Social Analytics in an Enterprise Context: From Manufacturing to Software Development

Evangelos Arvanitakis, Michael Petychakis, Evmorfia Biliri, Ariadni Michalitsi-Psarrou, Panagiotis Kokkinakos, Fenareti Lampathaki, Dimitrios Askounis
School of Electrical and Computer Engineering,
National Technical University of Athens
Athens, Greece
{agg.arvanitakis, mpetyx ,ebiliri, amichal, pkokkinakos ,flamp, askous}@epu.ntua.gr

Abstract— **Although customers become more and more vocal in expressing their experiences, demands and needs in various social networks, companies of any size typically fail to effectively gain insights from such social data and to eventually catch the market realm. This paper introduces the Anlzer analytics engine that aims at leveraging the "social" data deluge to help companies in their quest for deeper understanding of their products' perceptions as well as of the emerging trends in order to early embed them into their product design phase. The proposed approach brings together polarity detection and trend analysis techniques as presented in the architecture and demonstrated through a simple walkthrough in the Anlzer solution. The Anlzer implementation is by design domain-independent and is being tested in the furniture domain at the moment, yet it brings significant added value to software design and development, as well, through its experimentation playground that may provide indirect feedback on future software features while monitoring the reactions to existing releases.**

Keywords-social data analytics; social media monitoring; trend analysis; opinion mining; natural language processing; sentiment analysis; manufacturing; software development

I. Introduction

With the proliferation of Web 2.0 and the increasing tones of user generated content from the numerous social platforms, new opportunities for engagement and collaboration have materialized. Exploiting this web of information, allowing for what is most commonly referenced as collective intelligence, is of uttermost importance for any enterprise in order to effortlessly realize the needs, satisfaction and buying trends of their customers and to ultimately, accommodate their consumers' wishes. The massive amounts of user generated content from social media open up new perspectives for businesses and manufacturers to understand their consumers' needs and preferences [18]. Nowadays, customers mostly interact with social media in order to express their opinions, complaints or fulfillment for products and services, as for example DELL did when changing their social media policies early on [19]. These data are publicly or easily available through APIs and therefore constitute a great source of information about customer behavior. The accessible information can be harnessed and transformed into knowledge in order to help identify a brand's impact, predict upcoming trends and ultimately enable smarter production and marketing decisions to be made. It is an undeniable reality that all social media platforms gain daily new users on their platforms, which combined with the content of existing users and agents on the web, create an exponential growth of user-generated content. It is clear that businesses can directly benefit [12], [14] from this growth of information, but on the other hand it is extremely difficult for anyone to handle this in an even close to real-time manner [20], [13]. Choosing only one of the most popular social platforms, such as Twitter, Facebook or Instagram, and trying to keep track of all the data produced daily and understand it by running some manual analysis, is characterized as challenging, if not prohibitive [21].

In order to fully utilize the almost infinite amounts of user generated textual data on the various social media, special text processing techniques have been developed, that are usually responsible for processing and classifying the user posts/tweets. Those methods are most commonly known as Natural Language Processing (NLP), and typical techniques include keyword extraction, tokenisation, stemming, Part of Speech (POS) tagging, named entity recognition and other. Micro-blogging data on various social media, for example in Twitter, has developed some unique characteristics that makes it extremely difficult, even impossible in some cases to analyse it, and those need to be evaluated upfront: small size and informal nature, lack of syntax and grammar rules, spelling mistakes, usage of colloquial language, emoticons and excessive usage or total lack of punctuation are only some of the unique features of social media data which can lower the performance of NLP methods if not properly handled.

In the present paper, we elaborate on the PSYMBIOSYS Anlzer Engine which is designed to provide "social" data insights to enterprises regarding their products, services and software. It is envisioned to provide a cloud-based, highly customizable and domain-independent solution, with a user-friendly interface, in order to search efficiently and identify trends and polarity over retrieved statuses, including opinions and perspectives shared across different social networks.

The paper is organized as follows. In section 2 we present the related work as that was identified and studied as part of the background and state of the art we built upon. Section 3 focuses on delivering the requirements and the design principles, while also providing the system architecture. Section 4 concerns the application of the methodology for software development, while Section 5 presents the conclusions of our work and next steps.

