Jaakko Ojalehto

Building Online Reputation Management Tool

The purpose of this graduation study is to document the problems faced, decision taken and solutions found upon creating an online reputation management tool called Prism for an Internet marketing agency NetBooster. The tool replaces an older tool used by the company and it is completely redesigned and rewritten.

Online reputation management tools are automated applications that search the Internet for articles and discussions relevant to a brand, person or company. The articles and discussion are saved to a database and analysed further. The analysis includes qualitative and quantitative analysis by using the actual text of the articles as well as the available meta data.

Technologies and tools used in the project include PHP, Yii Framework, jQuery and Cuty-Capt. Many external services are introduced including tools provided by search engine companies such as Google, text analytics application programming interfaces such as AlchemyAPI, website categorization service SimilarGroup and Internet audience monitoring service Alexa Internet.

Social Media analysis is discussed as social media reputation management is very different from other online reputation management. Multiple social media platforms such as Facebook and Twitter are discussed as well as the metrics used to analyse social media.

The user interface of Prism is introduced and explained with multiple illustrations. Also the technologies used for creating interactive charts are explained and evaluated.

This graduation study should help to overcome some of the problems faced when building an online reputation management system, using text analytics services or creating a web application that includes charts.

Preface

I always wanted a challenging software development project with real business value as my final thesis project and I am forever grateful to all the people who helped me in getting the project I wished for and pushing through it.

Especially I would like to thank Niko Juntunen, Nils Carlsson and Pascal Dioh for all the work you have done that made it possible for me to work for this project in Paris.

I would also like to thank Guillame Balloy, Christophe Morvant, Emmanuel Arendarczyk, Laura Mortier, Pascal Dioh, Thomas Bernot, Yann Gabay and other employees of Net-Booster who have contributed in the Prism project.

I would also like to thank Simo Silander and Jonita Martelius for all the support and help I have received. Special thanks for allowing me to work on the thesis from abroad.

In Paris 30 April 2012

Jaakko Ojalehto

Abbreviations and Acronyms

.NET	Multi-language software framework for Windows Platform developed by Microsoft
ASP.NET	Web application framework which is part of .NET software framework
C#	Modern strongly typed object oriented programming language developed by Microsoft
Cron	Time based job scheduler in Unix-like systems
CSS	Cascading Style Sheets is a language used to describe the look and formatting of a HTML or other markup document
CSS Framework	Prepared CSS library that inludes solutions to the most commons problems and tasks
cURL	Command line tool for transferring data by various protocols such as HTTP and HTTPS
CutyCapt	Command line tool to create an image or PDF of a website
DOM	Document Object Model is a language independent presentation of a HTML or other markup document
Elinks	Text based www browser
GD	Multi-language graphics library for creating and editing images programmatically
HTML	HyperText Markup Language is a markup language used for creating web pages
HTTP	HyperText Transfer Protocol an an application protocol enabling the data transfer in the Internet
Java	Object oriented programming language developed by Oracle
JavaScript	Prototype based scripting language mainly used in websites and is run by the client browser
JSON	JavaScript Object Notation is a human readable standard for representing data
MVC	Model-View-Controller is an architectural pattern used in software engineering
ORM	Object Relational Mapping enables database rows to be turned into objects and the other way around
PDF	Portable Document Format is a file format for displaying documents containing text and images

PHP	Hypertext Preprocessor is a server-side scripting language mainly used in web development
PNG	Portable Networks Graphics is a bitmapped image format which offers lossless compression
Qt	Cross-platform application framework used for application development
RSS	Family of web feed formats used to publish frequently updated entries
Spring Framework	Application framework for Java mainly used for web development
SQL	Structured Query Language is a programming language used for managing relational database systems
Struts	Open-source java web application framework
Web application Framework	A group of code libraries to implement the basic functionalities of a web application such as ORM
WebKit	Layout engine designed for web browsers to render web pages
XHTML	eXtensible HyperText Markup Language is a subset of XML markup language that extends HTML
XML	Extensible Markup Language is a human readable standard for representing data

Contents

Appendices

Appendix 1. Prism version 2 functional requirements

1   Introduction

The purpose of this graduation study is to document the problems faced, decisions taken and solutions found upon creating an online reputation management tool that replaces an older online reputation management tool called Prism developed and used by NetBooster Group. In addition this graduation study analyses the purpose and use of reputation management tools and introduces the online reputation management tool market and the key differences between the competing products.

Online reputation management refers to the activity of monitoring the reputation of a person, brand or business in the Internet and addressing the issues causing negative reputation. This is done by gathering information on what is being written and talked about the client in the Internet, identifying the problems causing the negative reputation and acting appropriately to eliminate the cause of the bad reputation. Many companies, persons and brands are interested in this kind of information because it useful for them in various tasks related to marketing, brand management and customer service. (Wikipedia 2011i)

Online media monitoring is the part of online reputation management that focuses on searching and analysing the media output provided by various online publications. The focus is usually on big established news websites but sometimes more informal media such as blogs and forums are included as well. (Cision)

The purpose of an online reputation management tool is to provide information about the client's online reputation fast, reliably, accurately in easy to read format. Online reputation management tools accomplish this by trying to find all the articles and discussions, often called mentions, in the Internet and saving them into a database. After finding the relevant mentions they are analysed to provide aggregated data of the client's online reputation and finally the aggregated data as well as the found mentions are shown to the user of the tool as charts, lists, images and tables.

The problems introduced in this study were faced during the design or implementation of the tool. Most of the answers and solutions were found from various Internet

sources including studies, reports and company websites. In addition to that some of the solutions were found independently by trial and error.

The second chapter focuses on the other online reputation management tools on the market and the features that make them different from each other. Online reputation management tool is an important product for NetBooster but the older version of Prism is lacking in functionality, ease of use and reliability. The new version of Prism aims to improve in all of these areas. The objectives of the project are discussed further in the third chapter as well as the history of Prism and its position on the market in relation to the other tools.

The fourth chapter focuses on the architectural and technological decisions of the new version of Prism. The chapter covers the programming languages and frameworks evaluated and chosen for the production of the tool. The fifth chapter focuses on finding of the mentions from the internet. Different solutions are evaluated and few common problems are introduced and solved. The sixth chapter focuses on analysing the found mentions. Various metrics and their calculation problems are discussed. The seventh chapter focuses on the measurement of the social media. This chapter includes discussion of the different social media networks as well as the metrics used to measure the reputation of a client in the social media platforms.

Chapter 8 introduces the user interface of Prism and evaluates various user interface related libraries and technologies. The design decisions of various views of the tool are discussed and screenshots are provided.

2 Online Reputation Management Tools

Unlike broadcast and press media monitoring which is mostly conducted by searching and recording articles manually, online monitoring can be automatized to a great extent. Even though automatizing media monitoring is possible it requires special software and it is usually more effective to buy this instead of creating one.

Online reputation management tools are used by multiple different groups including celebrities, politicians, all kinds of companies from startups to big global groups, gov-

ernments and associations. Departments within a company including public relations, marketing and executives use online reputation management tools for different purposes. Because of the variety in users there is a wide array of products available for this purpose. The range of online reputation management tools ranges from simple free ad-hoc tools to complete software solutions. According to Mashable the cost of the paid services ranges from $ 1 to over $ 100 000. (Mashable 2008b)

2.1   Free Online Reputation Management Tools

According to Mashable the best free tools for online reputation management are services such as Google Alerts, Yahoo! Pipes, Technorati Blog Search, Twitter Search and other specialized search tools (Mashable 2008a). These tools offer an access to huge amounts of data but the drawback is all the work one has to do to find the relevant mentions and discussions and to analyse them. Paid tools offer a solution to the problem by bringing all data from multiple sources such as search engines and social media networks to one user interface and analysing them automatically.
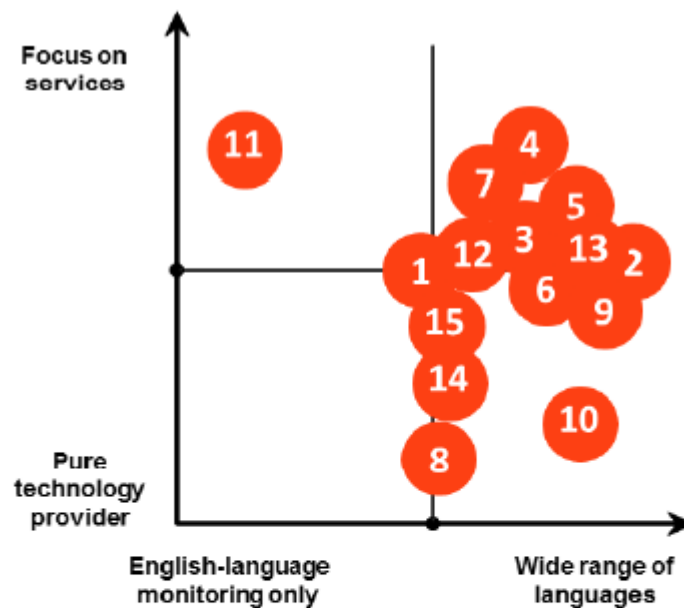
2.2   Commercial Online Reputation Management Tools

