companies listed at the MIB (stock exchange in Milano) and using a heuristic combining repeated word in the context of an increase or a decrease of the values, we generated an opinionated lexicon for the Italian language (in the financial domain). This lexicon has been merged with a "classical" computational lexicon of Italian (the Italian version of the NooJ tools, that DFKI is running.[5]), and it is being manually checked against longer texts, also in different domains.
On the polarity data, we are also consulting the list of words deemed to be positive, negative, as they are defined in the subjectivity lexicon of Loughran and McDonald Financial Sentiment Dictionaries[6]. This for the English text, but we also provided for

[4] Entity polarity: combination of previous scores to have an overall polarity for the entity in a day (Godbole et.al, Zhang and Skiena)
[5] http://www.nooj4nlp.net/pages/nooj.html
[6] http://www3.nd.edu/~mcdonald/Word_Lists.html

Italian translation. We are currently implementing a sentence level polarity computation for Italian newspapers, and porting the strategy to Tweets.

2.5 User interface and examples of market investigations

The Financial Use Case aims at investigating any link or relationship between financial market and investors sentiment derived from text mining. The sentiment is to be treated as a second source of information for trading and investment decisions. As a proof-of-concept, three different examples are analysed and discussed below.

The TrendMiner user interface allows to show time-based sentiment and activity on a particular topic of interest and compares them visually with the time series of a financial instrument (price, index, etc.). Since the user interface has been designed at a prototype level and permits only to show few changes in sentiment over time, we can display daily values for max one month and if we analyse a longer period we get weekly readings. In addition, the interface is unique for all the TrendMiner Use Cases so it does not include sophisticated financial features and quantitative tools. It allows to analyse financial markets and investor sentiment at the "explorative" level but this permits in any case to get very interesting results. The following figures show how the user interface works according to the available components. Figure 1**Fehler! Verweisquelle konnte nicht gefunden werden.** lists the track created to investigate specific topics.

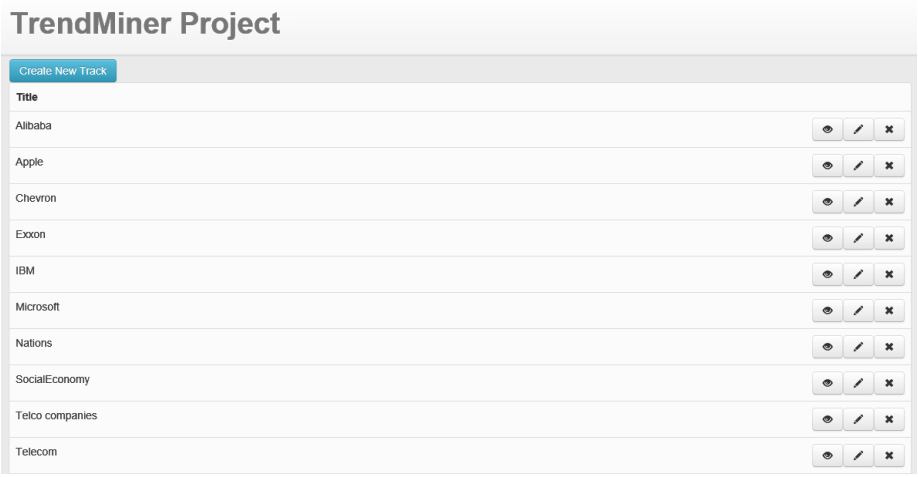


Figure 7 – Cockpit for track management

2.5.1 Social-economy

As a first example, we choose to investigate a social-economy situation creating a track (with the parameters showed in Figure 8) to which corresponds the entity distribution showed in Figure 9.

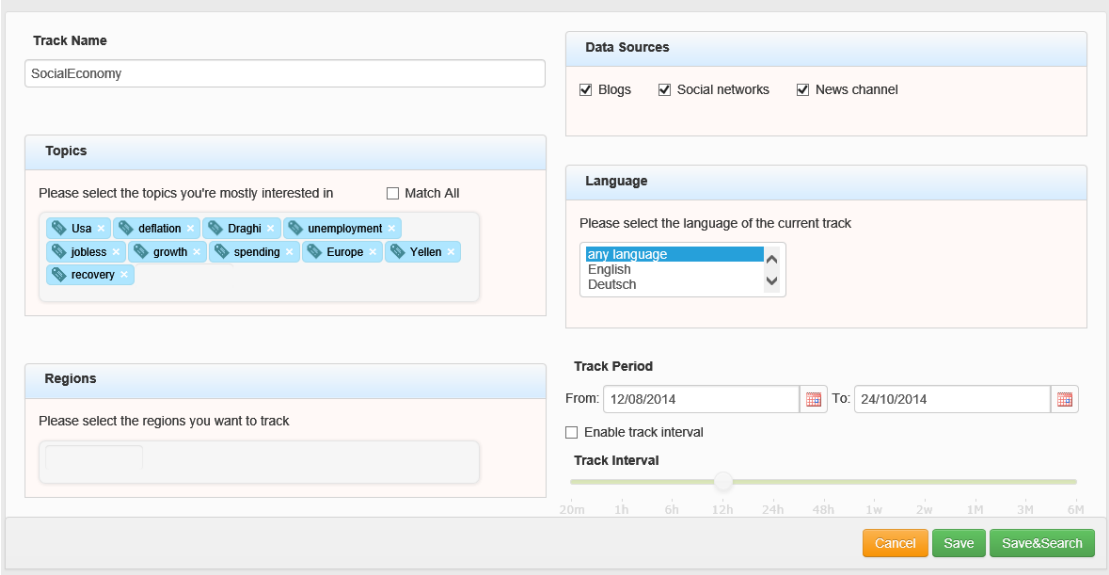


Figure 8 – Track setup

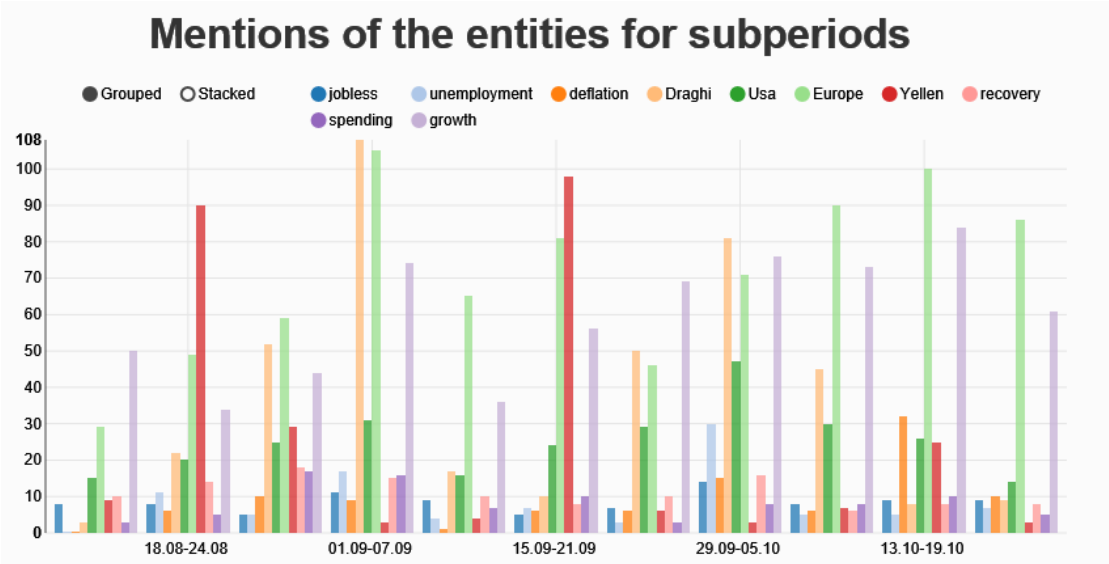


Figure 9 – Number of Entities for each topic

As a first result, we can see that entities like Draghi, Yellen, growth, Europe, Usa are more mentioned than topics like jobless, unemployment, deflation, recovery. We see also that Draghi and Yellen are mentioned when they approach their periodic speech, and, apart these events, the interest of the financial community is less evident. Instead, the topic growth is evident over all the observed period, and this also applies to the topic Europe which is more evident than USA.

We also tried more topics like austerity, claims, households and others but their influence is low if compared with the above terms, and we experienced that lower activities have low impact over the sentiment readings. This first screening is important in order to establish which terms should be monitored over the time (they could be the topics with more weight on the sentiment changes) and which terms should be discarded instead.