II. Background

Social media is the "group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content" [1]. The continuous growth of

978-1-5090-0990-9/16 $31.00 © 2016 IEEE
DOI 10.1109/ARES.2016.99

social media provides unprecedented opportunities to understand involved individuals and organisations and mine behavioural patterns on the produced data [2]. Users are no longer just passive information receivers, but actively contribute in information creation and diffusion. An abundance of web and mobile applications, including social media platforms, blogs and forums, e-shops and other websites for product reviews, enable the envisioned role of Web 2.0 as a medium that leverages user participation, openness, and network effects.

Social Media serve various and diverse purposes, spanning from social networking (e.g. Facebook, LinkedIn) and microblogging (e.g. Twitter) to photo/video sharing (e.g. YouTube, Flickr). The immense potential of leveraging user generated content in social media is evident from the abundance of commercial tools that offer various statistics on such data (e.g. Lexalytics, Klout, LikeAlyzer, Locowise, Pylon, SocialRank etc..). Most of the popular social networks (e.g. Facebook, Twitter, LinkedIn) provide themselves an analytics service, usually accompanied by an API, however external tools and services are being developed to cover more complex and targeted needs (e.g. Bluenod specializes in visualizing Twitter user communities). However, the majority of these tools are primarily addressed to marketers, with brand reputation management and promotional campaign monitoring services, but has little to offer related to product-service engineering and design purposes.

The social media landscape is extremely diversified, in terms of purpose, content and involved stakeholders and social media mining encompasses techniques from various domains, including computer science, data mining, machine learning, social network analysis, network science, sociology and ethnography and statistics. Typical social media mining applications include trend prediction, influencer identification and sentiment analysis.

Among social media mining techniques, this paper mainly focuses on sentiment analysis, commonly referred to as opinion mining as well. Application-oriented approaches in sentiment analysis usually adopt a simplistic view of emotions, i.e. limit their scope into polarity (pleasantness) categorization. Computation of arousal (strength) estimation is also included in some cases, with sentiment often being represented in a Likert scale, e.g.: Very Negative, Negative, Neutral, Positive, Very Positive. However, there are some approaches that maintain the link to the world of emotions. [3] and [4] use the following four categories in emotion recognition (in applications that use speech, facial expressions and other multimodal information): Anger, Sadness, Happiness, Neutral. Researchers in [5] propose a framework to analyze customer emotions as part of the product development lifecycle, especially targeting the early stages of the design process. In their research they cite and describe a variety of emotion measures, from Likert scales to custom emotion dimensions that are linked to specific adjectives and affective verbs used in customer discussions and statements regarding products. This type of language and context dependence points to the need of more P-S specific emotion models. The need to adapt the emotion model to its context reveals one more side of the problem's complex nature: Although more popular and easy, simplified polarity and arousal classification approaches are not always sufficient to understand the sentiments invoked by products and services to the customers.

The realization of the problem's manifold nature has nurtured new holistic approaches that attempt to position themselves closer to emotion theories. As an example, researchers in [6] defined an affective emotion categorization model inspired by Plutchik's wheel of emotions which organizes primary emotions around four dimensions: Pleasantness, Attention, Sensitivity and Aptitude. Approaches following this direction include difficult NLP processes, like co-reference resolution, dependency parsing and other procedures related to the syntactic analysis of the text. Gangemi, Presutti, & Reforgiato Recupero [7] propose a domain independent, completely unsupervised methodology to extract such rich sentiment information (i.e. for whom the sentence is negative/positive and according to whom), employing a rule-based approach that combines knowledge representation with NLP.

The plethora of different approaches to sentiment analysis, including supervised learning, statistical methods, linguistic methods (lexicon based) and rule-based, proves on one hand the interest in the field and on the other hand the number of open issues that remain to be solved.

Sentiment analysis can provide more actionable insights when closely linked to specific products/services or even features, in order to guide engineers and designers towards identifying the truly important issues, as these are casually expressed in online discussions. In order to achieve this, sentiment analysis is often applied together with topic/feature extraction processes. Although both functionalities can be seen as classification problems, they have inherent differences and difficulties to tackle. As an example, although the presence of certain words/phrases inside a text and/or the concepts related to the majority of topic-related words can be relatively trustworthy indicators of its topic, whereas affective words/phrases were shown to be less valuable in predicting the emotional content of a text. Moreover, topic identification is less sensitive to word and sentence positioning inside a given text. [8] Statistical techniques (e.g. Latent Dirichlet Allocation (LDA) [9]) are commonly used to discover implied underlying topics.