Because online reputation management tools have an array of different kind of users, the features of the tools vary from product to product. Econsultancy Digital Marketers United has identified 47 features that differentiate online reputation management tools from each other. As written in the report conducted by Econsultancy Digital Marketers United there are approximately 140 different commercial online reputation management tools on the market and half of the vendors have more than two years of experience. (Econsultancy 2011)

2.3   Differentiating features

According to Econsultancy Digital Marketers United one of the major differentiating features of the online reputation management tools are the sources where the products find the mentions and discussions. Some tools focus on traditional news websites, while others give more focus to blogs, forums or social media. Some tools like Visible Intelligence focus solely on social media and are often called as social media management tools. As seen in Figure 1 two other important features outlined by Econsultancy Digital Marketers United are the language support and focus of services or technology.

Some online reputation management tools are able to analyse only English mentions and discussion while others are able to work in multiple languages.

Figure 1. Chart illustrating the services and technology focus as well as the language support of 15 most prominent online reputation management tools. (Econsultancy 2011)

As seen in Figure 1, some tools are sold as technology while others focus more on service by providing consultancy, manually written reports and customer service by phone. The advantage of having a consultant or analyst is that the client can get much more out of the tool and save resources as there is somebody explaining, cleaning and verifying the data.

Table 1. Comparison of 15 most prominent online reputation management tool providers' business models. (Econsultancy 2011)

Business model and approach	Alterian	ASOMO	AT Internet	Attentio	BrandsEye	Brandwatch	Cymfony	Infegy (Social Radar)	Mark Sentinel	Meltwater Group	Onalytica	Radian6	Sentiment Metrics	Sysomos	Visible Technologies
Research	●	●	●	●	●	●	●	●	●	●	●	●	●		●
Consulting	●	●			●	●	●	●	●	●	●	●			●
Agency services		●			●		●	●	●	●	●	●	●		●
Search Engine Optimisation / Online content services	●				●						●			●	
Software (hosted)	●		●	●	●	●	●	●	●	●	●	●	●	●	●
Software (installed)			●	●						●			●		
White label offering	●		●	●		●	●	●	●			●	●	●	

As seen in Table 1, Econsultancy Digital Marketers United has indentified 15 most prominent online reputation management tool providers and compared their business model approach. Tool providers like AT Internet, Attentio and Sysomos do not offer consulting and agency services at all but focus on technology. (Econsultancy 2011)

2.4 Automatic Analysis

Related to the technological focus is also the depth of the automatic analysis the tool is able to perform. Some tools provide high quality analysis that offers aggregated information of all the mentions and discussions found by the tool. This data can be displayed in easy to read charts, tables and key performance indicators that can together form a dashboard where one is able to see the current situation and the evolution of online reputation over time. Sophisticated analysis features can offer information about the topics and sentiments of the relevant mentions found from the Internet.

2.5 Online Reputation Monitoring Tools

One key differentiating factor is also the focus on management and monitoring. Some tools only offer features to monitor the online reputation management while tools like Radian6 offer features to take part into the discussion. Tools that offer only monitoring

features are often called online reputation monitoring tools and the ones going one step further are called online management tools. (Wikipedia 2011i)

3   Prism Online Reputation Management Tool

Prism is an online reputation management tool developed by NetBooster Group. Prism is sold by NetBooster Group to client companies as a service that includes the tool, consultancy and customer service. The service has a monthly fee without any additional charges. Prism has been in use by multiple companies in France and by some companies worldwide. The clients who have used Prism include companies such as one of the world's major oil and gas groups Total S.A. and one of the world's leading hotel groups Accor.

3.1   NetBooster Group

NetBooster is a France based digital marketing agency founded in 1998. NetBooster delivers services to 24 countries and has offices in Germany, Spain, Finland, Italy, the Philippines, China and Brazil. NetBooster also owns majority of the shares of Guava Group operating in United Kingdom, Sweden and Denmark. NetBooster offers a wide array of services such as search engine optimization, pay-per-click advertising, display advertising, affiliate marketing, website development, analytics consultancy, social media marketing and facebook marketing services. NetBooster also offers an array of tools and services including ROI Platform, Visibility Platform and Prism. (NetBooster 2011a, buypennystocks 2012)

3.2   History of Prism

The first version of Prism was released in the March of 2009. It included features like topic, sentiment and category analysis. Also support for social media platforms including Facebook and Twitter were added shortly after the release.

The first version of Prism used Google Custom Search API to find mentions from the Internet causing the tool to find mentions only from a predefined set of websites. The tool did not include any kind of automatic analysis and all analysis had to be conducted manually. These two characteristics limited the usefulness of the tool and made it more of a database to store information that was found and analysed by human. This made

the first version of Prism very service oriented tool where the consultant had a lot responsibility over the quality of the online reputation management of each campaign.
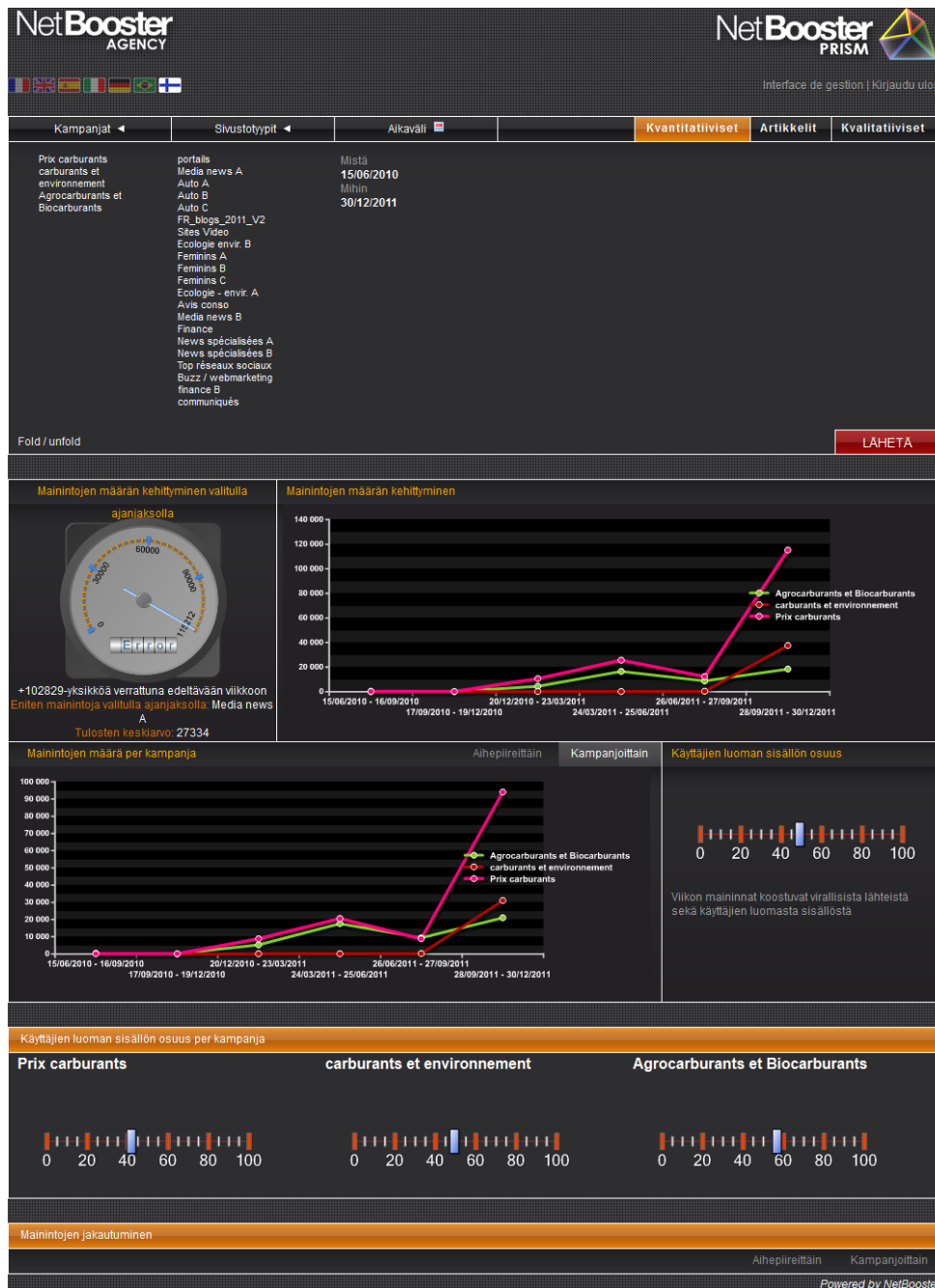

