Social sentiment indices powered by X-scores

Social Sentiment Indices powered by X-Scores

Brian Davis*, Keith Cortis†, Laurentiu Vasiliu‡, Adamantios Koumpis†, Ross McDermott* and Siegfried Handschuh†
*INSIGHT Centre for Data Analytics, NUI Galway, Ireland
Email: name.surname@insight-centre.org
†Universität Passau, Passau, Germany
Email: name.surname@uni-passau.de
‡Peracton, Dublin, Ireland
Email: laurentiu.vasiliu@peracton.com

***Abstract*—Social Sentiment Indices powered by X-Scores (SSIX) seeks to address the challenge of extracting relevant and valuable economic signals in a cross-lingual fashion from the vast variety of and increasingly influential social media services; such as Twitter, Google+, Facebook, StockTwits and LinkedIn, and in conjunction with the most reliable and authoritative newswires, online newspapers, financial news networks, trade publications and blogs. A statistical framework of qualitative and quantitative parameters called X-Scores will power SSIX. This framework will interpret economically significant sentiment signals that are disseminated in the social ecosystem. Using X-Scores, SSIX will create commercially viable and exploitable social sentiment indices, regardless of language, locale and data format. SSIX and X-Scores will support research and investment decision making for European SMEs, enabling end users to analyse and leverage real-time social media sentiment data in their domain, creating innovative products and services to support revenue growth with focus on increased alpha generation for investment portfolios.**

Keywords*–*social sentiment index, cross-lingual, social networking services, social media analytics, statistical analysis, sentiment analysis, big social and news data, natural language processing, artificial intelligence

I. Introduction

The emerging use of social media data as part of the investment process has seen a rapid increase in uptake in recent years, as by examined Greenfield, GNIP[1]. The lag between Social Media Monitoring and Social Media Analytics - "Brand Analytics" and Finance specific analytics applications has narrowed. The Social Finance Analytics sector has built on the base developed by Brand Analytics and has evolved the ecosystem to focus on investment decision-making. The growth of trading specific social networks like StockTwits has also provided highly valuable structured social data on trading discussions, which was not accessible previously on general social media communities. This new data source has provided a vital pipeline of thoughts, words and decisions between people; connecting and interacting as never before. This collective pulse of conversations and emotional attitudes acts as a gauge of opinions and ideas on every aspect of society. Finance specific social media applications provide asset managers, equity analysts and high frequency traders with the ability to research and evaluate subtle real-time signals, such as sentiment volatility changes, discovery of breaking news and macroeconomic trend analysis. These data streams can be incorporated into current operating models as additional attributes for executing investment decision-making, with a goal to increase alpha and manage risk for a portfolio.

The European research project Social Sentiment Indices powered by X-Scores (SSIX)[2], seeks to assist in this challenge of incorporating relevant and valuable social media sentiment data into investment decision making by enabling X-Scores metrics and SSIX indices to act as valid indicators that will help produce increased growth for European Small and Medium-sized Enterprises (SMEs). SSIX will extract meaningful financial signals in a cross-lingual fashion from a multitude of social network sources, such as Twitter, Google+, Facebook, StockTwits and LinkedIn, and also authoritative news sources, such as Newswires, Bloomberg, Financial Times and CNBC news channel; transforming these signals into clearly quantifiable sentiment metrics and indices regardless of language or locale. Financial services' SMEs can customise SSIX indices enabling them to provide meaningful domain specific insight to design more efficient systems, test trading and investment strategies, better understand risk and volatility behaviour of social sentiment and identifying new investment opportunities.

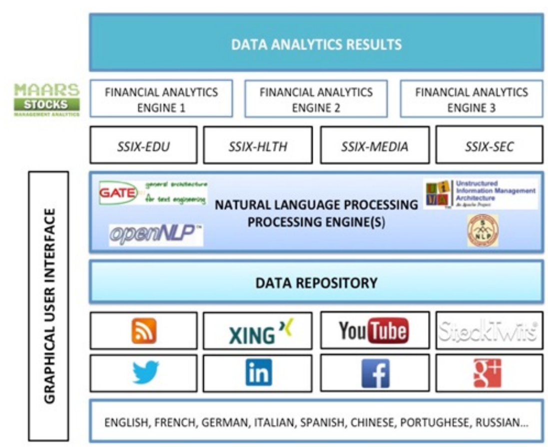

Figure 1. SSIX platform architecture

SMEs can exploit the open source SSIX tools and methodologies to provide financial analytics services or alternatively resell custom SSIX Indices as valuable financial data products to third parties, thus leading to growth and increased revenue