III. Social Data analytics in Product-service design

A. Architecture

The PSYmbiosys Anlzer is an open-source tool that allows businesses to gain access to a collection of various insights that derive from social media activity. It enables them to identify sentiments related to their products and brands, detect user trends, monitor discussions around specific topics of interest and obtain valuable information regarding the social media platforms and accounts that influence their market segment the most. These provide an opportunity for businesses to prevent potential damage to their corporate reputation, respond to a product general feedback and adapt to the ever changing market trends.

To achieve that, PSYmbiosys Anlzer utilizes state of the art open-source software all throughout its underlying architecture. The latter, can be segmented into five (5) different layers and engines as demonstrated in Figure 1. These interact and communicate harmoniously with each other in order to constitute the entirety of the PSYmbiosys Anlzer platform. We will now describe the role, responsibility and workflow of each layer individually.

Figure 1: Anlzer Architecture

1) Data Providers

In order to perform the sentiment mining, Anlzer is in need of a social media entry dataset to act upon. This essentially translates to a need of social media platforms from which data can be freely derived. To satisfy this need, we have chosen to support data retrieval from the three (3) major social networking services: Facebook, Twitter and Instagram, based on their popularity and open APIs offering. Keywords and hashtags found inside post activities of selected accounts flow to our platform to contrive a unified pool of social data. To automate the latter, connectors were developed in Python to be latched to the APIs of the services above and fetch data according to the predefined user preferences. It's worth mentioning these social connectors differ from one another due to the API limitations of each provider. Twitter particularly offers a Streaming API that guarantees the monitoring of tweets containing specific terms or being made from specific accounts, which are then automatically forwarded to Anlzer. On the contrary, Facebook and Instagram don't support such an interface, which unfortunately results into a polling necessity implemented via schedulers that run once per day, fetching data from the desired pages and accounts.

As soon as the collected data reach our platform, they are instantly preprocessed before being passed to the Anlzer's storage system. The preprocessing is divided into two (2) separate phases.

The first phase is a collection of sub processes that remove and swap unneeded data. It includes repeated character and inappropriate content removal, replacement of urls and usernames with static placeholders and a global lowercase conversion.

The second phase handles the differentiation of the chosen social services response. Although all responses are in Javascript Object Notation format (JSON), the keywords that describe each particular type of data differ from one another, resulting into a mismatches across different data providers. To resolve this, we preprocess and re-format the data in order to conform a common schema that encapsulates the needs of the platform. The type of information that we chose to store includes basic data such as the text, hashtags, user and mentions, as well as additional metadata such as created time, language, post IDs and accounts/pages. The output of the preprocessing is a normalized collection of entities that are subjectable to the basic query and analyzation mechanisms of the platform. This output is then forwarded to the storage layer as soon as it becomes available.

2) Storage System

In order to make the unstructured, JSON formatted, collection of data prone to fast queries and searches, we must save it in a database. To further align with the API responses and to minimize the search time, we introduced a NoSQL document database to the platform. Out of the many available open source options, we chose Couchbase Server [15] due to its scalability and fast performance, partnered with the flexible data schema that could be easily extended and/or changed for future documents. It's worth mentioning that since our data is stored as JSON objects, the need for allocated memory space increased, but didn't outweigh the query benefits.

3) Processing & Indexing

The benefits of a NoSQL cluster can be easily seen when partnered with Elasticsearch [16]. The latter is an engine that provides real-time distributed indexing and full-text search for any JSON documents belonging to such a cluster. To achieve that we utilized Couchbase's cross datacenter replication (XDCR) to propagate document updates made in Couchbase into Elasticsearch. The search index on the Elasticsearch cluster is always kept up to date with the data in Couchbase but with an access to a rich query domain-specific language (DSL) for both simple and advanced queries. For the index configuration, a mappings file was created defining the ways in which our documents should be handled and mapped to the Search Engine.

We should note that the integration of the Elasticsearch engine resulted into almost twice as fast keyword and hashtags searches, a reduction in time that eventually propagated to the end user.

4) Sentiment Analysis

In order to perform trend and sentiment analysis, multiple natural language processing techniques must first be applied so as to transform the initially unstructured content into a structured form. Indicatively, this data processing step includes tokenization, n-gram creation, emoticon replacement, stop word filtering, stemming, but also more complex processes, such as Part-of-Speech (POS) tagging. All these processes are developed using the Python natural language processing toolkit NLTK [11].