Figure 2. Campaign view of the first version of Prism

As seen in Figure 2, the first version of Prism had a dark, moody and somewhat unappealing user interface. It was not very easy to use either as most of the menus and buttons did not stand out from the rest of the page. The function of the user interface items was not clear enough in many cases. Also the administering and configuration

views of the tool were hard to use and many trivial tasks were impossible or very tedious to perform. In addition, the first version of Prism was very slow and suffered from errors that caused it to crash.

3.3 Objective and Features of Prism

The objective of the Prism project is to provide an easy to use web interface for online reputation management consultants and their clients to monitor the online reputation of their subjects of interest as well as to be able to address issues affecting the reputation in a negative way.

This is achieved by finding the mentions about the subjects of interest in the Internet, analysing them statistically and providing a visual representation of the analysed data. The finding of the mentions should be fully automatic after the initial setup and configuration of the subjects of interest. The statistical analysis of the acquired data should be automatic too but consultants should be able to perform statistical analysis too and be able to override as well as disable the automatic analysis. The visual representation of the data should be simple enough to show the overview of the reputation just by logging into the system but also sophisticated enough to enable filtering, zooming into and comparison of the data easily.

The main focus on the Prism tool is on the monitoring and analysis of data while the actual reputation management and tackling the reputation problems is left to the client and the consultant.

More detailed list of the functional requirements of the second version of Prism can be found from the Appendix 1. Functional requirements were divided into three groups including web monitoring, administration and social media monitoring and engaging. The first group of requirements describe what kind of data is to be displayed to the end user and how the articles and discussion need to be grouped and analysed. The second group of requirements define what kind of data has to be imported to Prism from the social media platforms and what kind of social media related interaction the user can perform in Prism. The last group of requirements lists all the tasks the users of Prism have be able to perform in order to setup and administer the surveys, user accounts and all other entities in Prism.

There are many things that the new version of Prism aims to improve from the old version while the basic features remain the same. One of the major aspects that need improvement is the user interface. The user interface visible to the client aims to be easier and faster to use as well as give an impression of trust and professionalism while the interface visible to consultants and administrators aims to be faster and easier to use. Also the setting up and configuring of the campaigns is to be easier, fast and more straightforward. The major features to be added in the new version is the ability to find mentions from other websites than the ones given to the tool and the possibility to white label the tool which means changing the logo and colours of the tool. White labelling enables the tool to be sold to agencies that can then resell it further. Another major feature to be added is the ability to perform analysis automatically without human interaction. Automatic analysis saves time and money as well as enables better coverage of the analysis.

4 Technologies Powering Prism

There are many technological decisions that need to be taken when building an online reputation management tool. This chapter focuses on evaluating and choosing the programming language and web application framework used to implement the new version of Prism. User interface related technologies and libraries are evaluated and chosen in the chapter number eight.

4.1 Programming Language

The programmer of the Prism project being familiar with C#, Java and PHP programming languages which are all suitable for web application development gave the project team the freedom to choose between the three back end programming languages. Choosing C# would practically mean using Microsoft's proprietary ASP.NET web application framework. The advantage of this is that the earlier version of Prism is also written with C# using ASP.NET so old codebase could be reused. Unluckily the old codebase is largely undocumented and outdated so it should be refactored which reduces the efficiency of using the old code. The disadvantage of using ASP.NET is the commitment to use Microsoft's proprietary interactive development environment, compiler,

web server and other technologies. NetBooster is also strategically moving away from the .NET architecture. (Wikipedia 2011b)

Java on the other hand is actively used by NetBooster in various projects with Spring and Struts frameworks. There are many Java web application frameworks available but only Spring was considered because it is the only one the project team had experience with. The main concern on using Java is the large overhead caused by learning and using the highly complex frameworks such as Spring. Java frameworks often have separated templating, testing, security and database frameworks steepening the learning curve considerably (Wikipedia 2011c). The project team agreed that Java with its frameworks introduces too much overhead and complexity to the code that using it is not efficient for the Prism project. (Wikipedia 2011e)

PHP was chosen as the programming language to be used because the programmer has the most experience using it and its frameworks compared to the other languages. PHP also offers a wide variety of frameworks with different feature set and complexity (Wikipedia 2011c). Unlike in Java frameworks most of the PHP frameworks offer all the basic functionality in one interoperable package without the need to learn multiple frameworks (Wikipedia 2011c).

4.2 Web Application Framework

Web application frameworks are used to eliminate the need to write similar low level programming code again in every project. Web application frameworks include the code needed to perform the basic features of a web application including but not limited to database operations, inserting, editing, viewing and deleting items, internationalization and localization, login and security, testing, caching and form validation. There are more than 20 PHP web application frameworks that are still being updated to choose from. Each of the frameworks is based on different paradigms and contains different features implemented in their own way. (Wikipedia 2011c)

Required features of the web application framework of Prism include Ajax, MVC, ORM, internationalization and localization, security, templating and validation support. Scaffolding and caching framework are good features to have, but are not a requirement. (Wikipedia 2011c)

Based on a previous experience CodeIgniter, Kohana, Zend Framework and Yii were evaluated to be used as the web application framework for the Prism.

As seen in Table 2. CodeIgniter does not have full internationalization and localization support nor ORM features, therefore it does not fulfil the requirements and cannot be used. (Wikipedia 2011c)

Table 2. Comparison of PHP web application frameworks (Wikipedia 2011c)

Name	Ajax	MVC Push/Pull	Internaliazation and localization	ORM	Security	Template	Caching	Form Validation	Scaffolding
CodeIgniter	Yes	Push	Mostly	No	Yes	Yes	Yes	Yes	Yes
Kohana	Yes	Push	Yes	Yes	No	Yes	Yes	Yes	No
Yii	Yes	Push & Pull	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zend Framework	Yes	Push & Pull	Yes	Yes	Yes	Yes	Yes	Yes	No

Kohana on the other hand does not include the required security features and user management and they must be installed as a plugin. There are multiple plugins available for this and none of them works out of the box for Prism. (Wikipedia 2011c)

Zend Framework on the other hand fulfils all the requirements but has other disadvantages. Zend Framework follows use at will paradigm where all of its components are loosely coupled to each other. This has an advantage that one can only use the features of Zend framework needed in the project but when looking for a solution that offers most of the common features of a web application framework in one package. Although this use at will approach introduces unnecessary complexity to the low level code. Zend Framework is also a product by Zend technologies. Even though Zend Framework itself is free to use, other products such as servers and development environments supporting Zend are commercial products. (Wikipedia 2011n)

Yii framework fulfils all the requirements and it also includes scaffolding features. By using scaffolding the framework automatically creates files containing the required code to perform add, edit, view and delete operations. These files are created based on the database schema created. Scaffolding is extremely useful when prototyping a product. After scaffolding the files can be modified to fulfil the requirements of the

project. Yii was chosen as the web application framework of the project. (Wikipedia 2011c, Wikipedia 2011j, Yii Software LLC)

Yii is based on Model-View-Controller architecture like many other PHP frameworks. The model architecture takes use of database access objects, PHP Data Objects and enables access to different database management systems through unified PHP-based interface. Yii offers an object-oriented method for creating SQL queries and takes advantage of Active Record pattern which is a well-known Object-Relational Mapping approach. Yiis input validation is based on prebuilt and custom validators that are set to a property of each data object. Yiis authentication and authorization supports multiple access control methods including role-based access control. Yii has a built-in support for jQuery and jQuery UI taking advantage of real-time Ajax update enabling user interface controls such as autocomplete input fields. (Yii Software LLC)

5   Mention Discovery

This chapter focuses on finding the relevant mentions from the Internet which is a major challenge for all media monitoring and management systems. As shown in Figure 3 it is the first step needed to gather the key information needed by the tool. There are ultimately two ways of doing this. First is to crawl the web pages as search engines do and analyse the web page content for related keywords. The second option is to use the various application programming interfaces provided by web services to search the content in each service. Both of the methods are very time and resource consuming processes.