[1]"Social Media in Financial Markets: The Coming of Age...". D. Greenfield, GNIP - https://gnip.com/pages/social-media-and-the-markets-the-coming-of-age-whitepaper/

[2]http://ssix-project.eu/

for SSIX industry partners within the consortium and beyond. Beyond the financial application, the SSIX approach and methodologies can have broader impact across geopolitical and socio-economic domains, generating multifaceted and multi-domain sentiment index data for commercial exploitation. Figure 1 presents the overall architecture design of the SSIX platform.

The objectives of the SSIX project are to:

1) **Develop the "X-Scores" statistical framework**, which will analyse metadata from indexed textual sources to capture the signature of social sentiment, generating a sentiment score. Statistical methods will include regression, covariance and correlation analysis. These X-scores will be used to create the custom SSIX Indices.
2) **Create an open-source template for generating custom SSIX indices** that can be tailor-made with domain specific data parameters for specific analysis objectives, such as Economics, Trading, Investing, Government, Environmental or Risk profiling.
3) **Create a powerful, easy to implement and low latency "X-Scores API"** to distribute the raw sentiment data feed and/or custom SSIX Indices that will allow end users to easily integrate SSIXs sentiment data into their own systems.
4) **Enable end users to do cross-lingual target and aspect oriented sentiment analysis** over any significant social network using user defined dedicated SSIX Index.
5) **Enable various public/private organisations and institutions to create a SSIX Index** and integrate them with their proprietary tools in an easy to use manner.
6) **Explore the domain of SSIX Indices and X-Scores beyond its primary focus of Finance applications.** Research has shown there is a positive correlation between social media sentiment and a financial securities performance, but it is more difficult to measure a broad topic such as, welfare of a region or community. X-Scores will seek to provide metrics which can filter out the noise and provide real quantifiable data, which can give insight via a custom SSIX Index into domains diverse as Education (SSIX-EDU), Media trust (SSIX-MEDIA), Economic sociology (SSIX-ECOSOC), Security (SSIX-SEC) and Health (SSIX-HLTH).
7) **Empower and equip SMEs within the emerging Big Data Financial News sector** to better compete with established larger industry players via technology transfer involving stable, mature and scalable open source semantic and content analysis technologies.
8) **Trigger, nurture and maintain a SSIX and X-Scores commercial ecosystem within and beyond the project lifecycle.**
9) **Pierce language barriers with respect to untapped and siloed multilingual financial sentiment** content by harvesting cross lingual Big Social Media and News Data.

By number crunching news text and social networks data feeds regarding a company, product or various financial products (such as, stocks, funds, exchange-traded funds (ETFs), bonds etc.,) in a mathematical and statistical way, our approach will allow investors and traders to combine SSIX generated indices with their own proprietary tools and methodologies. We envisage empowering the end-user, such as financial data providers, financial, institutions, investment banks, wealth management houses, asset management professionals, online brokers, professional traders and individual investors with the ability to make more informed and better and safer financial decisions. Finally, SSIX could help in identifying unwanted or dangerous trends that could be signalled to financial regulators in advance in order to take appropriate measures, potentially preventing unhealthy and toxic trading behaviour, thereby safeguarding economic growth and prosperity.

II. Related Work

A. Sentiment Analysis on Financial Indices

In [1], Bormann defines several psychological definitions about feelings, in order to explain what might be meant by "market sentiment" in literature on sentiment indices. This study is very relevant to SSIX, since it relates short and long term sentiment indices to two distinct parts of sentiments, namely emotion and mood; and extracts two factors representing investor emotion and mood across all markets in the dataset.

The FIRST project [2] provides sentiment extraction and analysis of market participants from social media networks in near real-time. This is very valuable towards detecting and predicting financial market events. This project is relevant to SSIX, since the tool consists of a decision support model based on Web sentiment as found within textual data extracted from Twitter or blogs, for the financial domain. The relationship between sentiment and trading volume can provide the end-user with important insights about financial market movements. It can also detect financial market abuse e.g., price manipulation of financial instruments from disinformation. Unlike SSIX, only social networking services are used for extracting and analysing sentiment, whereas the developed tool cannot be easily customised to support media sources, target specific companies or select the required language. In this respect, SSIX provides a template methodology and source code to create in a consistent manner the sentiment index for any type of financial product and financial derivatives. Also the outcome is easily integrated within other analytics tools as a data stream with values between 0 and 100 that will define the ranges of that specific sentiment.