Anlzer offers fine-grained sentiment analysis, classifying documents into the following five emotional categories: angry, sad, neutral, happy and excited. Although previous versions of Anlzer [22] used supervised machine learning for sentiment classification, in the new release a new lexicon-based approach is introduced mainly for two reasons: (a) help the enterprise analys get a better understanding of the real data and the casual terminology of online discussions which will be leveraged to create a mapping to the official domain terminology and (b) to lift the burden of training which was almost impossible to properly achieve for all five emotion categories. Therefore, the affective lexicon SentiWordnet [10] is used to compute sentiment polarity scores for certain extracted keywords. SentiWordnet, however, provides only positive, negative and objectivity level scores for a given word, depending on the various contexts in which it may be encountered. The scores for the first 3 such contexts (called synsets in SentiWordnet) are retrieved, normalized, combined and compared with predefined thresholds in order to perform an initial document categorization. The thresholds are decided upon experimentation on the given domain terminology and real datasets.

Apart from the sentiment insights, Anlzer also offers metrics that can help identify product and service trends, as well as correlations among them based on the frequency of common appearances in online discussions.

5) *Interface Level*

The user is able to interact with the platform via a web application that implements a Client – Server distributed structure model. For better decoupling and further optimizations in terms of speed, this application is completely detached from the other engines, but still has the ability to connect with them via HTTP requests. The server is based on Python's Django framework, which utilizes the MVC architectural pattern to structure the application. It communicates with a relational PostgreSQL database where the users, projects and reports are stored and through dedicated APIs it has the capability of connecting with the other engines and layers of the platform. The client is implemented as a custom web interface written entirely in HTML5, CSS3 and JavaScript. Through the latter, the user can create projects, configure reports and view the sentiment analysis results. For the analytics that are available to the end user we utilize Kibana 3 [17], a software platform that is seamlessly integrated with ElasticSearch and offers a wide variety of views and interactions with statistical data. In addition, we also extended our arsenal of visualization libraries by adding D3, one of the most known open-source javascript libraries that simplifies the creation of powerful analytical data. The latter provided an easy to user API, allowing the development of custom, robust, and optimized web graphs supported by all web browsers. All analytics and statistical data are available to the user through the aforementioned web interface, enabling him to take advantage of the optimized queries and interactive visualization.

B. *Approach and Considerations*

To optimize development and reduce refactoring cycles to a bare minimum, a process of platform requirement identification was initiated. This procedure lasted five (5) months, during which sessions with business representatives took place to ensure that the approach was in line with their expectations and needs. Upon careful examination, detailed mockups were drawn, specific scenarios were agreed and finally the functional and non-functional requirements were properly specified.

From the latter, the most challenging was the achievement of domain independency in terms of context. As opposed to most academic work in sentiment analysis, our engine shouldn't be focused on specific datasets, in order to be agile enough to keep its accuracy rate across different domains. Due to the fact that such a cross-domain knowledge isn't available in a general case, the system requires iterative optimizations in order to enhance its overall performance.

Generally, all processes need to be implemented by the company's personnel, most of whom have no programming knowledge. This is why the PSYMBIOSYS Anlzer platform should provide an easy-to-use query interface, removing the need for any development knowledge skills. The goal is to give the user the ability to perform queries with multiple levels of filtering across different fields and documents, without him being aware of the underlying query construction complexity.

Delving into the platform interaction, the need for flexibility becomes apparent. Different companies, different departments under the same company or even different people in the same department will use Anlzer in different ways. This is why the report creation and data visualization mechanisms have been designed in such a way, as to allow maximum personalization. In addition, user type diversification has been also implemented, restricting specific accounts from performing certain actions, making the platform even more manageable. Currently, Anlzer supports company accounts and personal accounts. Company accounts are created through a normal registration process and have administrative control over the projects and reports of the company, while personal accounts are created from within a company account and are restricted to a basic set of actions without any administrative options available to them.

Since the platform can handle a variety of micro-tuned querying, some of which is being constantly repeated on a weekly or monthly basis, the need for availability of historical data became apparent. Browsing through the data and interacting with the corresponding charts needs to be also enhanced by the ability to store query results, keep them for future reference or even compare them with one another without the need of repeating the process multiple times.