Figure 3. Flowchart illustrating the order of data processing in online reputation management tools.

The problem with the first approach is that according to a study conducted by Maurice de Kunder the web contains over 12 billion web pages and the number is growing so rapidly that the number of computers needed to find and analyse them would be huge (Maurice De Kunder 2011).

The problem with the second approach is that according to ProgrammableWeb there are over 4000 thousand web services with public application programming interfaces each of them requiring their own implementation to the tool searching for the mentions (ProgrammableWeb 2011). This causes the amount of programming needed to search in all the services to take hundreds of thousands of work hours. And even if all the web service APIs could be implemented the tool still found only a small portion of all the mentions on the Internet.
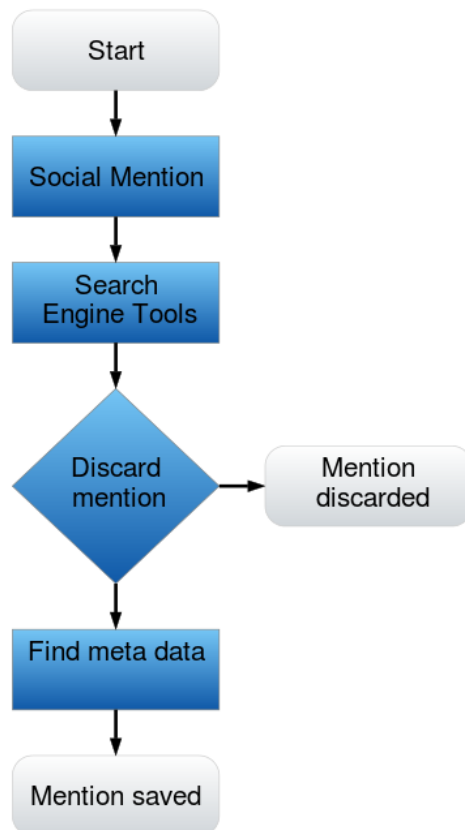
Figure 4. Flowchart illustrating the finding of the mentions

Prism uses third party services that solve most of the problems related to finding the mentions. As shown Figure 4, the services used by Prism include Social Mention API and tools provided by search engines. These tools combined provide a source that covers most of the websites in the Internet. After finding the mentions their quality and relevancy is verified automatically by Prism. If the quality and relevancy of the mention are sufficient Prism searches for meta data of the mention and saves it to the database.

5.1   Finding Mentions Using Social Mention

Social Mention is a web service that uses application programming interfaces of 81 social media platforms to search mentions by keyword and other search parameters. This includes the most used platforms such as Facebook, Twitter and Reddit as well as tens of platforms with a smaller number of users. The service also offers its own application programming interface to access the data programmatically. The usage of the Social Mention service removes the need to implement multiple application programming interfaces because they are already aggregated to the Social Mention API. (Cianciullo, J)

As seen in Figure 5, Social Mention offers a search engine style user interface where the user inputs a search phrase and the application outputs a list of mentions found. In addition to the mentions Social Mention also analyses the sentiment, top keywords, top users and other metrics of the mentions.
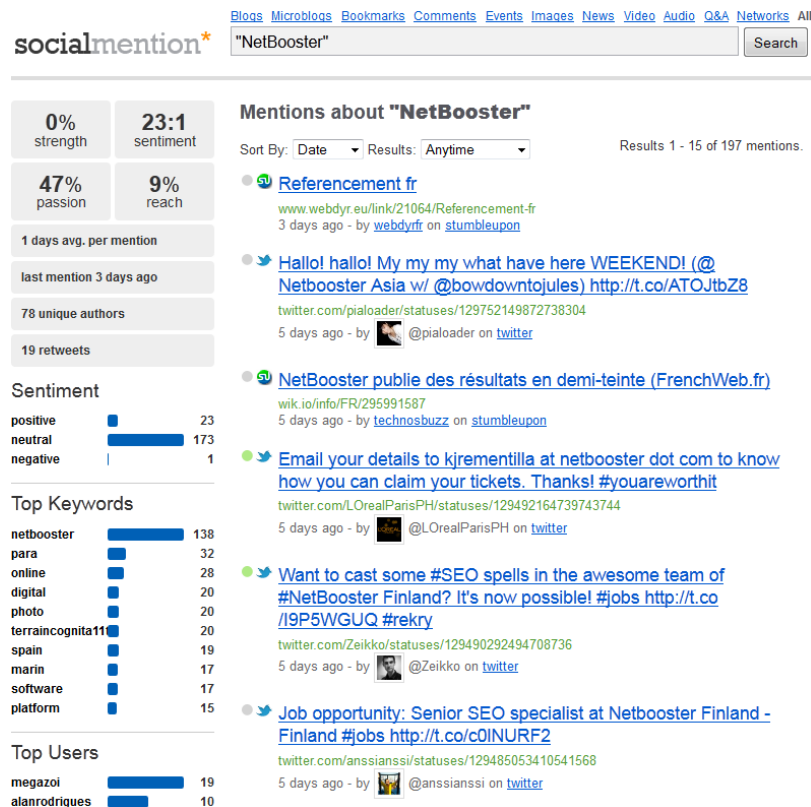
Figure 5. Social Mention search results for "NetBooster" search phrase

Prism uses Social Mention to find new mentions by keywords. Social Mention API offers a parameter that can be used to define the timespan of mentions searched (Cianciullo, J., 2010). Only new mentions are received using the timespan parameter through the API each time. The service returns a number of details of each mention including but not limited to title, description, link, timestamp and language (Cianciullo, J., 2010). The service also returns information about the author of the post in including the user-name, image and a link to his profile in the social media platform (Cianciullo, J., 2010). All this data is used for further analysis by Prism.

By using Social Mention Prism is able to find all the relevant mentions from the most popular western social media platforms easily without having to implement API clients for each platform. This approach makes the Prism easier to build and maintain but introduces a dependency to an external service and a single point of failure. If Social Mention suffers from a downtime Prism cannot find any mentions for the period of the downtime. Social Mention service does not include many eastern social media plat-forms including Chinese microblogging service Sina Weibo which, according to Wikipe-dia, has over 250 million registered users (Wikipedia 2011l). This makes using Social Mention effective only when operating in western countries.

5.2   Using Search Engines to Find Mentions

Search Engines are applications that search information from the Internet and offer a simple user interface to search the information using keywords (Wikipedia 2011m). In other words search engines address the problem of having to search the relevant arti-cles from the 12 billion web pages of Internet. Search engines are an easy way for human users to access the information in the Internet but harnessing the power of search engines to be used by another application is not so straightforward.

First of all the user interfaces or search engines are designed to be used by human not by computer. Creating an application that uses a search engine user interface takes some time but the problem is that the user interfaces of the search engines get up-dated and changed very often (Third Door Media 2011). The user interface can also be slightly different every time the engine is used. Most search engines also prohibit the use of automated tools that use the search engine instead of a human user. For exam-ple Google prohibits this in their terms of service (Google Inc. 2007). Many Search En-

gines including Google are also taking measures to prevent the automated usage of their search engine (Goole Inc. c).