The sentiment, determined according to of WP2 and WP3 tools, is shown in Figure 10. Here the average value is 0.24; an increasing trend is seen in the first period with a peak in the middle of September. After that the sentiment drops. A possible interpretation of this pattern could be an increasing expectation from Draghi and Yellen speeches of mid-September and a decreasing interest after this event. The FTSE100 time series (Figure 11) seems to be in line with the sentiment.

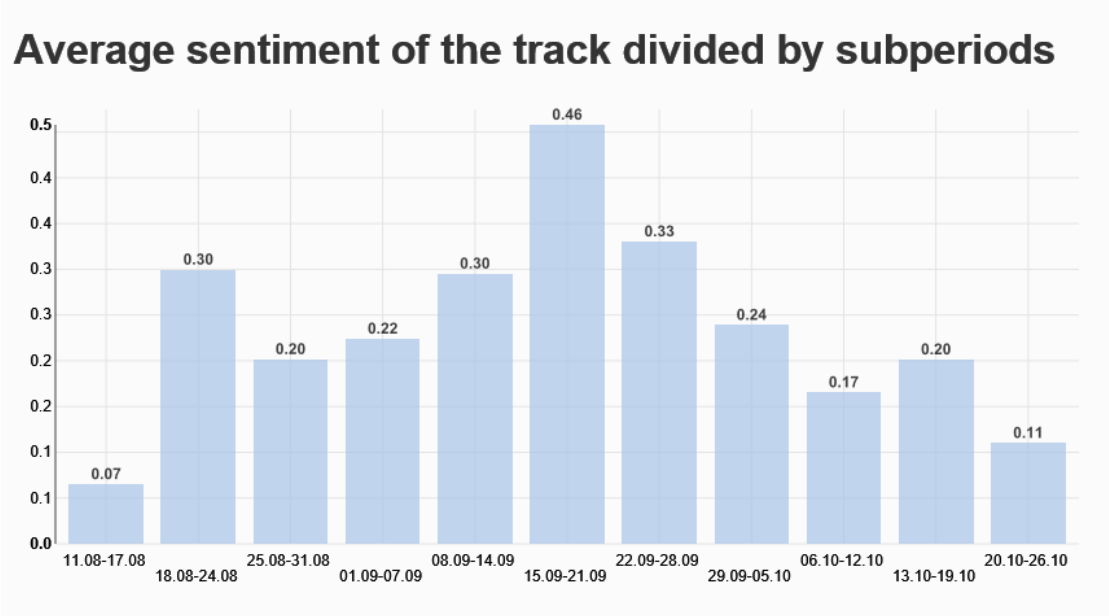
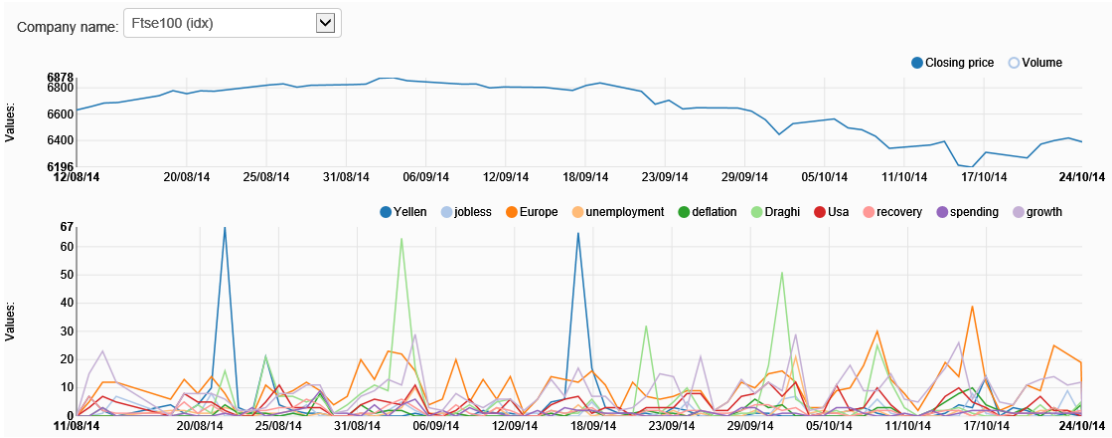


Figure 10 – Sentiment pattern



The following figures show how the user interface works according to the available components. The cockpits in Figure 1**Fehler! Verweisquelle konnte nicht gefunden werden.** and **Fehler! Verweisquelle konnte nicht gefunden werden.** show how a query is built according to topics, data source, locations, language and time period.

Figure 11 – Topics activity and FTSE100 time series

Figure 12 – Co-occurrence of entities

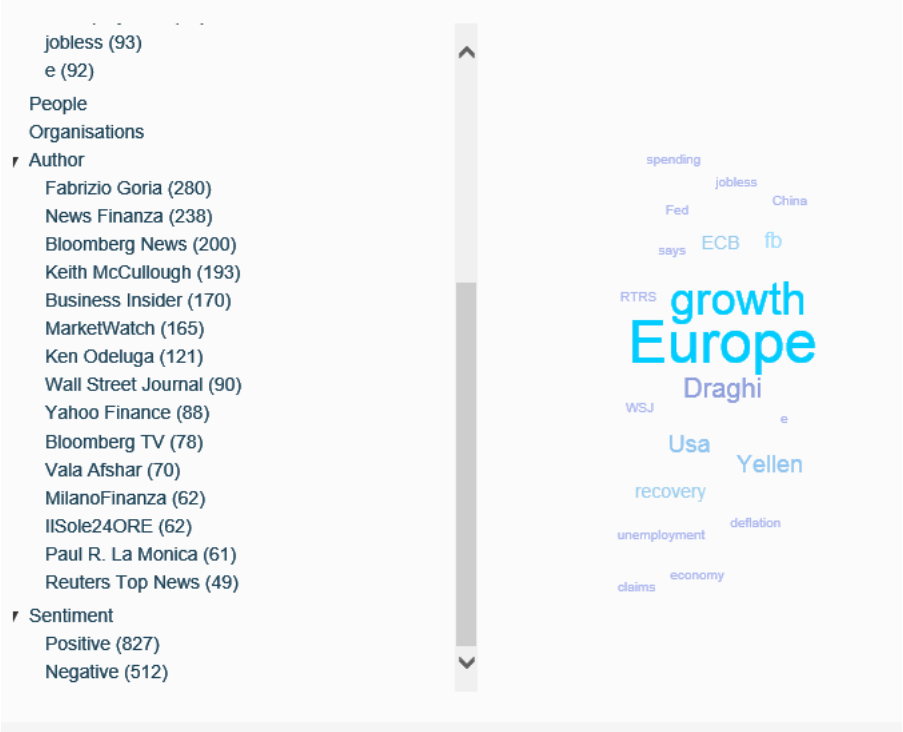


Figure 13 – Topics, word cloud and source indications

2.5.2 Initial Public Offering (IPO)

As a second example we investigated the IPO of Alibaba Group which took place on September 19, 2014. The parameters of the track are shown in Figure 14.

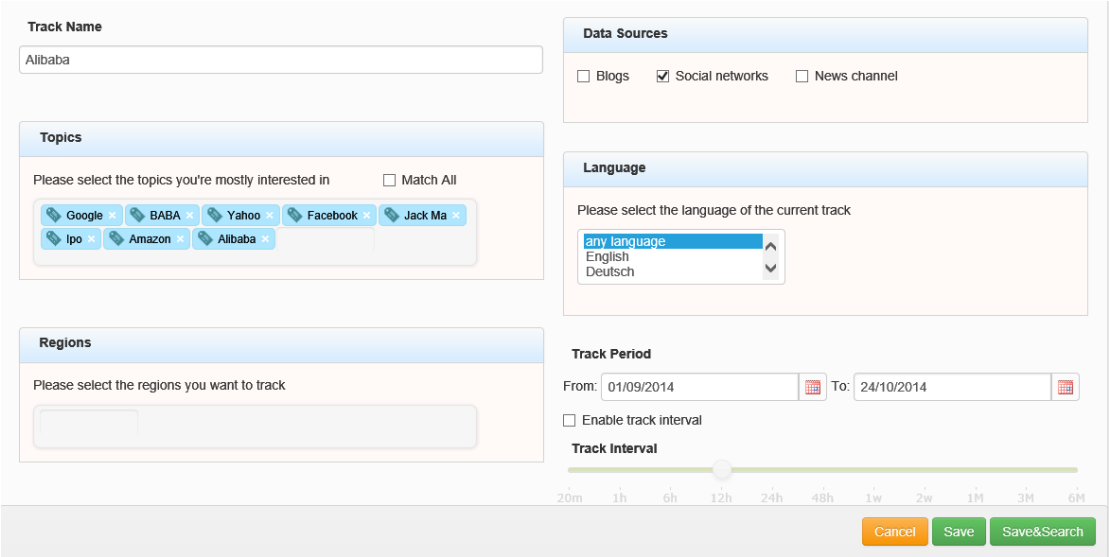


Figure 14 – Alibaba track creation

The entities chart of Figure 15 shows a high interest around the IPO date. We compared Alibaba against some peers, in order to see if some discontinuity happens in the period. As can be seen, when the listing is over all the peers seem to have the same mood as it is logic, and this is an important result in terms of reliability of our dataset.