Mirowski et al. [3] presents an algorithm for topic modelling, text classification and retrieval from time-stamped documents. It is trained on each stage of its non-linear multi-layer model in order to produce increasingly more compact representations of bags-of-words at a document or paragraph level, hence performing a semantic analysis. This algorithm has been applied to predict the stock market volatility using financial news from Bloomberg. The volatility considered is estimated from daily stock prices of a particular company. On a similar level, in [4] the authors present StockWatcher through a customised, aggregated view of news categorised by different topics. StockWatcher performs sentiment analysis on a particular news messages. Each message can have either a positive, negative or neutral effect on the company. This tool enables the extraction of relevant news items from RSS feeds concerning the NASDAQ-100[3] listed companies. The sentiment of the news messages directly affects a company's respective stock price. SSIX, will extract meaningful financial signals from multilingual heterogeneous (micro-blogging and conventional) content sources and not just news items.

Gloor et al. introduces a novel set of social network analysis based algorithms for mining unstructured information from the Web to identify trends and the people launching them [5]. This work is relevant, since the result of a three-step process produces a "Web buzz index" for a specific concept that allows for an outlook on how the popularity of the concept might develop in the future. A possible application of this system might be for financial regulators who try to identify micro- and macro-trends in financial markets e.g., showing the correlation between fluctuations in the Web buzz index for stock titles and stock prices. Similarly, the Financial Semantic Index estimates the probability that on a particular day, an article in the financial press expresses a positive attitude towards financial markets. This is measured through the emotional tone of the mentioned article [6]. It is relevant to SSIX, since it provides a certain viewpoint of the media environment the market participants consume. In the case of SSIX, it targets to transform the extracted information into multiple clearly quantifiable social financial sentiment indices regardless of language and data format. This will improve the trading and investment accuracy through the combination of various fundamental and technical parameters together with sentiment ones.

B. *Cross-lingual mining of information*

The MONNET project provides a semantics-based solution for integrated information access amongst language barriers [7]. MONNET is relevant for SSIX, since one of its major innovations is the provision of cross-lingual ontology-based information extraction techniques for semantic-level extraction of information for text and (semi) structured data across languages by using multilingual localised ontologies. It provides real-life applications that demonstrate the exploitation potential in several areas, such as financial services. In fact, one of the project's use-cases deals with searching and querying for financial information in the user's language of choice. On the other hand, it focused on cross-lingual domain, thus failed to target other important aspects, e.g., mining the extracted information. SSIX will help identify unwanted/dangerous trends that could be signalled to financial regulators in advance, in order to potentially prevent unhealthy trading behaviour. Hence, SSIX indices can be used as 'early warning' signals for traders, investors and regulator agencies, such as European Central Bank, EU states national banks and rating agencies.

TrendMiner, another European project [8], presents an innovative and portable open-source real-time method for cross-lingual mining and summarisation of large-scale social media streams, such as weblogs, Twitter, Facebook, etc. One high profile case study was a financial decision support (with analysts, traders, regulators and economists). In terms of novelty, a weakly supervised machine learning algorithm is utilised for automatic discovery of new trends and correlations, whereas a cloud-based infrastructure is used for real-time text mining from stream media. This project is relevant to SSIX given that it provides several multilingual ontology-basedsentiment extraction methods.

The main goal of the LIDER project [9] is to create a Linguistic Linked Data (LLD) cloud that is able to support content analytics tasks of unstructured multilingual cross-media content. This will help in providing an ecosystem for a new Linked Open Data based ecosystem of free, interlinked and semantically interoperable language resources (e.g. corpora, dictionaries, etc.) and media resources (e.g. image, video, etc.). It also aims to make an impact on the ease and efficiency with which LLD is exploited in processes related to content analysis with several use cases in multiple industries within the areas of social media, financial services and other multimedia content providers and consumers. One limitation is that LIDER aims to make an impact on the LOD cloud and not to further transform any extracted signals into clearly quantifiable social sentiment indices, as in the case of SSIX. Such indices are targeted to any equities, stock indices or derivatives.