On the other hand some search engines offer application programming interfaces for applications to perform search queries. The APIs are usually restricted or have limitations compared to the native user interface of the search engine. Microsoft's Bing search engine offers this kind of service but its possibilities and restrictions are not clearly documented (Microsoft). Google used to offer similar API but it was discontinued in December 2006 (Google Operating System Blog). However Google offers other services that can be used to find relevant mentions but they all have some drawbacks.

Google Alerts is a service which provides an RSS feed to all the new articles Google has found matching a specified search phrase. This kind of service is exactly what is needed to find the relevant mentions. The drawback of this service is that one can only have a very limited number of Google alerts running and there is no API to create the alerts. (Google Inc. d)

Google Custom Search API is a service that allows a developer to build their own customized search engine by using Google's technology. The custom search engine can process an unlimited number of queries and search phrases. The drawback is that the custom search engine only searches from a defined set of websites that must be configured upon the creation of the custom search engine. (Google Inc. e)

Even though search engines provide these tools to perform searches and analyse the search results programmatically it is difficult to use them to find all the relevant mentions in the Internet. Search engines have their own interests to protect their business and prevent others from building user interfaces taking advantage of their technology.

5.3   Discarding Irrelevant Mentions

It was noticed during the development of Prism that even though the tools provided by search engines and Social Mention API have a lot of features and filters to find the relevant mentions matching the search phrase there are a lot of irrelevant mentions in the results. In the same way as performing a Google search gives good results on the first pages and a huge number of poor quality results after that. Thus analysing and

validating of the mentions before saving them to the database increases the quality and relevance of the mentions significantly.

The validation process starts by Prism checking if the website still exists because search engine indexes and Social Mention API can include websites that have been already deleted or moved. This is done by Prism automatically downloading the website where the mention is found. Prism uses a library and PHP extension called cURL to download the Hypext Text Markup Language (HTML) of the website which can then be processed with PHP string functions. If the website still exists and the HTML is successfully downloaded it will be processed with a PHP library called PHP Simple HTML DOM Parser. This library turns the HTML into PHP objects that can be searched from by using selectors in the same way as in CSS and jQuery. This makes the searching and manipulation of the HTML much easier and reliable than using regular expressions because regular expressions fail easily if the HTML is malformed. PHP Simple HTML DOM Parser is used to remove all the content from the HTML that is believed to be irrelevant to find mentions from. These elements include links, input forms, checkboxes and dropdown menus. Mentions are often found from these elements but are usually totally irrelevant. After all the irrelevant HTML has been removed with PHP Simple HTML DOM Parser the HTML is searched for the mentions by the same search phrase that was used in the first place to find the page and if no results are found the mention is discarded. This has proven to be a very effective way to ensure the quality and relevance of the found mentions. (The PHP Group a, Chen, S. C)

5.4 Finding Mention Meta Data

The mentions that were not discarded at the HTML check are saved to the database. In this step Prism needs to find and save the metadata of the mention, actual text body and create a PDF file and images of the website. This data will be used in the further automated analysis as well as shown to the end user through the Prism user interface.

The meta data is saved by using the data from Social Mention API and tools provided by the search engines and the HTML of the page. The HTML usually contains specific meta elements to describe its contents and other HTML elements such as title and

headings. The contents of these elements can be easily found with PHP Simple HTML DOM Parser and saved to the database.

After saving the meta data the processing continues by Prism creating a PDF file and a screenshot of the website. Prism creates the PDF file by using a command line tool called CutyCapt. CutyCapt uses WebKit web browser engine and Qt application framework to render the page and turn it into a PDF file. CutyCapt is a state of art application and the PDF files created are very close to how the website is viewed in a browser. CutyCapt is able to load the CSS and JavaScript of the page but not Flash. Because of all the JavaScript, CSS and images CutyCapt downloads and runs creating a PDF file takes a considerable high amount of time. The time consumed can be anything from a fraction of a second to minutes, depending on the website. In addition to the PDF files Prism also creates two PNG images of different size that will be shown in the user interface to the end user. CutyCapt is also able to create PNG files but not simultaneously with PDF files without loading the website multiple times. To prevent having to spend time and resource to run CutyCapt again the PNG images are created from the PDF files with linux command line tool called convert. Convert creates the images fast and exactly as they are in the PDF files. The only downside that was found during the development is that the image does not look exactly the same as the original website because of the limitations such as fixed page width and lack of interactivity in the PDF format. (Höhrmann, B., WebKit Team, Nokia Corporation, LinuxManPages)

5.5 Multi-Instanced Processing

There are hundreds of keywords Prism needs to search mentions for every day which results in thousands of found and processed mentions every day. Processing of a single mention can take several minutes due to the launching of CutyCapt, WebKit and loading of the whole website. During the development it was noticed that this rapidly leads into a situation where Prism cannot keep up and process all the mentions found.

To solve this problem Prism needs to process multiple websites simultaneously. This is achieved easily in PHP and Apache environment because it is multithreaded by nature. One PHP script runs only in one thread but there can be multiple instances of the same script running at any given time. There are a few things that need to be taken care of in the application level. One of them is the prevention of processing of the same

search phrase by two instances at the same time. Prism handles this by saving a time-stamp to the database when updating of a search phrase is begun. If there exists a start timestamp for a search phrase new instances are not started to process this search phrase. The timestamp is cleared when processing is complete.

Processing start timestamps can also be used to control the number of processing instances. The number of processing instances should not exceed the number of processors devoted for processing of the mentions. New processing instances are created regularly automatically by the server itself. This is done by using a time based job scheduler called Cron that is found in most Unix-like operating system. Cron runs a command line web browser called elinks that requests a PHP script on the server itself. This PHP script calculates the number of processing instances running by using the timestamps in the database. If the number of running instances is lower than the maximum allowed number defined in the Prism configuration the PHP script runs elinks again requesting the script that actually processes the search phrases. Same kind of multithreading could be achieved by taking advantage of the Process Control support of PHP and calling pcntl_fork function.

The multi-instanced processing speeds up the system significantly but causes one more problem. If a processing script fails it will leave the timestamp to the database preventing this search phrase to be processed again and new processing threads to be started. There are many reasons why processing scripts may fail because they are dependent on several external applications such as CutyCapt, WebKit and cURL as well as services such as AlchemyAPI and social media platform APIs. Prism handles this problem by clearing the timestamp and retrying to update the same search phrase if a processing script fails. This is done by catching all the exceptions and performing necessary operations in the catch block. However, in PHP this is not enough because a script can fail because of a PHP error instead of an exception. Script failures called by PHP error cannot be caught by catch blocks because the whole PHP script will terminate.

Prism handles PHP errors by registering a custom PHP shutdown function that will be called when the running of a PHP processing script ends regardless of the reason. This shutdown script will reset the timestamp and create a run elinks again to request the

same script. To prevent infinite recursion by a script calling itself Prism uses elinks with 1 second timeout configuration. This way the calling script will continue after the called script has been started. PHP has a useful feature that scripts are terminated if the calling browser cancels the requests or times out. In this case it causes a problem by terminating the script that was just called. This behaviour can be disabled by setting PHP environment variable called ignore_user_abort to true. (Wikipedia 2011d, The PHP Group d, The PHP Group c)

6 Mention analysis

This chapter focuses on the second major problem, mention analysis, faced by many online reputation management tools. The goal in analysing the mentions is to provide the user an overview of the online reputation of without having to read the mentions. This is useful because the number of mentions can be several thousand each day and reading them would take an excessive amount of time.

There is an array of metrics that are to be analysed. The most obvious and simple is the number of mentions. Prism also uses a more complex metric depicting the number of Internet users who have read the mention, this metric is called visibility index. These metrics are quantitative, in other words they depict the amount of buzz in the online media. Prism uses also a number of qualitative metrics that measure the content of the mentions. Prism analyses the language, topic and sentiment of the mentions to provide meaningful qualitative metrics. Prism also groups websites where the mentions are found to categories to provide an overview of the type of websites generating the web presence. All these metrics together turn the gathered mentions into meaningful information which is then displayed as graphs and tables to the consultants and clients.