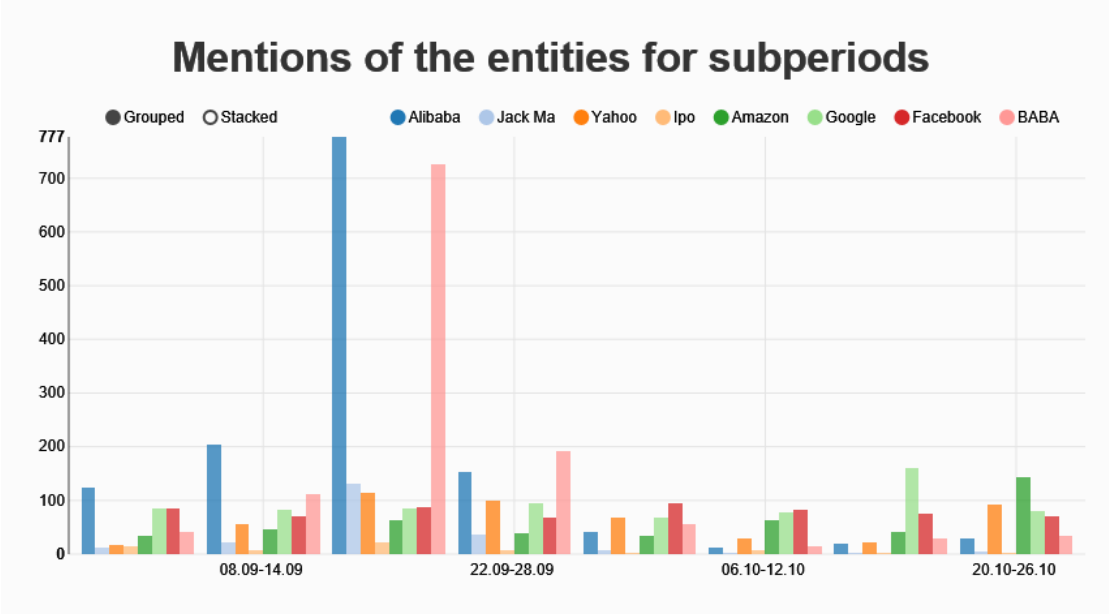


Figure 15 – Alibaba entity distribution

The sentiment chart (Figure 16) shows an average value of 0.47 that is higher than the average of the social-economy case. On the other hand, the trend seems the opposite to the one of the stock price, meaning that sentiment roughly increases when stock price decreases.

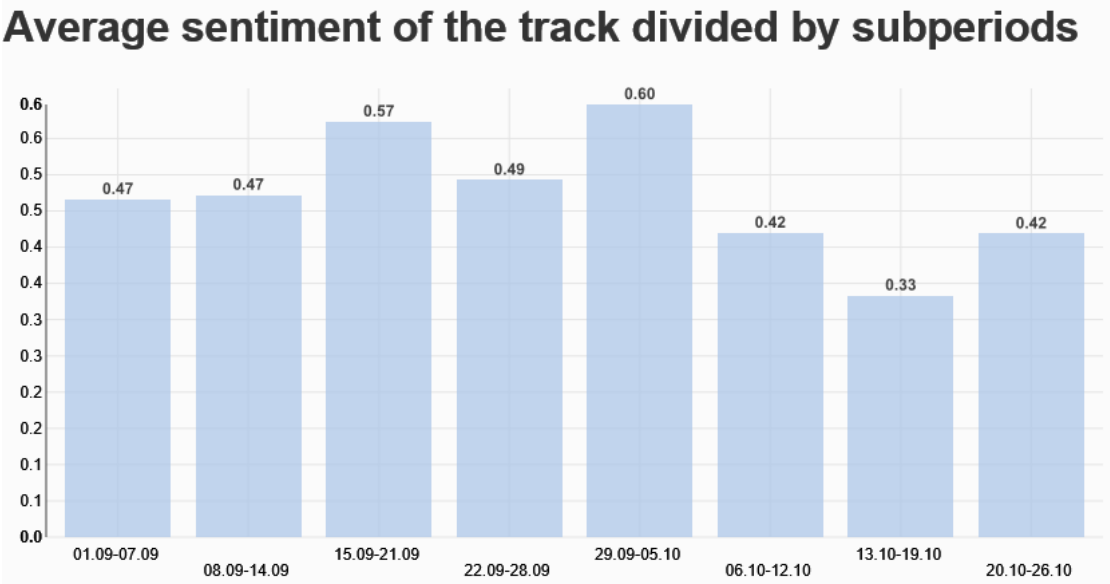


Figure 16 – Alibaba sentiment pattern

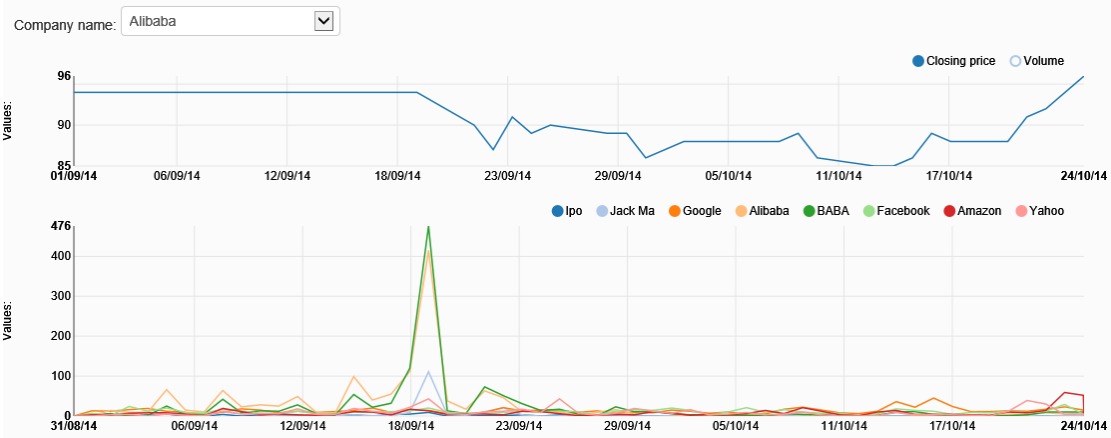


Figure 17 – Alibaba activity and stock exchange price

The stock price decreases immediately after the listing date and increases at the end of the observed period. Figures 18 to 22 provide a detailed view around the listing date. This happen thanks to a specific functionality has been implemented as an additional diagram on top of the main timeline filter, allowing the user to focus on specific period by moving or changing time window selector.

During the days close to the listing date, activities are driven by Alibaba's topic while the sentiment strongly decreases the day after the listing. This behaviour may be expected as the attention of investors typically drops right after an important event (Simon, 1955). The sentiment then starts to grow again daily but in this case Alibaba entities are comparable with the one of the peers and the same behaviour happens on the focus around the more recent period (Figure 19-20). No indication can be drawn on the visual link between sentiment and stock price after the zooms, instead some relation could be observed between peaks of activity and stock prices during a normal trading period as is shown in Figure 21.