Scalability is the last but most important requirements of the Anlzer tool. Due to the nature of social media platforms, important news can go viral in a matter of hours with huge amounts of data produced in a short time span. It is thus important that the technologies used are scalable and can support real time retrieval, storage and processing of data – when this capability is offered by the external APIs - without crippling the engine or altering the experience of the user.

C. *Demonstration*

To interact with the engine a user must first complete a registration process which differs based on the type of account that is being created. As described, company accounts follow the common registration process offered through the Anlzer interface and they act as administrators inside the company. Additional personal accounts can be created through the company account's admin panel that have restricted permissions. This is done to ensure that multiple users can log in to the system but only a single account per company can alter the fundamental content. A company account can create as many personal accounts as it requires and has full authority over them, deleting them as it pleases. The admin panel of the company's account can be seen in Figure 2. Since personal accounts have a subset of the company accounts' permissions available to them, we will be describing the latter as our typical user.

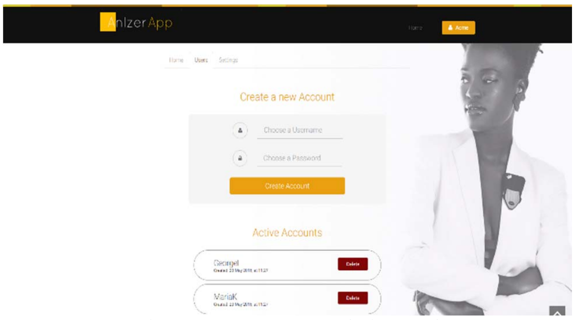


Figure 2: Company's Admin Panel

After a user is logs in, he is redirected to his home page from where he can either initiate a new project or view his existing projects. There is also a more analytical view of the currently existing projects of the account as tiles inside a grid, which when hovered display a description of the specific project (Figure 3). From this page the user can either create a new project or click an existing one in order to be redirected to the projects main page.

Creating a new project is a 2-step procedure. The first step is to define the title and description of the project, along with the keywords, hashtags and social media platforms that are to be tracked. The second step is the initial training of the project. Although this step is optional, it is of paramount importance that Anlzer gets trained to respond well to the user's specific domain. The project creation process can be seen in figure 4.

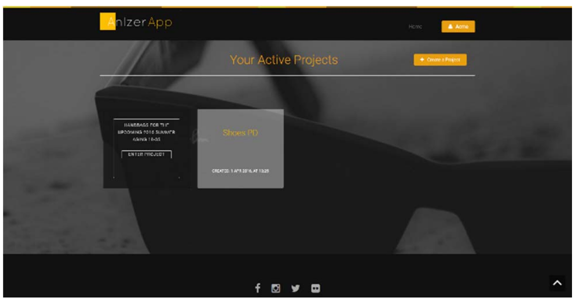


Figure 3: User's list of projects

After a project is created or when the user clicks a project, he/she gets transported to the project's main page.

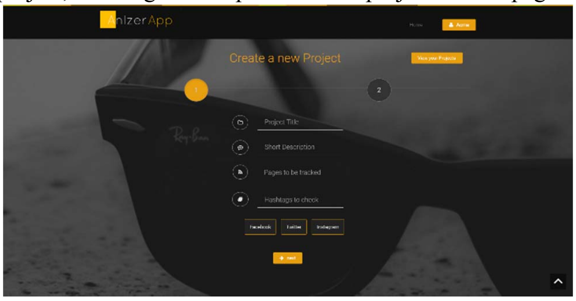


Figure 4: Project Creation process

From there he can view the list of existing project reports, add comments to the project, re-train it or even update it. A report could be perceived as a collection of filters bound to the dataset that is linked with the project. This dataset gets initialized during the project's creation process and is derived from the social providers, pages and hashtags that the user selects for the project. So in other words, a project defines the data pool that will be used, while a report adds a form of filtering to the original pool. For example, a project might refer to clothing and a report could isolate specific dress types, colors or even dates from the original dataset. The sentiment results, analytics and overall statistical information are always connected with reports. This means that even when the user doesn't want to add any filters to the data fetched from the social platforms he/she must still create a report and simply select no filters. The user can create an infinite number of reports, and the creation process is similar to a typical project creation with the difference that the user, instead of adding keywords, hashtags and providers, he filters out the project's corresponding data during a specified period of time. Clicking a specific report from the list, would direct him/her to the report's view where the engine results are available.