The analysis can be performed manually as in the old version of Prism or automatically. Manual analysis has the drawback that someone has to read through all the articles where mentions are found and analyse them one by one which takes a lot of time and resources. The automatized approach is based on analysing the text using linguistic, statistical and machine learning technologies to turn the human generated text to a structured data (Wikipedia 2011o). This process is called natural language processing (Wikipedia 2011h).
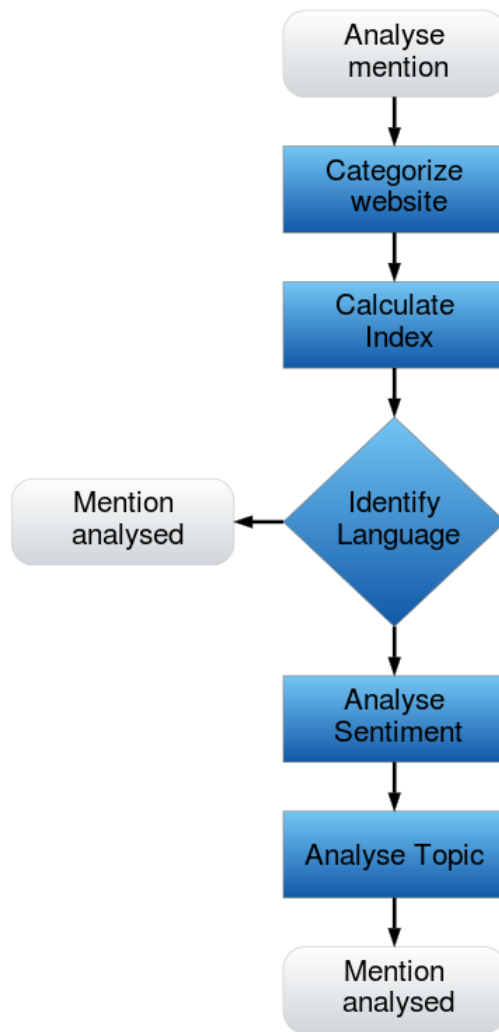
Figure 6. Flowchart illustrating the mention analysis

Prism uses third party services that analyse the mentions computationally. As seen in Figure 6, there are several steps that need to be taken to analyse the different aspects of the mention. The implementation of these steps is covered in detail in the following five chapters.

6.1   Number of Mentions

The number of mentions is a simple metric which is calculated by the plain number of mentions in the selected time period. The number of mentions does not accurately depict how known the brand, enterprise or person is in the social media because some mentions are more visible in the internet than others. Mentions in a website with millions of visitors each day have much more influence to the online reputation than mentions in a website with tens of visitors per day.

6.2   Website Categorization

Prism categorizes all the websites to provide information what kinds of websites generate the online reputation. In Prism website categorization can be performed manually or automatically. Categorization is performed globally so that all clients use the same website categorization database. However the categorization can be customized for each client by defining new set of categories and categorizing websites to different categories for one client only. Automated categorization is performed only in the global level but manual categorization can be performed both in the global level or client level.

The automatic categorization is a complex concept and implementing that natively on Prism would have required too much resources but luckily there is an online service called SimilarGroup that performs this kind of analysis already. SimilarGroup offers an API that gives a textual representation of a website's category. SimilarGroup uses automated technologies to categorize websites. SimilarGroup has categorized over 25 million websites and more than 30 000 websites are categorized each day. The Internet consists of more than 500 million websites with unique hostname so SimilarGroup has categorized roughly 5% of all the websites (Netcraft Ltd. 2011). Even though the number of categorized websites is very small it is very meaningful because those 5% draw most of the traffic in the Internet. SimilarGroup has a hierarchical category tree with a huge number of categories. However using a great number of categories is not useful in Prism because the categories are used by humans to understand and analyse the information. Prism only uses the highest level of categorization resulting SimilarGroup to group all the websites to fourteen categories which is good number for human to process. The categories are very general such as "News" and "Business". (SimilarGroup Ltd. 2011)

After four months Prism has found over 27 000 websites and found a category for about 7 000 of them from Similar Group making about 25 % of the website categorized. In the four months Prism has found about 200 000 mentions and 70 000 from websites that are categorized by SimilarGroup thus about 35% of the mentions are in categorized websites.

The manual categorization is used to categorize the websites that SimilarGroup is unable categorize. It would still take too much resources to categorize all the uncategorized websites but focusing only on the websites where a great number of mentions has been found raises the number of mentions in the categorized websites easily. To do this efficiently Prism offers a list of uncategorized websites that has been sorted by the number of mentions found in each website as seen in Figure 7. The list also has a dropdown menu to categorize the websites fast. This enables the consultant to categorize the uncategorized websites that contain the most mentions fast and easily.

Uncategorized websites

Domain	Number of mentions	Category
forum.autocadre.com	793	Uncategorized ▾
www.dijonscope.com	686	Uncategorized ▾
friendfeed.com	673	Uncategorized ▾
www.forum-auto.com	670	Uncategorized ▾
www.prixalapompe.com	617	Uncategorized ▾
www.shopromo.com	490	Uncategorized ▾
www.codespromotion.fr	468	Uncategorized ▾
www.midilibre-annonces.com	463	Uncategorized ▾
www.banque-pdf.com	386	
assistance.b-and-you.fr	379	

Reference
Business
Science
Uncategorized
Computers
News
Arts

Go to page: < Previous | 1 | 2 | 3 | 4 | 5 | 0 | Next >

Figure 7. User interface for categorizing uncategorized websites in Prism

In addition of increasing the coverage of the categorization manual categorization also improves the accuracy. Using the automatic categorization in conjunction with the manual one provides an accurate metric in an inexpensive way. Automatic categorization gives also freedom to the frequency when manual categorization has to be done.

6.3   Visibility Index

Because the number of mentions does not give an accurate figure how known the enterprise, brand or person is in online media there is a need for metric that depicts how many users have actually seen or read the mention. Prism uses a metric called Visibility Index which takes into account the number of visitors per each website.

The first problem of calculating the visibility index is the measurement of the website traffic. While measuring the traffic of a single self-controlled website is easy measuring the traffic of a website controlled by another party is impossible. A company called Alexa Internet, Inc. which is owned my amazon.com offers a solution to this problem. Alexa offers global web metrics including the approximate number of users of all the major websites per country. Metrics offered by Alexa are based on data that is gathered by users of Alexa toolbar. Alexa toolbar, as seen in Figure 8, is a browser plugin that sends Internet usage data from the client's browser to Alexa. In other words Alexa has a diverse group of people browsing the Internet and sending them the information of websites visited. As a return Alexa toolbar offers tools to the users of the toolbar using the data they gather from the users. The accuracy of the data provided by Alexa is not perfectly accurate because Alexa toolbar users represent only a subset of all the Internet users. The subset of users may be biased to specific user groups such as users with a specific profession or cultural background. (Alexa Internet Inc. a, Alexa Internet Inc. b, Wikipedia 2011a)
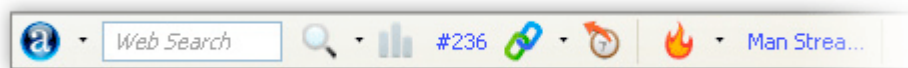
Figure 8. Alexa toolbar (Alexa Internet Inc. b)

In addition to providing the information with the toolbar Alexa also has an application programming interface to access information programmatically. The API access is part of commercial Amazon web services. The cost of retrieving the usage information of 1000 websites is $ 0.15. Thus retrieving the usage information for all the 70 000 websites found in the four month test period of Prism costs $ 10.5. The data Alexa provides is inexpensive and very valuable for any service performing any kind of global Internet analysis. Prism uses services of Alexa to determine the approximate number of users of a website per country. (Amazon Web Services LLC)

In addition to the approximate number of users per country visibility index also takes into account the category of a website. Each category has a manually set weight value from one to one hundred. The index of a category is used to scale all mentions that are found from a website categorized into the category. The idea behind the category weight is that different types of websites have different number of pages in them. For

example reference sites such as Wikipedia have a huge number of pages while blogs can have much smaller number of pages. Well known websites such as Wikipedia also have a huge number of users but the number of page views on a single page is relatively small because of the huge number of pages on the website. Thus the category weight tries to take into account the size difference between different types of websites.

The visibility index of a single mention is count by multiplying the score provided by Alexa with the category weight. This gives an index that has no absolute meaning. Thus the index itself does not tell anything but it can be used for a comparison between different datasets. The visibility index can be count separately for different countries using each country's Alexa score. Alexa also offers a global score if using a country specific index is not useful for the client.