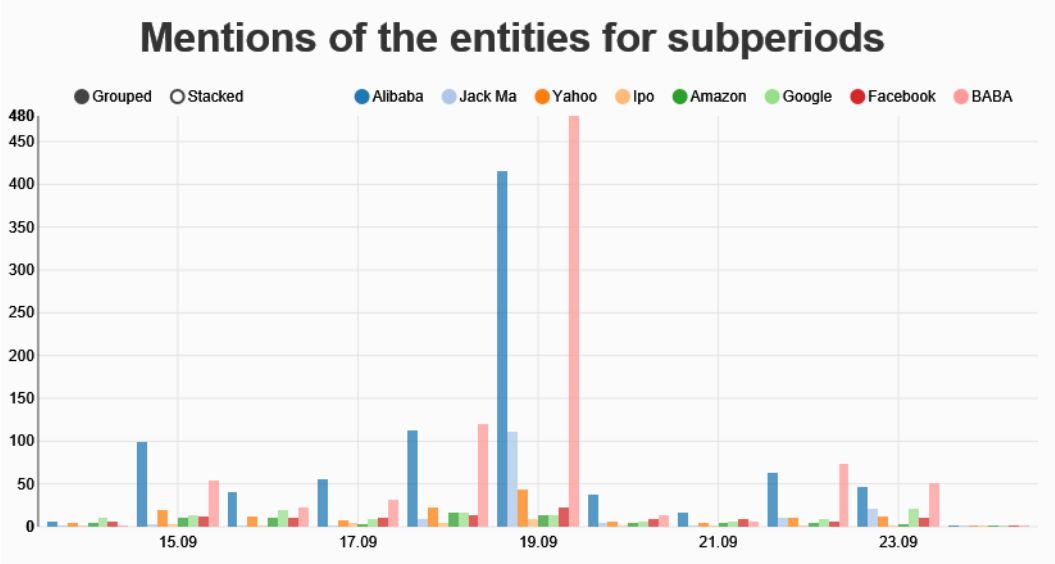


Figure 18 – Alibaba entities aroud the listing date

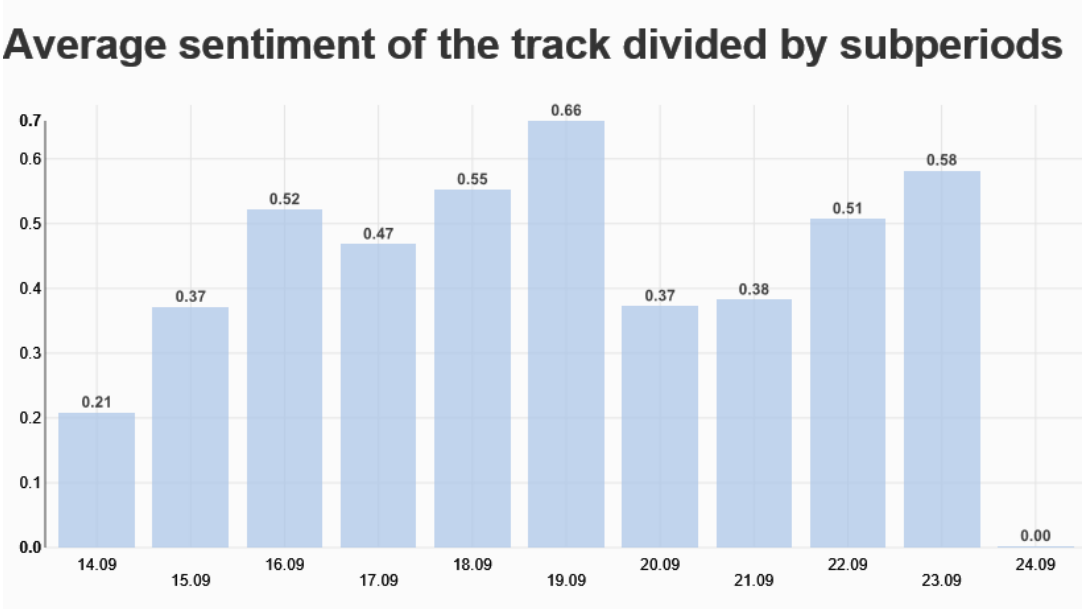


Figure 19 – Alibaba sentiment around the listing date

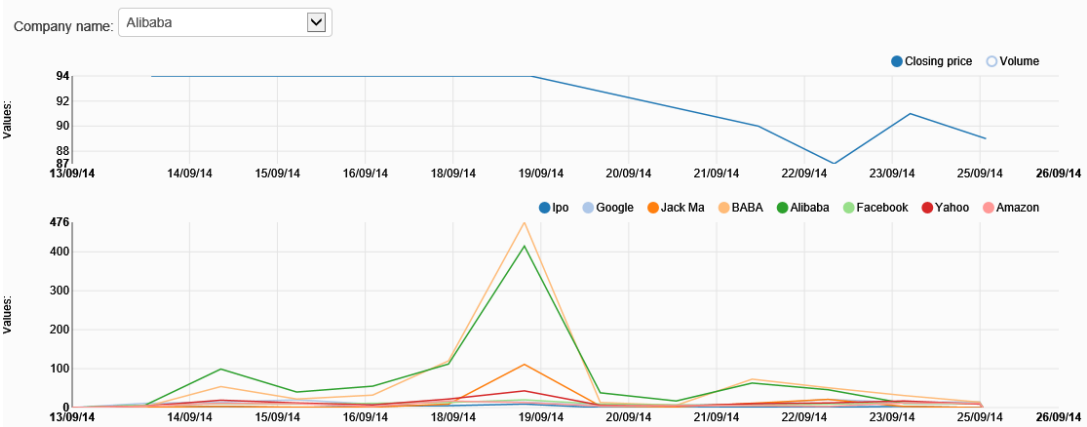


Figure 20 – Alibaba activity and stock price around the listing date

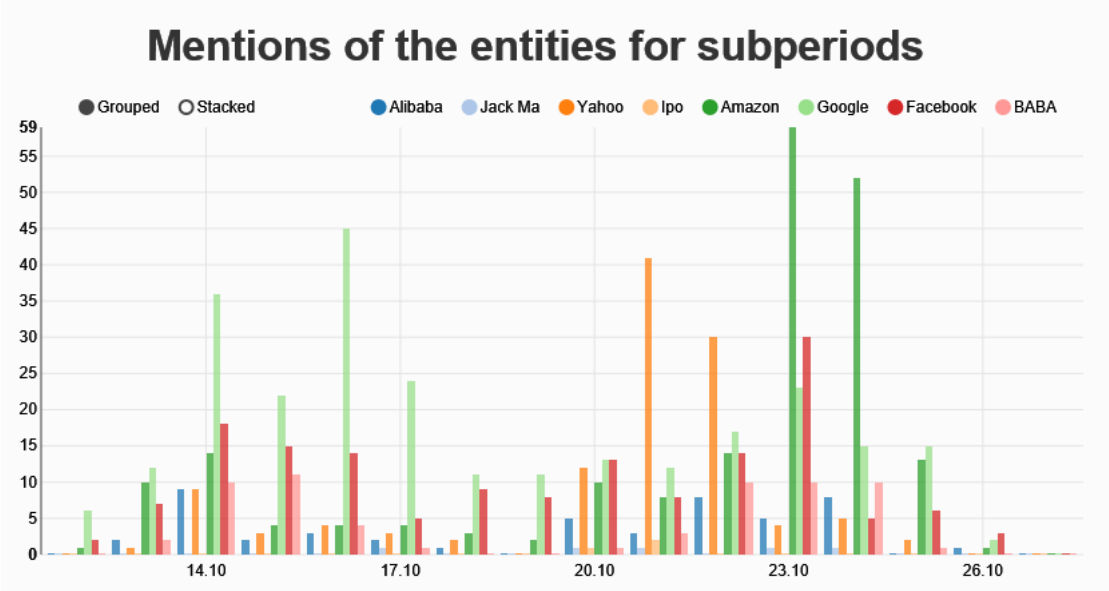


Figure 21 – Alibaba recent entity chart

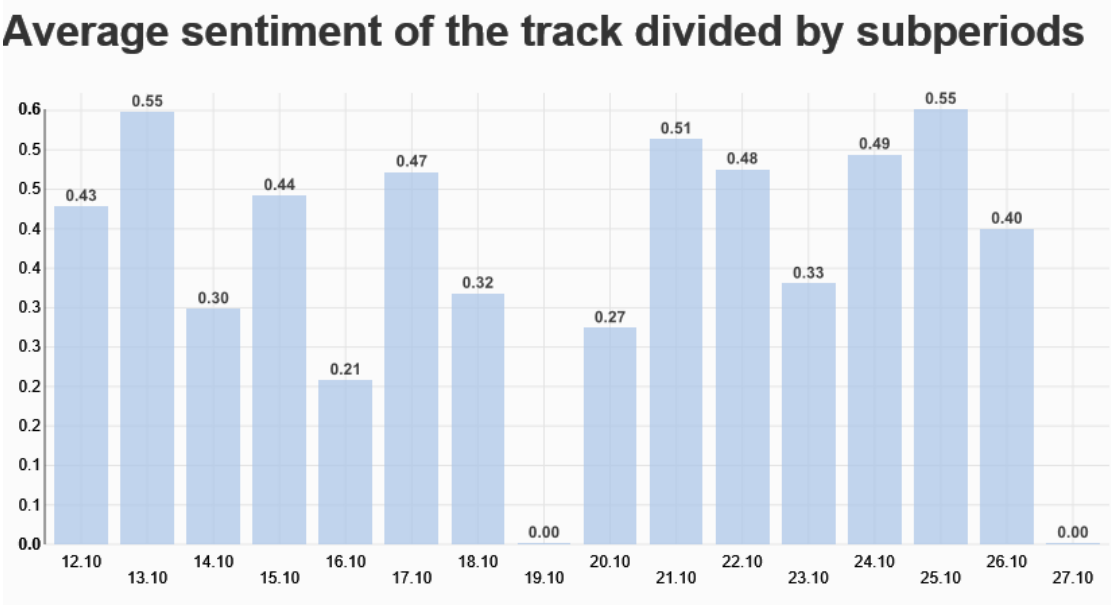


Figure 22 – Alibaba recent sentiment pattern

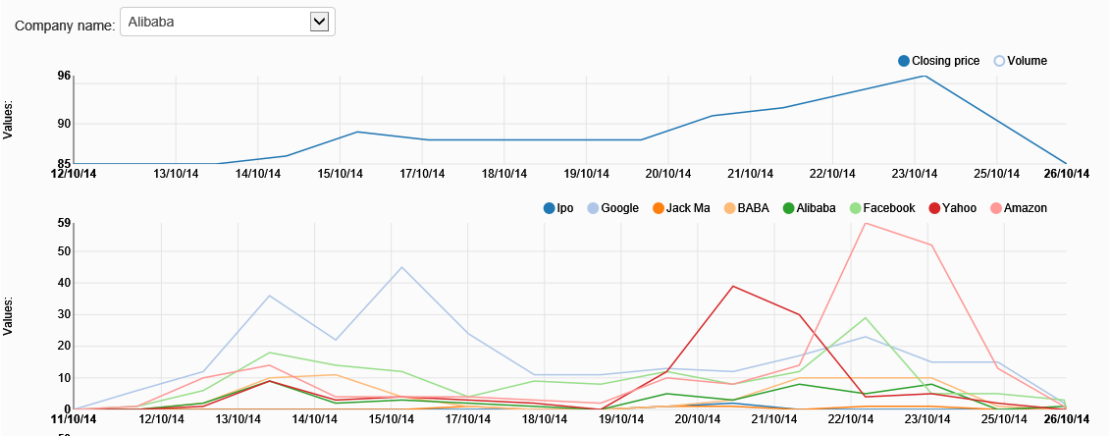


Figure 23 – Alibaba recent activity and stock exchange chart

2.5.3 New product announcement

As a third example we analysed the Apple stocks around the date of the iPhone 6 announcement. The track is shown in Figure 24.