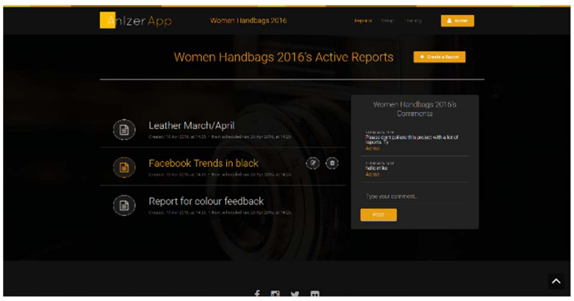
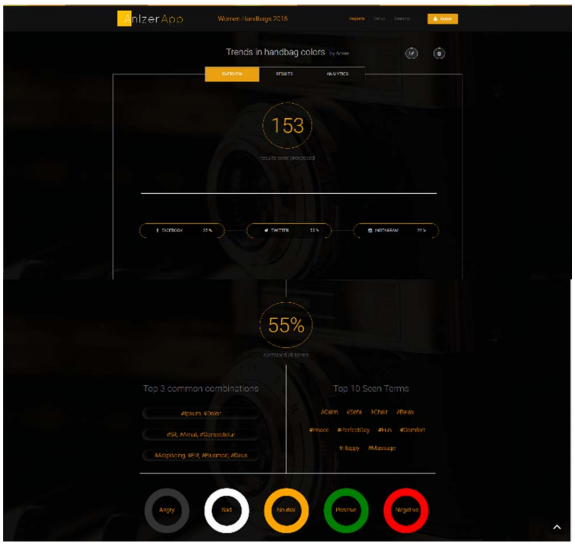


Figure 5: Project page with its reports

The report's page is divided into three (3) tabs. The overview, the results and the analytics. In the first tab the user can see aggregated statistical data that act as a quick look into the platform's output. In the second tab the user can get to see a document-to-document analysis, showing the sentiment assigned to each entry inside the report's filtered dataset. If the user feels that a specific document was assigned an incorrect sentiment, he can click on it to correct it, thus initiating an engine re-training process. In the third tab, a look into advanced charts and analytics are provided to the user, helping him get a more in-depth understanding of the analyzed dataset. These are viewable in figures 6, 7 and 8.



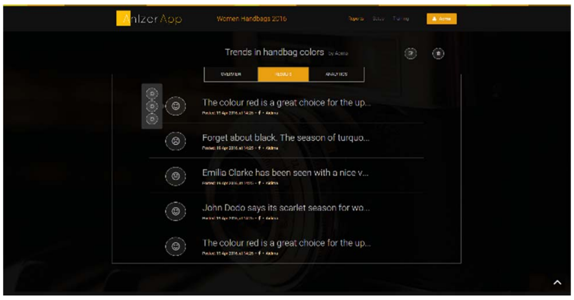


Figure 7: Report's Sentiment Analysis Results

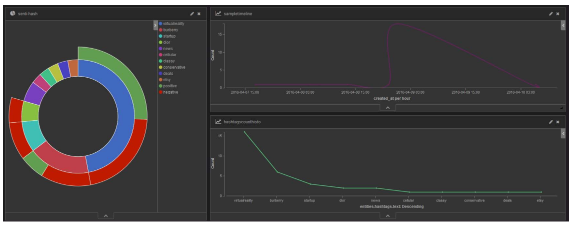


Figure 8: A Glimpse of Report Statistics

IV. Applying Anlzer in Software Design

In the era of cloud & mobile applications where the feedback is immediate and directly shared onto social media creating an unprecedented network effect, software development is taking new impetus. It is instrumental for a company to closely follow and monitor the public's opinion on their product in order to both fix potential issues and get inspired for new products and features. In this context, the functionalities of Anlzer (as already presented in detail in the previous sections) can be ported in the software / IT domain.

Anlzer attaches added value to the agile development process of continuous iterations for new concepts that ideally would need to be integrated within the software product. Following the steps presented, a product manager can create a project to monitor various potential software features and validate the ones that the community values as the most important ones. This way, tracking and receiving the public perception is vastly simplified for the software on social media and thus is it easier to respond on time to criticism and the various comments in order to avoid damage of the company branding and its public image perception.

V. Conclusions and Next Steps

In this paper, the Anlzer analytics engine that extracts user-generated content from selected social networks in order to make it valuable and actionable by manufacturers and enterprises, in general, was outlined. The social data experimentation playground it offers through an inherently simple, generic and customizable approach is listed among its added-value features. Currently, the Anlzer engine is applied and evaluated in the furniture domain in the contact of the PSYMBIOSYS Factories of the Future Project, yet it has the credentials for expansion to any industry beyond manufacturing, i.e. software engineering.