Using the metric provided by Alexa and the manual weighting of the categories visibility index gives a rough estimate of the amount of buzz in the Internet. It is a more meaningful metric than the number of mentions because it takes into account the number of users in each website and the number of pages in each website category.

6.4   Sentiment Analysis

Sentiment analysis also known as opinion mining is a type of natural language processing technique that aims to determine the attitude of the author of a piece of content. Sentiment can mean the contextual polarity of the document or the emotional state of the author when writing. Sentiment can be described with different kind of adjectives, numerical scales or graphs but the simplest one is to show if the overall sentiment is negative, neutral or positive. Sentiment analysis can be performed manually or automatically. Performing sentiment automatically analysis contains many computational and linguistic problems because same words can have totally different meaning in different instances and sentences. For this reason sentiment analysis that is done by simply statistically analysing the words is not very accurate. To improve the accuracy of the sentiment analysis the grammar can also be analysed to understand the real context and meaning of the words. Also machine learning technologies can be used to improve the accuracy. (Wikipedia 2011k)

Sentiment analysis is a very important metric in reputation analysis because it can be used to determine the quality and type of the publicity. The quality and type of publicity are after all one of the most interesting attributes of reputation. In addition to providing the overall sentiment of the reputation, sentiment analysis can be used to filter out all the neutral and positive mentions. This makes it possible to find the mentions that require an action by the brand or enterprise being mentioned. Thus sentiment analysis is a key element when the usage of media management tool is extended from monitoring to the actual management and addressing of the issues.

There are about one hundred tools and services that can be used to perform sentiment analysis. Some of them are test services created by universities and companies, some are free services and some commercial products. A great number of the services perform sentiment analysis for Twitter mentions only. Twitter is widely used as a resource for sentiment analysis because of the high volume of mentions and the high speed that Twitter users react to new products and news. One of the best known twitter sentiment analysis tools is called twitter sentiment. It offers a minimal user interface similar to search engines to search for tweets as seen in Figure 9. In addition of the tweets Twitter sentiment also shows their sentiments and the aggregation of sentiment of all the found Tweets. (Twitter Sentiment 2011)
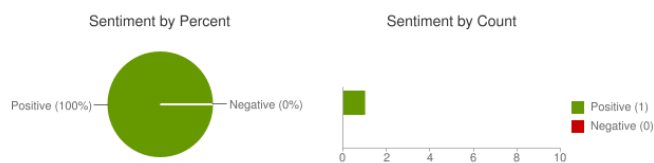
Figure 9. Twitter Sentiment user interface (Twitter Sentiment)

As can be seen in Table 3, there are numerous commercial services offering sentiment analytics such as Lexalyctics, Saplo, Clarabridge, OpinionCrawl, OpenDover and OpenAmplify. Many of these companies offer other text analytics services in addition to the sentiment analysis. Most of the companies offer sentiment analysis only in English. Few services such as Clarabridge, OpinionCrawl and alchemy have added support for other languages too. Some of these companies offer limited testing or even permanent free but limited access to their tools. Usually test accounts are acquired by sending an email to the company, but some have also an automatic registration process.

Table 3.   Supported topic and sentiment analysis languages of 6 text analytics services. (OpenAmplify, Saplo AB, Orchestr8, Clarabridge, Semantic Engines LLC, Byelex Multimedia Products)

Product	Topic Analysis	Sentiment Analysis	Other
OpenAmplify	English	English	Intent, Decisiveness, Emotions, Demographics, Categorization
Saplo	English, Swedish	In development	
AlchemyAPI	English	English	Text extraction, language detection, categorization
Clarabridge	No	English, Spanish, Brasilian, Portuguese, French, German, Russian and Mandarin	
OpinionCrawl	No	English	
OpenDover	English	English	

Prism uses one of the commercial services called AlchemyAPI developed by Orchestr8. AlchemyAPI is a product consisting of different kind of natural language processing services. Orchestra8 has been found in 2005 and is based in Denver, Colorado in United States of America. AlchemyAPI can analyse sentiment of a mention from plain text, HTML or just by receiving the URL to the mention. The service returns a numerical representation from -1 to 1. Closer the return is to -1 the more negative the sentiment is and the other way around for positive. The AlchemyAPI is also able to combine the sentiment analysis with its keyword and entity analysis services. In other words AlchemyAPI is able to find the sentiment for each keyword or entity mentioned in the article. Currently the AlchemyAPI's sentiment analysis supports only English language. AlchemyAPI offers 1000 free service queries per day but a license is required to per-

form more than that. Prism also supports manual sentiment analysis that can be performed for both unanalysed and previously analysed articles by clicking corresponding smiley in the user interface as seen in Figure 10. (Orchestr8 d, Orchestr8 a)
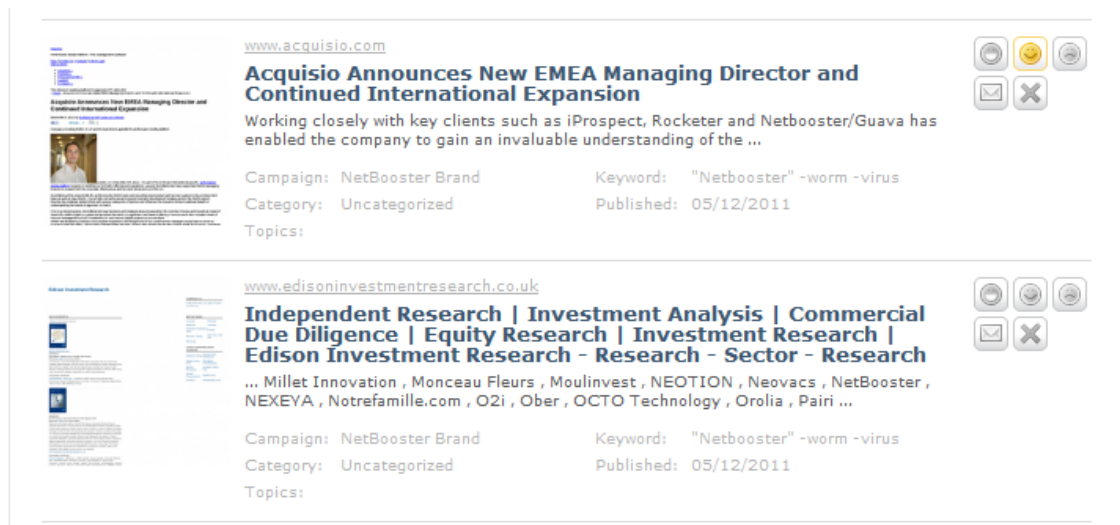
Figure 10.   User interface for performing sentiment analysis in Prism

Sentiment analysis provides information whether customers and press actually like a product or a service. This information is gathered much faster and from a larger group of people than is possible by surveys and other traditional ways of gathering feedback. Even though sentiment analysis can be conducted automatically which makes it less expensive than manual analysis it is still not inexpensive. Sentiment analysis performed by a commercial service usually costs something between 0.01 € and 0.1 € per analysed mention. Sentiment analysis is still worth the cost because it saves a lot of time for person analysing the reputation and makes it easy to find the mentions that require an action by PR or marketing team.

6.5   Language Identification

Language identification is a form of natural language processing that focuses on determining which language given content is written in. In other words language identification is a kind of text categorization used to get more information about pieces of content. Language identification usually relies on statistical approaches such as identifying frequent words and letters and comparing that to previously given learning data. Other forms of grammatical and machine learning techniques can also be used to improve the accuracy of the analysis. (Wikipedia 2011f)

Language identification is important in online reputation management tool because it helps to focus the reputation analysis and management to the mentions that matter. The problem in keyword based mention finding is that many keywords have multiple meanings in different languages. Also many brand and company names have different owners in other parts of the world. Language analysis can be used to filter out the unwanted mentions that are actually dealing with something totally different than what the user is interested in. In the case of global brands and companies language analysis can also be used to analyse mentions written for each language separately and compare the results with each other.

Implementing a language identification service for a small number of languages is not a huge undertaking but when the number of detected languages goes up the complexity of the system increases radically. There are thousands of languages and even hundreds of common languages in the Internet and many of them need different kind of techniques to differentiate them from other similar languages. The computational problems in the field of language identification are easily solved with modern computers that can analyse text faster than it's produced. The problems with text identification mainly revolve around the questions of choosing the correct statistical metrics used to differentiate languages making it more of a linguistic problem than computational.