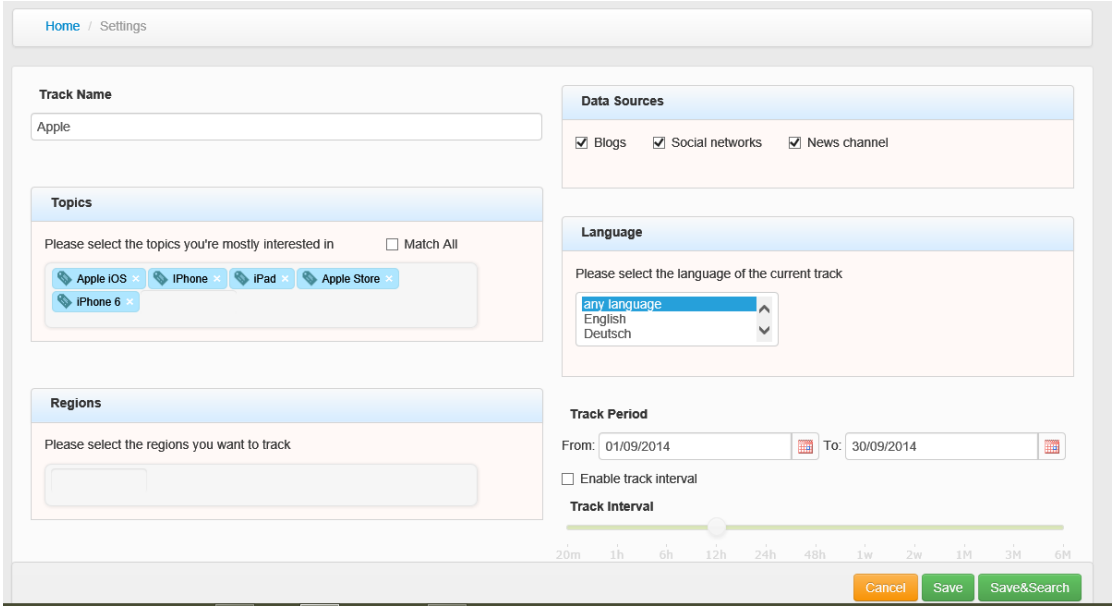


Figure 24 – Apple track

The topics have been chosen by selecting the terms most related to this stock as Figure 25 confirms. It shows the weights of each entity during the selected period and where iPhone 6 is the most relevant.

The chart of activity combined with the stock price shows a peak in the activity around the 8th of September. This corresponds, on one hand to a negative change in sentiment , while on the other hand to a positive reverse in the stock price. This is an interesting configuration to examine in order to decide for a *long* position on the stock. The same also happens from the third to the forth week, even if the peaks and the increase of stock price are less evident