The AnnoMarket project has delivered a cloud-based platform for unstructured data analytics services, in multiple languages [10]. This text annotation market is delivered via annomarket.com and has been in public beta as of April 2014. The services being offered can be adopted and applied for many business applications, e.g., large-volume multi-lingual information management, business intelligence, social media monitoring, customer relations management. It includes several text analytics services that would be of benefit to the SSIX project. Similarly, OpeNER will provide a number of ready to use tools in order to perform some NLP tasks (entity mentions detection and disambiguation, sentiment analysis and opinion detection) that can be freely and easily integrated in the workflow of SMEs [11]. this project aims to have a semi-automatic generation of generic multilingual (initially for the English, French, German, Dutch, Italian and Spanish languages) sentiment lexicons with cultural normalisation and scales through the reuse of existing language resources. SSIX goes beyond text analysis on unstructured data, since an "X-Scores" statistical framework will be implemented to capture the signature of social sentiment from indexed textual sources. These scores will help create custom SSIX Indices that can be tailored for a particular domain depending on specific data parameters. This will provide a meaningful insight to drive trading, investment decisions and strategies, and create new investment opportunities.

III. SSIX Templates

SSIX templates will empower both the public and private sectors to develop innovative disruption-enabling mobile and cloud services and products, to leverage the massive amount of sentiment data that is constantly produced and published on various social media networks within multiple domains such as Finance, Economy, Government, Politics and Health.

The SSIX templates will be able to gauge the actual voiced sentiment from social media conversations, specifically emotional attributes, such as (but not restricted to) optimism and pessimism. These sentiment signals can be analysed to evaluate their influence on real world economic/social/political outcomes and can act as valid indicators. An ideal paradigm that can benefit from the integration of SSIX templates is the field of investment decisions. Traditionally, research on securities, such as stocks, fixed income and foreign exchange

[3]http://www.nasdaq.com/markets/indices/nasdaq-100.aspx

relied on applying a Fundamental and/or Technical Analysis approach to determine the most efficient and lowest risk investment decision for a given amount of expected return. In this scenario, market sentiment is derived from the aggregation of a variety of these two disciplines (Fundamental and Technical analysis), including attributes, such as price action, price history, economic reports/data, market valuation indicators, fund flows, sentiment surveys (e.g. ZEW Indicator of Economic Sentiment - A Leading Indicator for the German Economy[4], commitment of traders report analysis, analysis of open interest from the futures market, seasonal factors and national/world events. As a consequence, it is difficult to get a reliable and easy to interpret measure of a securities sentiment score without using a selection bias and almost impossible to measure a niche sector efficiently; this type of sentiment classification tends to be a lagging indicator to price movement but can act as confirmation.

The growth of social media APIs and the application of news analytics has provided a new method allowing sentiment analysis from a social media perspective to be carried out on financial securities, which has been proven to show a positive correlation to price performance[5]. This data can be analysed to gain a greater understanding of sentiment behaviour and its correlation to price volatility for an individual security/sector or the entire market. By using this new sentiment data source, SSIX can deliver unique sentiment indices using X-Scores (a statistical framework of qualitative and quantitative parameters, such as regression, covariance and correlation analysis), such as the 'Social Sentiment Index for Healthcare' - SSIX_Health or the 'Social Sentiment Index for Technology' - SSIX_Tech, which will show the sentiment levels for their corresponding sectors, quantifying how market participants feel. X-Scores metrics can used in conjunction with industry standard technical parameters to analyse securities, such as Moving Average Convergence-Divergence (MACD), Relative Strength Index (RSI), Moving Averages (MA), Exponential Moving Average (MVA), Pivots Points, etc. SSIX X-Scores will provide real quantifiable data and tools to anticipate volatility and to analyse past performance, which will help develop alternative and more efficient approaches to reduce risk. SSIX can be used to identify trading signals, helping to make more informed investment decisions, resulting in a more efficient use of capital while reducing any associated risk. SMEs will be able to integrate the SSIX framework data into their own models for use in any area of application where sentiment analysis is used.

IV. Big Social and News Data Management

Data retrieved from digital social networking and news sources provides significant data samples to the Natural Language Processing (NLP) component of SSIX. The entire process is developed through the following steps:

- Data download and gathering from different digital platforms (social networks, blogs, news sites, etc.) with different techniques (API usage, CSV download, Web scraping, etc.);
- Data cleaning and filtering to isolate significant information;
- Data processing to produce analysed and enriched data (smart data);
- Data sampling to extract pieces of smart data intended to be used by NLP component.