The next steps we intend to pursue in order to advance our work in Anlzer include:

- Effective, gradual incorporation of additional data sources that are relevant for an enterprise and convey sentiment (e.g. customer satisfaction surveys, blogs, enterprise social networks).
- Leveraging the results from the lexicon-based sentiment analysis as seeds for the application of semi-supervised machine learning towards more efficient sentiment classification.
- Detection of "influencers" per domain and social network, building on big data analysis technologies, in order to enrich the underlying trend analysis and provide more credible results.
- Multi-lingual features allowing for analysis of user-generated content in other languages beyond English.

Acknowledgment

This work has been funded by the European Commission through the Horizon 2020 FoF Project PSYMBIOSYS: Product-Service sYMBIOtic SYStems (Grant No. 636804).

The authors wish to acknowledge the Commission and all the PSYMBIOSYS project partners for their contribution.

References

[1] A. M. Kaplan, M. Haenlein. "Users of the world, unite! The challenges and opportunities of Social Media." Business horizons 53, no. 1 (2010): 59-68.

[2] Zafarani, Reza, Mohammad Ali Abbasi, and Huan Liu. Social media mining: an introduction. Cambridge University Press, 2014.

[3] Wu, Chung-Hsien, and Wei-Bin Liang. "Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels." Affective Computing, IEEE Transactions on 2, no. 1 (2011): 10-21.

[4] Busso, Carlos, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. "Analysis of emotion recognition using facial expressions, speech and multimodal information." In Proceedings of the 6th international conference on Multimodal interfaces, pp. 205-211. ACM, 2004.

[5] Khalid, Halimahtun M., and Martin G. Helander. "Customer emotional needs in product design." Concurrent Engineering 14, no. 3 (2006): 197-206.

[6] E. Cambria, A. Livingstone, A. Hussain. "The hourglass of emotions." In Cognitive behavioural systems, pp. 144-157. Springer Berlin Heidelberg, 2012.

[7] Gangemi, Aldo, Valentina Presutti, and Diego Reforgiato Recupero. "Frame-based detection of opinion holders and topics: a model and a tool." Computational Intelligence Magazine, IEEE 9, no. 1 (2014): 20-30.

[8] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and trends in information retrieval 2, no. 1-2 (2008): 1-135.

[9] Latent Dirichlet Allocation. (2015, 10 11). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation

[10] SentiWordnet. http://sentiwordnet.isti.cnr.it/

[11] NLTK 3.0 Documentation http://www.nltk.org/

[12] Kietzmann, Jan H., et al. "Social media? Get serious! Understanding the functional building blocks of social media." *Business horizons* 54.3 (2011): 241-251.

[13] D. Maynard, K. Bontcheva, D. Rout. "Challenges in developing opinion mining tools for social media." *Proceedings of@ NLP can u tag# usergeneratedcontent* (2012).

[14] M. Ogneva "How Companies Can Use Sentiment Analysis to Improve Their Business". Mashable (2010), http://mashable.com/2010/04/19/sentiment-analysis/

[15] Couchbase Server http://www.couchbase.com/couchbase-server/overview

[16] ElasticSearch http://www.elasticsearch.org/

[17] Kibana 3 http://www.elasticsearch.org/overview/kibana/

[18] A. M. Kaplan, M. Haenlein. Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons. Volume 53, Issue 1, January–February 2010, Pages 59–68

[19] E. Qualman. Socialnomics. How social media transform the way we live and do business. Second Edition.

[20] N. Michaelidoua, N. Siamagkab, G.Christodoulides. Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands. Industrial Marketing Management. Volume 40, Issue 7, October 2011, Pages 1153–1159

[21] A. Holzinger. Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What Is the Benefit of Bringing Those Two Fields to Work Together? Availability, Reliability, and Security in Information Systems and HCI. Lecture Notes in Computer Science Volume 8127, 2013, pp 319-328

[22] E. Biliri, M. Petychakis, I. Alvertis, F. Lampathaki, S. Koussouris, D. Askounis. "Infusing social data analytics into future internet applications for manufacturing." In Computer Systems and Applications (AICCSA), 2014 IEEE/ACS 11th International Conference on, pp. 515-522. IEEE, 2014.