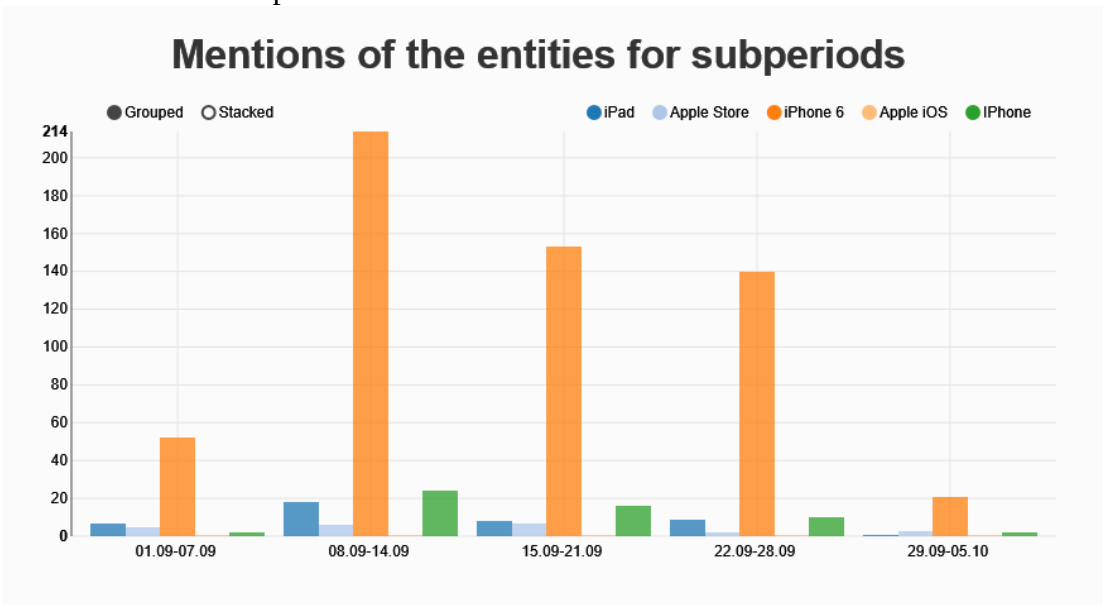


Figure 25 – N° of entity mentions in a period

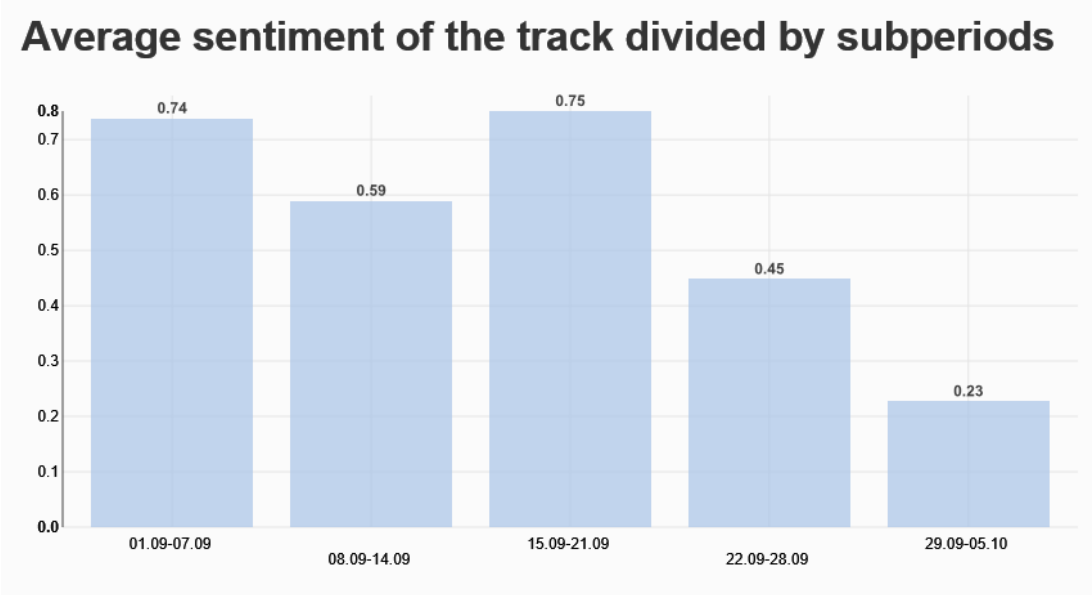


Figure 26 – Sentiment of the track

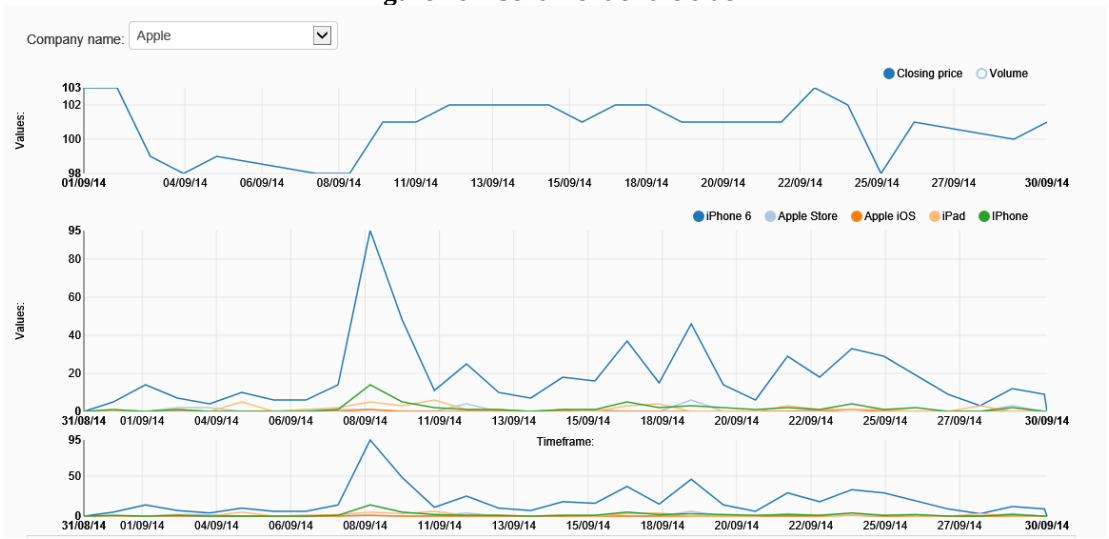


Figure 27. Stock market and twitter data in the time frame

This trend could be explained by the fact that the iPhone 6 announcement was seen by the market as the most significant driver of the stock price for a long time. A reduction of the sentiment at the same time of the peak in activity is considered by some traders/investors as an acceptable behaviour. This may happen (Simon, 1955) [7] because an increasing information is exchanged among the actors when a new product (the iPhone 6) could drive the future stock price. In general in this phase, expectation and activity increase around the new factor and decrease when more details are known and the decision investment has been taken. Another theory, which leads to the same conclusion, consider the investor sentiment as a contrarian (Thiorp, 2004) indicator which foresee a bullish market when the sentiment reach low values and viceversa.

Of course, this is not a unique stategy for a *long* position in the stock. Strategies for short positions could be also detected. As a result, the trader/investor can base his

7

strategy not only considering the market price but also with the information provided by the text mining. The application of the outcome of Trendminer on finance, therefore, must be seen as a decision support tool for traders/investors. Such a tool is based on a flexible architecture which can be tuned according to the user needs.

3 Discussion on test results and possible follow up

From the three examples analyzed the following considerations can be drawn:

Sentiment can be considered as an an additional source of information to drive investment decisions. This can be used togheter with the consolidated tools of technical analysis. However at this stage quantitative approaches for the computation of the sentiment are still missing. It is not yet clear whether the absolute value of a sentiment can be associated to a bullish or to a bearish market. This requires a long tuning phase where it could be useful to refer to the sentiment levels and market returns published weekly by the American Association of Individual Investors (AAII). These levels are the result of surveys and their statistics are summarized in the following table (Thorp, 2004) Note that averages and deviations derive from a long history as the series started in 1987.

Historical Averages & Extremes of AAII Member Sentiment
July 24, 1987, to September 2, 2004

	Average	Median	Standard Deviation	Extremes			
				Highest		Lowest	
	(%)	(%)	(%)	(%)	Date	(%)	Date
Bullish	38.8	38.0	11.3	75.0	Jan. 6, 2000	12.0	Nov. 16, 1990
Neutral	33.4	34.0	8.6	62.0	June 3, 1988	7.9	Dec. 14, 2000 & June 19, 2003
Bearish	27.9	27.0	9.2	67.0	Oct. 19, 1990	6.0	Aug. 21, 1987

The history approach is also the support for evaluations based on the changes of sentiment, because significant changes can be evaluated in terms of standard deviations. Contrarian strategy is one of the most aimed in the market and try to discover situations in which extremely bullish or extremely bullish configuration happen in order to decide to go long or short in investment. Actually, this configurations are investigated through matemathical and statistical tools but a second source value coming from moods and opinions could be of great influence.

Of course, either the visual links and the quantitative readings will lead in any case to subjective valuations for investments, because the amount of which a sentiment must change in order to shift from a bullish configuration to a bearish one is very subjective. This stresses the concept that a decision support system is the most valuable aid for investment choices.

Given this considerations, our examples of market investigations assess some important points:

- The validity of underlying dataset. All the examples, although not similar in the content, have found feedbak and compliance in the system. This means that this dataset must be maintained and improved;

- Some topics must be identified and followed during the time. This helps to find out average values and quantitative changes in sentiment. This topics must be started aroud financil instruments with a consistent “volume” of

moods around them, and therefore the most important indices and, in a second time, the stocks with high volume exchanged;

- An investigation must be done about the forecasting power of the sentiment when quantitayive values will arise. In our examples we used daily prices but we don't know if the sentiment indicator will be more appripriate to forecast weekly or infra-day market values. Or if long daily time series are necessary to analyze the next infra-day behaviour

- An interface specialized on the needs of finance would be more flexible and more indicated.

The previous developments are based, however, on a good architecture that allows since now to make interesting visual valuations.

Bibliography and references

D2.2.2 Multilingual, Ontology-Based IE from Stream Media - v2

D5.3.1 Architecture for distributed text annotation - v2; Real-time Stream Media Processing Platform - v1

D6.1 Requirements analysis on the shaping process of sentiments and opinions on financial decisions

D4.2.2 Multi-paradigm search software; Cross-lingual web UI for trends/sentiments in streaming media - v.2.

N. Godbole, M. Srinivasaiah, and S. Skiena. *Large-scale sentiment analysis for news and blogs*. In Int. Conf. on Weblogs and Social Media (ICWSM 2007), 2007.

Simon, H. A. A behavioral model of rational choice. Quarterly Journal of Economics 1955

W. A. Thorp, CFA *Using investor sentiment as a contrarian indicator*. CI Features September/October 2004

W. Zhang and S. Skiena. *Trading strategies to exploit blog and news sentiment.* In Fourth Int. Conf. on Weblogs and Social Media (ICWSM), 2010

ANNEX 1 – DATASET SAMPLE

Twitter Accounts

Name	URL	Language	Name of selected lists	Amount of users
StockTwits	http://twitter.com/StockTwits	EN	experienced-investors,global-macro-traders,swing-traders,futures-traders,technical-analysis,forex-traders,options-traders,day-traders,personal-finance,finance-bloggers,investing-news,suggested,	417
Dominic Jones	https://twitter.com/irwebreport	EN	public-companies,ir-associations,	399
Trader Club Tirol	https://twitter.com/traderclubtirol	DE	european-news-german,traders	163
FinanzNachrichten.de	https://twitter.com/FN_en	EN/DE		
Ken Odeluga	https://twitter.com/TheSquareMile	EN	denizens-of-thesquaremile	289
Ed Bradford	https://twitter.com/Fullcarry	EN	alpha,bonds	66
Steven Perlberg	https://twitter.com/perlberg	EN	bi-finance-twitter-list	111
Desi Finance	https://twitter.com/desifinance	EN	the-stocktwits-edge	45
Fred Mazo	https://twitter.com/Fred_Mazo	EN	traders-and-trading	58
Pat Allen	https://twitter.com/RockTheBoatMKTG	EN	investmentmngrs-execs,broker-dealers,investment-managers,	148
Market Folly	https://twitter.com/marketfolly	EN	marketfolly-alpha-list	63