A. Big Data Challenges

In SSIX, multiple kinds of data are constantly collected, which process is continuous for the duration of the project. The following are types of data in question:

- Public available data from social networks
- Datasets part of the Linking Open Data (LOD) cloud
- LLD Cloud resources
- Public data available from domain-specific SMEs
- Survey data collected from independent events, such as technology summits, conferences, etc., or organised events, such as workshops, focus groups, etc.
- Economic trends outlined by the SSIX framework from analysis/mining of data
- Language Resources either automatically acquired or reused from SentiWordNet[6] and EuroSentiment[7]

Several challenges also arise due to the diverse nature of the gathered data. SSIX is able to deal with the three main challenges coming from the big data field namely, high volume, high velocity and high variety.

- High volume: constant growing of the data repository is managed through adoption of scalable technologies and architectures. The space required for the storage can be easily increased on request, while the technologies used are suitable to manage big quantities of data (e.g. Cassandra, Hadoop).
- High velocity: big stream of data is collected and managed with specific technologies and adequate processing capabilities. The project adopts high-performing servers with possibility to scale the computing power.
- High variety: the gathered data comes from multiple sources. In this case, each data source is treated separately. When required, an unstructured data model is implemented, in order to store information that can vary over time.

B. From Big Data to Smart Data

Figure 2 illustrates the flow that all the data will follow before entering the SSIX platform for further NLP and analysis, which process transforms the data retrieved into smart data.

Each process is explained in more detail as follows:

- BIG DATA: indicates all the information available on different external platform in form of data sources (e.g. social networks, blogs, news sites, etc.)
- DOWNLOADER: the data are gathered from the different data sources using techniques, such as API usage, CSV download and parsing, web pages scraping, etc.

[4] http://www.zew.de/en/publikationen/Konjunkturerwartungen/Konjunkturerwartungen.php3

[5] "Twitter is now a leading indicator of movement (up and down) of specific stocks - we can prove it." Social Market Analytics. https://activetraders.socialmarketanalytics.com/proof

[6] http://sentiwordnet.isti.cnr.it/

[7] http://www.w3.org/community/ontolex/wiki/Eurosentiment

- DATA FILTERING -¿ FILTERED DATA: a first process of noise removal and data processing that will produce a layer of filtered data.
- DATA PROCESSING -¿ SMART DATA: in this phase of the process, all the data will be parsed and transformed into smart data.
- DATA SAMPLING -¿ SAMPLES FOR NATURAL LANGUAGE PROCESSING: the last step will consist in the extraction of significant data samples destined for NLP.

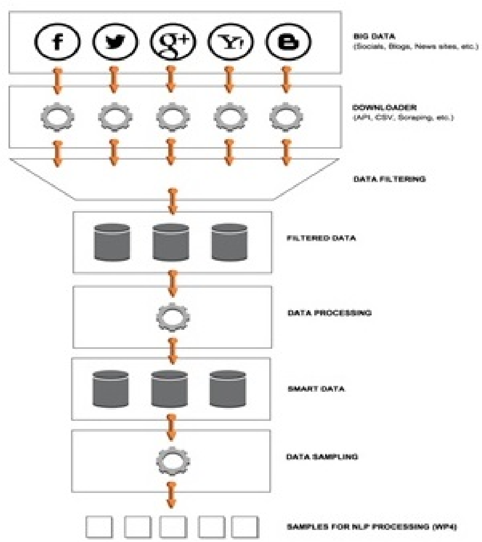

Figure 2. SSIX platform data-flow

All the smart data will be archived into a high performing repository. A cluster of servers will produce significant samples retrieved from the smart data repository that will be taken and streamed to the SSIX platform by and End Point component. The first prototype will use three physical servers to implement the architecture presented in Figure 3.