Chris Selland	https://twitter.com/cselland	EN	investors	500
Justin Wolfers	https://twitter.com/justinwolfers	EN	economists	11
Naomi Klein	https://twitter.com/NaomiAKlein	EN		
Matina Stevis	https://twitter.com/matinastevis	EN		
Ed Bradford	https://twitter.com/fullcarry	EN		
Chris Adams	https://twitter.com/chrisadamsmkts	EN		
ReformedBroker	http://twitter.com/ReformedBroker	EN		
richi	http://twitter.com/richi	EN		
nichcarlson	http://twitter.com/nichcarlson	EN	ticker	103
jimcramer	http://twitter.com/jimcramer	EN		
stockhaven	http://twitter.com/stockhaven	EN		
businessinsider	http://twitter.com/businessinsider	EN	bi-verticals,bi-strategy	
Convertbond	http://twitter.com/Convertbond	EN		
pensionpartners	http://twitter.com/pensionpartners	EN		
srussolillo	http://twitter.com/srussolillo	EN		
DougKass	http://twitter.com/DougKass	EN		
Hedgeye	http://twitter.com/Hedgeye	EN		
RedDogT3Live	http://twitter.com/RedDogT3Live	EN		
carlquintanilla	http://twitter.com/carlquintanilla	EN	stocks-sectors	26
howardlindzon	http://twitter.com/howardlindzon	EN		
KeithMcCullough	http://twitter.com/KeithMcCullough	EN		
seeitmarket	http://twitter.com/seeitmarket	EN		
optionmonster	http://twitter.com/optionmonster	EN		
Rocco_TheStreet	http://twitter.com/Rocco_TheStreet	EN		
ericjackson	http://twitter.com/ericjackson	EN		
MattRosoff	http://twitter.com/MattRosoff	EN		

TraderPlanet	http://twitter.com/TraderPlanet	EN		
PrivCo	http://twitter.com/PrivCo	EN		
tradefast	http://twitter.com/tradefast	EN		
chessNwine	http://twitter.com/chessNwine	EN		
ritholtz	http://twitter.com/ritholtz	EN		
jonfortt	http://twitter.com/jonfortt	EN		
LaMonicaBuzz	http://twitter.com/LaMonicaBuzz	EN		
allstarcharts	http://twitter.com/allstarcharts	EN		
tlmontana	http://twitter.com/tlmontana	EN		
bySamRo	http://twitter.com/bySamRo	EN		
cnbcfastmoney	http://twitter.com/cnbcfastmoney	EN		
valuewalk	http://twitter.com/valuewalk	EN		
SquawkStreet	http://twitter.com/SquawkStreet	EN		
ValaAfshar	http://twitter.com/ValaAfshar	EN		
andrewnyquist	http://twitter.com/andrewnyquist	EN		
Minyanville	http://twitter.com/Minyanville	EN	trading	124
OptionsProfits	http://twitter.com/OptionsProfits	EN	vips	22
MarketBeat	http://twitter.com/MarketBeat	EN		
ppearlman	http://twitter.com/ppearlman	EN		
BrianSozzi	http://twitter.com/BrianSozzi	EN		
Paul Krugman	https://twitter.com/NYTimeskrugman	EN		
Nouriel Roubini	https://twitter.com/Nouriel	EN		
Modeled Behaviour	https://twitter.com/modeledbehavior	EN		
Milano Finanza	http://twitter.com/MilanoFinanza	IT		
Borsa e Finanza	https://twitter.com/borsafinanza	IT		
News Finanza	https://twitter.com/newsfinanza	IT		

Tito Boeri	https://twitter.com/Tboeri	IT		
Fabrizio Goria	https://twitter.com/fgoria	IT/EN	economics-finance	169
Sole24Ore	https://twitter.com/sole24ore	IT		
Sole24Ore	https://twitter.com/24finanza	IT/EN	trader,economisti,il-sole-24ore-talk24,finanza,	51
Sole24Ore	https://twitter.com/24finmerc	IT		
Sole24Ore	https://twitter.com/MondoMercati	IT/EN		
Sole24Ore	https://twitter.com/Nova24Tec	IT		
Sole24Ore	https://twitter.com/24OreEnglish	EN		
Bloomberg	https://twitter.com/BloombergNews	EN		
BloombergTV	http://twitter.com/BloombergTV	EN	central-banks,luxury-brands,autos,tech-leaders,djia-companies	164
Reuters	https://twitter.com/Reuters	EN	reuters-insider,reuters-business,reuters-health,twitter-ipo	204
FT Alphaville Team	https://twitter.com/ftalpha/lists	EN	bi-best-finance-people	117
The Wall Street Journal	https://twitter.com/WSJ	EN	wsj-money,money-investing,investing,wsj-markets,	123
MarketWatch	http://twitter.com/MarketWatch	EN	trading-deck,marketwatch-staff,wsj-money	134
Financial News	https://twitter.com/eFinancialNews	EN	top-financial-tweeters	115
CNNMoney	https://twitter.com/CNNMoney	EN		
Yahoo! Finance news	https://twitter.com/YahooFinance	EN		
Google Finance news	https://twitter.com/GoogleFinSvcs	EN		
NYSE (stock exchange)	https://twitter.com/NYSEEuronext	EN		
NASDAQ (stock exchange)	https://twitter.com/NASDAQ	EN		

LSE (stock exchange)	https://twitter.com/LSEGplc	EN		
Deutsche Borse (stock exchange)	https://twitter.com/DeutscheBoerse	EN		
Bombay Stock Exchange	https://twitter.com/BSE_News	EN		
CBOE (option exchange)	http://twitter.com/CBOE	EN		
CBOE_TV (option exchange TV)	http://twitter.com/CBOE_TV	EN		

Financial News Providers

Name	Language	URL	RSS feeds
Bloomberg	EN	http://www.bloomberg.com/	
BBC	EN	http://www.bbc.com/news/business/	http://feeds.bbci.co.uk/news/business/rss.xml
Yahoo! Finance news	EN	http://finance.yahoo.com/	http://finance.yahoo.com/news/rssindex/
Yahoo! Finance news	IT	http://it.finance.yahoo.com/notizie	http://it.finance.yahoo.com/notizie/rssindex/
Google News	EN	http://www.google.com/finance/market_news?ei=dIF9UMjVL8zGwAPH7AE	http://news.google.com/news?cf=all&ned=us&hl=en&topic=b&output=rss
Google News	IT		http://news.google.com/news?pz=1&cf=all&ned=it&hl=it&topic=b&output=rss
Financial times	EN	http://www.ft.com	http://www.ft.com/rss/companies
Financial times	EN		http://www.ft.com/rss/markets
Financial times	EN		http://www.ft.com/rss/comment
Financial times	EN		http://www.ft.com/rss/personal-finance
Reuters	EN	http://www.reuters.com/	http://feeds.reuters.com/reuters/companyNews?format=xml
Reuters	EN		http://feeds.reuters.com/reuters/businessNews
Reuters	EN		http://feeds.reuters.com/reuters/hotStocksNews
Reuters	EN		http://feeds.reuters.com/reuters/bondsNews
Reuters	EN		http://feeds.reuters.com/reuters/globalmarketsNews
Reuters	EN		http://feeds.reuters.com/reuters/governmentfilingsNews
The Wall Street Journal	EN	http://uk.wsj.com/home-page	http://online.wsj.com/xml/rss/3_7014.xml