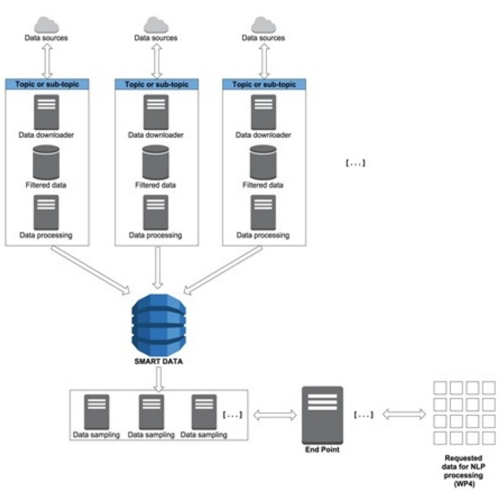

Figure 3. SSIX platform data architecture

The schema defined in Figure 3 illustrates the ideal architecture delegated to retrieve data from the identified data sources, in order to process it and to create data samples for the NLP phase. The business case studies that will be executed in the duration of the project (such as the one discussed in Section VI) will be managed by a cluster of machines that will include: i) a software component that will interface with the different data sources, which will retrieve the data from them; ii) a repository of filtered data; and iii) a software component for data processing.

V. Natural Language Processing Services and Analysis

Analysing trends in social media content results in the process of a very large number of comparably short texts in near real-time. Therefore, the major challenge for the implementation of the NLP pipeline is in the orchestration of the different analysis components in a way that is potentially scalable in a cluster of servers that is able to handle hundreds of messages per second. Special care has to be taken to provide the NLP process as a distributed near real-time computation system that can reliably process unbounded streams of data. SSIX implements this process based on Apache Storm[8]. Apache Storm is a framework that offers the foundations of distributed stream processing. Moreover, SSIX addresses the following major objectives:

- Automatic execution planning of NLP analysis processes: based on the descriptions of existing analysis components, available input and infrastructure, and desired output, SSIX automatically computes an appropriate execution plan (“topology” in Apache Storm);
- Standardised API for analysis components: a common problem in NLP processing is that there are many components for different, but related tasks, but they all implement completely different APIs, making it hard to combine them efficiently in a process. SSIX provides a standardised API and a standardised component description format to simplify integration of existing and additional analysis components.
- Sufficient collection of initial components: a big challenge in building this pipeline is to provide a sufficient collection of initial components so that we can (1) validate our execution model and API, and also provide examples for developers, (2) provide a process for real-time analytics, and (3) integrate with queuing and database technologies provided by SSIX. Figure 4 provides an overview architecture of the NLP pipeline.

A. *Multilingual Language Resource Acquisition and Management*

The multilingual language resource acquisition and management occurs in two phases:

1) Identification and resource of existing language resources for adaption for SSIX business cases (one business case example is discussed in more detail in Section VI), i.e. exploitation of multilingual sentiment and domain specific lexica from European projects, such as EuroSentiment [12] –which provides a shared language resource pool for fostering sentiment analysis from a multilingual, quality and domain coverage perspective– or the adaptation of

[8]http://storm.incubator.apache.org/

LLD resources and carry out any necessary localisation of monolingual resources where target language equivalents are scarce, such as Asian languages.

2) Exploration of unsupervised and/or semi-automatic corpus based methods for acquisition of multilingual lexica to support entity and sentiment analysis tasks.

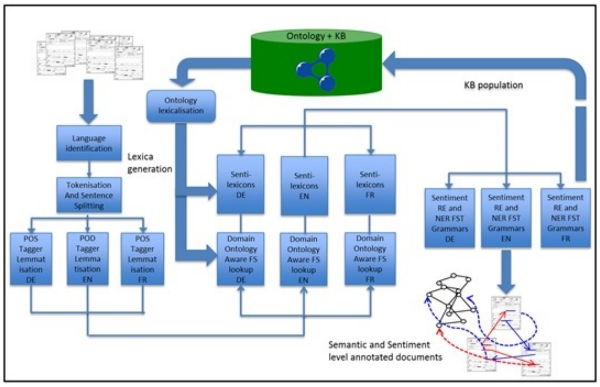

Figure 4. SSIX platform Knowledge-based NLP pipeline

VI. Business Case Study: Investment and Trading

The SSIX sentiment index template will be used to determine social sentiments on stocks and then incorporate them as independent parameters within Peracton's MAARS platform[9], for complex evaluation together with other financial parameters. Performance analysis will be made over historic data and real time, to determine the impact of SSIX indices. This case study will be made of the following 5 phases:

Phase 1: Establish data sources and targets in order to generate unique SSIX indices

Establish data sources and targets in order to generate unique SSIX indices In this phase we will identify the suitable data sources available on various social networks (Twitter, Facebook, LinkedIn, Google+). It is estimated that approximately 6,000+ stocks will be traced individually on US exchanges, such as NASDAQ, AMEX and NYSE.