The Wall Street Journal	EN		http://online.wsj.com/xml/rss/3_7086.xml
The Wall Street Journal	EN		http://online.wsj.com/xml/rss/3_7031.xml
The Wall Street Journal	EN		http://online.wsj.com/xml/rss/3_7432.xml
The Wall Street Journal	EN		http://online.wsj.com/xml/rss/3_7104.xml
The Wall Street Journal	EN		http://online.wsj.com/xml/rss/3_7425.xml
The Wall Street Journal	EN		http://online.wsj.com/xml/rss/3_7041.xml
Market Watch	EN	http://www.marketwatch.com/	http://www.marketwatch.com/rss/
Market Watch	EN		http://feeds.marketwatch.com/marketwatch/realtimeheadlines/
Market Watch	EN		http://feeds.marketwatch.com/marketwatch/marketpulse/
Market Watch	EN		http://feeds.marketwatch.com/marketwatch/bulletins
Market Watch	EN		http://feeds.marketwatch.com/marketwatch/pf/
Market Watch	EN		http://feeds.marketwatch.com/marketwatch/StockstoWatch/
Market Watch	EN		http://feeds.marketwatch.com/marketwatch/internet/
Market Watch	EN		http://feeds.marketwatch.com/marketwatch/financial/
Market Watch	EN		http://feeds.marketwatch.com/marketwatch/commentary/
Market Watch	EN		http://feeds.marketwatch.com/marketwatch/newslettersandresearch/
Forbes	EN	http://www.forbes.com/investing/	http://www.forbes.com/real-time/feed2/
Sole24ore	IT	http://www.ilsole24ore.com/	http://www.ilsole24ore.com/st/extra/rss.shtml
Sole24ore	IT		http://feeds.ilsole24ore.com/c/32276/f/643616/index.rss
Sole24ore	IT		http://feeds.ilsole24ore.com/c/32276/f/643609/index.rss
Sole24ore	IT		http://feeds.ilsole24ore.com/c/32276/f/643612/index.rss
Sole24ore	IT		http://feeds.ilsole24ore.com/c/32276/f/566673/index.rss
Sole24ore	IT		http://feeds.ilsole24ore.com/c/32276/f/566674/index.rss

Sole24ore	IT		http://feeds.ilsole24ore.com/c/32276/f/566675/index.rss
Sole24ore	IT		http://feeds.ilsole24ore.com/c/32276/f/643607/index.rss
Corriere della Sera – Economia	IT	http://www.corriere.it/economia/	http://xml.corriereobjects.it/rss/economia.xml
Wall Street Italia	IT	http://www.wallstreetitalia.com/	http://feeds.feedburner.com/wallstreetitalia
Teleborsa	IT	https://www.teleborsa.it/	http://news.teleborsa.it/NewsFeed.ashx
Agi	IT	http://www.agi.it/economia	http://agi.feedsportal.com/c/34339/f/625410/index.rss
Ansa	IT	http://www.ansa.it/	http://www.ansa.it/main/notizie/awnplus/economia/synd/ansait_awnplus_economia_medsynd_Today_Idx.xml
Repubblica	IT	http://www.repubblica.it/economia/	http://www.repubblica.it/rss/economia/rss2.0.xml
Milano Finanza	IT	http://www.milanofinanza.it/	

Blogs

Personal Finance & Trading	Economics
http://www.automaticfinances.com/	http://krugman.blogs.nytimes.com/
http://badmoneyadvice.com/	http://www.economicprincipals.com/
http://www.debtfreeadventure.com/	http://www.economonitor.com/
http://www.myliferoi.com/	http://www.roubiniblog.com/
http://wealthpilgrim.com/	http://www.nakedcapitalism.com/
http://seekingalpha.com/	http://aswathdamodaran.blogspot.it/
http://www.freemoneyfinance.com/	http://ftalphaville.ft.com/
http://blog.stocktwits.com/	http://www.voxeu.org/
http://www.hardrightedge.com/edge/	http://www.ritholtz.com/blog/
http://tradingsim.com/blog/	http://cafehayek.com/
Newspapers &news sources	**Italian blogs**
http://www.ft.com/intl/comment/blogs	http://www.soldionline.it/network-blog/
http://blogs.wsj.com/	http://www.finanzablog.it/
http://www.economist.com/blogs	http://blog.finanza.com/
http://blogs.marketwatch.com/thetell/	http://www.leoniblog.it/
http://blogs.ft.com/economistsforum/	http://www.pilloledibors a.it/
http://economix.blogs.nytimes.com/	http://www.trend-online.com/blog/
http://www.milanofinanza.it/talkback/blogMF/HPblog.asp	http://blog.finanza.com/chi-siamo/
http://www.blog.ilsole24ore.com/	http://www.ilfattoquotidiano.it/blog/lnapoleoni/
http://www.bloomberg.com/now/blog/	http://www.lavoce.info/
https://cs.thomsonreuters.com/blogs	http://www.finanzalive.com/

Companies

Name	Short Name	Official Web Page	RSS	Official reports	Twitter Account	News reports	Source of historycal prices
Apple Inc.	AAPL	http://investor.apple.com/	http://images.apple.com/main/rss/hotnews/hotnews.rss	http://investor.apple.com/stockquote.cfm	https://twitter.com/TheAppleInc	http://finance.yahoo.com/q?s=AAPL	http://finance.yahoo.com/q/hp?s=AAPL+Historical+Prices
Exxon Mobil Corporation	XOM	http://ir.exxonmobil.com/phoenix.zhtml?c=115024&p=irol-irhome	http://www.exxonmobil.com/Corporate/news_rss.aspx	http://ir.exxonmobil.com/phoenix.zhtml?c=115024&p=irol-finlanding	https://twitter.com/exxonmobil	http://finance.yahoo.com/q?s=XOM	http://finance.yahoo.com/q/hp?s=XOM+Historical+Prices
IBM	IBM	http://www.ibm.com/investor/?lnk=fif-inve-usen	http://www.ibm.com/investor/feeds/	http://www.ibm.com/investor/	https://twitter.com/IBM	http://finance.yahoo.com/q?s=IBM	http://finance.yahoo.com/q/hp?s=IBM+Historical+Prices
Microsoft	MSFT	http://www.microsoft.com/investor/default.aspx	http://www.microsoft.com/investor/InvestorServices/RSS/default.aspx	http://www.microsoft.com/investor/default.aspx	https://twitter.com/microsoft	http://finance.yahoo.com/q?s=MSFT	http://finance.yahoo.com/q/hp?s=MSFT+Historical+Prices
Chevron	CVX	http://investor.chevron.com/phoenix.zhtml?c=130102&p=irol-irhome	http://investor.chevron.com/phoenix.zhtml?c=130102&p=rssSubscription	http://www.chevron.com/annualreport/2011/financials/	http://twitter.com/Chevron	http://finance.yahoo.com/q?s=CVX	http://finance.yahoo.com/q/hp?s=CVX+Historical+Prices
BHP Billiton	BHP	http://www.bhpbilliton.com/home/investors/Pages/default.aspx		http://www.bhpbilliton.com/home/investors/reports/Pages/default.aspx	https://twitter.com/stopBHPBilliton	http://finance.yahoo.com/q?s=BHP	http://finance.yahoo.com/q/hp?s=BHP+Historical+Prices

Vodafone Group	VOD.L	http://www.vodafone.com/content/index/investors.html		http://www.vodafone.com/content/index/investors/investor_information/annual_report.html	https://twitter.com/VodafoneIR	http://finance.yahoo.com/q?s=VOD.L	http://finance.yahoo.com/q/hp?s=VOD.L+Historical+Prices
ENI	ENI.MI	http://www.eni.com/it_IT/investor-relations/investor-relations.shtml	http://www.eni.com/it_IT/media/media-news/media-news.shtml	http://www.eni.com/it_IT/investor-relations/bilanci-e-rapporti/bilanci_rapporti.page?type=bil-rap	https://twitter.com/eni	http://finance.yahoo.com/q?s=ENI.MI	http://finance.yahoo.com/q/hp?s=ENI.MI+Historical+Prices
Telecom Group	TIT.MI	http://www.telecomitalia.com/tit/it/investors.html		http://www.telecomitalia.com/tit/it/investors/reports-results.html	https://twitter.com/telecomitaliaTw	http://finance.yahoo.com/q?s=TIT.MI	http://finance.yahoo.com/q/hp?s=TIT.MI+Historical+Prices
Mediaset	MS.MI	http://www.mediaset.it/investor/home_it.shtml	http://www.mediaset.it/investor/documenti/2013/notizie_it.shtml	http://www.mediaset.it/investor/documenti/2013/bilanci_it.shtml	https://twitter.com/Mediasetitalia	http://finance.yahoo.com/q?s=MS.MI	http://finance.yahoo.com/q/hp?s=MS.MI+Historical+Prices
Mondadori	MN.MI	http://www.mondadori.it/Investor-relations	http://www.mondadori.it/Extra/RSS-Feed	http://www.mondadori.it/Investor-relations/Dati-finanziari	https://twitter.com/mondadori	http://finance.yahoo.com/q?s=MN.MI	http://finance.yahoo.com/q/hp?s=MN.MI+Historical+Prices
Mediolanum	MED.MI	https://www.mediolanum.com/ITA/12175.html		https://www.mediolanum.com/ITA/12257_12302.html	https://twitter.com/BancaMediolanum	http://finance.yahoo.com/q?s=MED.MI	http://finance.yahoo.com/q/hp?s=MED.MI+Historical+Prices