Phase 2: SSIX indices generation and storage

Once the data sources are established, SSIX engine will be instantiated to generate 6,000+ unique sentiment indices that trace 6,000+ US stocks. Such indices will be uniquely identified, such as SSIX_AAPL, SSIX_YHOO, SSIX_LNKD, SSIX_FB, etc. Once instantiated, the SSIX indices values will be generated first for every day and stored accordingly and then for every minute (if this will be technically feasible).

Phase 3: SSIX indices integration within MAARS

The 6,000+ generated index sentiment values, stored every day (and every minute) will be integrated within MAARS analytics and attached to the existing financial data stocks that are already stored within MAARS cloud.

Phase 4: Trading and Investment with SSIX indices

As sentiment data starts to be updated within MAARS analytics, simulations tests of investing and trading will be performed. There will be trading and investment tests with no SSIX sentiment data (control tests) and then in parallel, same investment and trading tests involving sentiment data.

[9]http://www.peracton.com/maars

Phase 5: Feedback Based upon Phase 4 tests, feedback will be provided with regard to the performance,and changes in results of investment / trading exercise due to using sentiment data.

VII. Conclusions

SSIX seeks to extract and measure meaningful financial sentiment signals in a cross-lingual fashion, from a vast multitude of social network sources, such as Twitter, Facebook, StockTwits, LinkedIn and public media outlets, such as Bloomberg, Financial Times and CNBC. It will generate custom X-Scores powered index for a given sentiment target or aspect i.e. company or financial product. The primary domain is finance although SSIX has scope for Environment, Health, Technology, Geopolitics and beyond. The X-scores will be used by the industrial partners and bundled with their financial analytics, in order to increase the accuracy of their output combined with either end of day financial data or, real time data feeds. SSIX will adapt existing mature, proven and scalable open source text mining tools in order and circumvent language barriers with respect to unexploited multilingual financial sentiment content by harvesting cross lingual Big Social Media and News Data. Semantic Analytics will be employed to generate SSIX indices.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme ICT 2014 - Information and Communications Technologies under grant agreement No. 645425.

References

[1] S.-K. Bormann, "Sentiment indices on financial markets: What do they measure?" Kiel Institute for the World Economy, Economics Discussion Paper 2013-58, 2013. [Online]. Available: http://www.economics-ejournal.org/economics/discussionpapers/2013-58

[2] "FIRST - large scale inFormation extraction and Integration infrastructure for SupporTing financial decision-making," *http://project-first.eu/*, 2013.

[3] P. Mirowski, M. Ranzato, and Y. LeCun, "Dynamic auto-encoders for semantic indexing," in NIPS 2010 Workshop on Deep Learning, Proceedings, 2010. [Online]. Available: http://yann.lecun.com/exdb/publis/pdf/mirowski-nipsdl-10.pdf

[4] A. Micu, L. Mast, V. Milea, F. Frasincar, and U. Kaymak, "Financial news analysis using a semantic web approach," in Semantic Knowledge Management: an Ontology-based Framework, Paolo Ceravolo, Ernesto Damiani, Gianluca Elia, Antonio Zilli (Eds.), November 2008, pp. 311–328.

[5] P. A. Gloor, J. Krauss, S. Nann, K. Fischbach, and D. Schoder, "Web Science 2.0: Identifying Trends through Semantic Social Network Analysis," in International Conference on Computational Science and Engineering. CSE '09., vol. 4, 2009, pp. 215–222. [Online]. Available: http://dx.doi.org/10.1109/CSE.2009.186

[6] Ontology2, "FSI: Financial Semantic Index," *http://financialsentimentindex.com/fsi/*, 2012.

[7] "MONNET - Multilingual Ontologies for Networked Knowledge," *http://cordis.europa.eu/fp7/ict/language-technologies/projectmonnet_en.html*, 2013.

[8] "TrendMiner," *http://www.trendminer-project.eu/*, 2014.

[9] "LIDER: "Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe"," *http://www.liderproject.eu/*, 2015.

[10] "AnnoMarket," *https://annomarket.eu/*, 2014.

[11] "OpeNER," *http://www.opener-project.org/*, 2014.

[12] "EuroSentiment," *http://eurosentiment.eu/*, 2014.