There are a number of online tools available that perform language identification. Many of the tools are developed by various universities or other non-profit organizations. The tools usually have a limited number of detected languages, simple user interface and no application programming interface. Google also used to have a simple tool for performing language detection as seen in Figure 11. Unfortunately the tool has now been moved to part of Google's commercial Translate API v2 (Google Inc. a, Agarwal, A.).

Figure 11.   User interface of discontinued Google language detection service

Alchemy API also provides a language identification service which is called language detection and it identifies 97 languages. The input of the service can be a uniform re-

source locator, hypertext markup language file or plain text. The output includes language code in three standardized formats, text representation of the name of the language and can be encoded in XML, JSON or RDF. (Orchestra8 LLC e)

Prism uses the Alchemy API language detection service to identify the language of each mention found. Regardless of the high precision and a great number of supported languages the service can't identify the language of all the mentions. Those mentions are simply identified as unknown language. Prism does not include a user interface to perform manual language identification because performing such task manually is very demanding and because it is not a critical feature in the system but just an additional feature.

Interesting feature of the Alchemy API is that sentiment analysis query also outputs the language the mention was written in. It makes sense as the language identification is needed in order to do proper sentiment analysis of a mention. The language identification information is returned even when the actual sentiment analysis has not recognized the sentiment. This gives the opportunity to reduce the number of performed queries by using the sentiment service to do both sentiment analysis and language identification. This saves both computational resources and money spent for the API responses. (Orchestr8 LLC)

Language identification can be used to provide deeper analysis of the mentions and filtering out unwanted mentions. It is also relatively inexpensive feature due to the fact that it is easy to perform computationally by using statistical analysis.

6.6   Topic Analysis

Topic analysis is an activity that results in the knowledge of what is being discussed in a body of text. There are multiple ways to resolve the topic being discussed about. The simplest way is performing keyword analysis to analyse what kind of words are used in the text. The next step is concept analysis where higher level concepts are found by looking for concepts that are related to the found keywords. Also entity analysis can be used to find out what organization, countries, people and other entities are being talked about in the text. Topic analysis is not a trivial task but can be automatized by using statistical analysis, machine learning and other text analytics technologies. Topic analysis can also be performed manually quite cost efficiently because human can get

a grasp of the topic of an article just by reading the topic and couple of sentences from the article. (Wikipedia 2010, Wikipedia 2011g)

The results of topic analysis are used to compare the volumes of discussions in the topics the client is interested in. This provides useful insight on what topics the public are interested in the client's domain. However the best benefit of topic analysis is achieved when it is used in conjunction with sentiment analysis. By using sentiment and topic analysis it is easy to see what topics the public is discussing in positive and negative tone. This enables division of labour by making it easy for specific people in the PR team to take care of the negative mentions in specific topics. Sentiment and topic analysis together can also provide insights whether customers like certain products or services.

As seen in Table 3, there are multiple services providing topic analysis in English but not many services available for other languages. Topic analysis is highly dependent on the language used and same analysis methods cannot be used with all languages. Many companies offer other services such as concept tagging and entity extraction that can be used to perform topic analysis to some degree. AlchemyAPI does not offer topic analysis directly but the entity extraction, keyword extraction and concept tagging can be used for same kind of purposes. Entity extraction service finds all the real world entities from the article such as persons, companies, countries, cities, etc. while keyword extraction finds the most important words describing the body of text. Concept tagging is equivalent to topic analysis searching for high level concepts that the body of text discusses. AlchemyAPI also has features called entity level sentiment and keyword level sentiment analysis. These features provide sentiment for each keyword or entity found. Naturally this approach provides much better accuracy than getting the overall sentiment of the article and assuming that this sentiment applies to all the keywords and entities found. (Orchestr8 LLC d)

Prism uses AlchemyAPIs concept tagging service to find out the topics of the mention. The number of concepts recognized by the AlchemyAPI is huge and the service often finds multiple concepts for each mention. During the four month test period Prism contains roughly 4000 unique topics all found by using AlchemyAPIs concept tagging. The great number of topics gives precise information of the topics discussed but makes it

difficult to aggregate this information and provide meaningful information to the user. It is also very difficult for a human to extend and modify the topic analysis performed by AlchemyAPI because of the huge number of unique topics and topics in each mention. Therefore the topic analysis by default is completely automatic and the consultant cannot modify it in any way. The consultant can however disable the automatic topic analysis or make it totally manual. Manual topic analysis requires the consultant to create a limited number of topics and then rate the sentiment of each topic in each mention as seen in Figure 12.

Figure 12.   User interface for performing manual topic and sentiment analysis in Prism

There are many ways to perform topic analysis. It can be performed in a low keyword level or a higher concept level. It is also possible for it to be done automatically and manually. Topic analysis itself offers very little value but combined with sentiment analysis it can offer very valuable insights.

7   Social Media Monitoring

Social media users produce a high number of mentions for Prism to analyse but in addition to that many social media platforms contain other features that are interesting for social media monitoring. Most social media platforms contain different kind of actions and relationships such as liking a company, following a celebrity or just being friend with another person. Social media platforms also usually offer APIs to access these actions and relationships programmatically. These actions and relationships form new useful metrics to be used in social media monitoring and they can be used to calculate even more metrics. These metrics make social media very interesting for online marketers but are often hard to track and use to their full extent. Prism gathers a lot of metrics in one place in easy to read format. It is often hard or impossible to track competitor's social media metrics but Prism is also able to do this and compare the results with clients own results.

7.1   Social Media Networks

As seen in Figure 13, there are numerous different social media networks used around the world. New Social Media networks are created all the time but most of them never gain the attention of the masses or die shortly after millions of users have registered to the site. Social media networks require a critical mass of people before the network starts to generate enough content for it to be interesting for users. According to Vincenzo Consenza who conducted a research using data provided by Alexa and Google Trends for Websites the trend for the past 3 years seems to be that Facebook is slowly becoming the most used social network around the world (Cosenza, V. 2011).
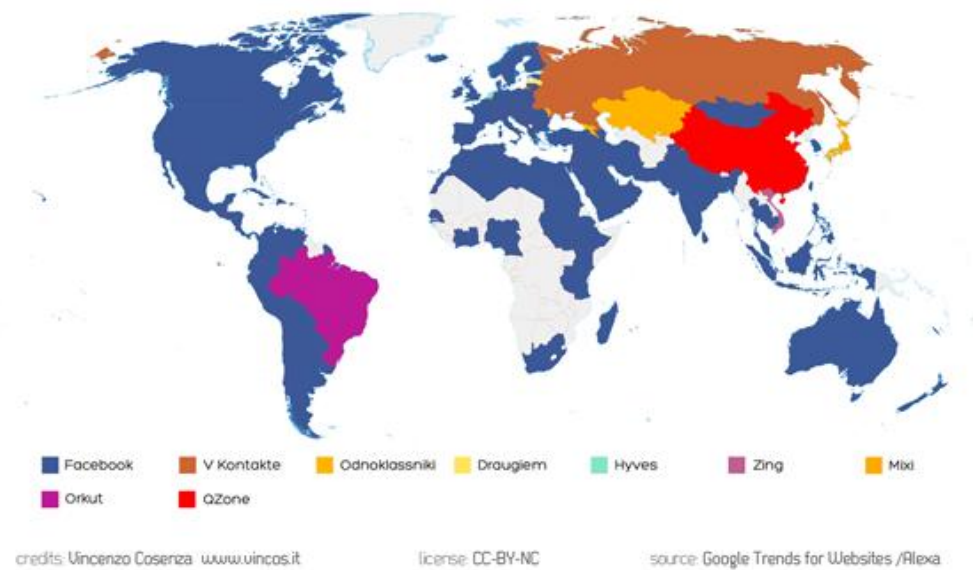
Figure 13.   The most used social media network by country (Cosenza, V. 2011).

Other smaller networks such as V Kontakte, Odnoklassniki are slowly losing their users to Facebook. There are also other Social Media Networks that are still growing such as Twitter, Google+ and Sina Weibo. The bottom line is that the world of social media networks is vibrant and changes rapidly. In order for Prism to monitor a social media platform it has to be able to use the social media's own API which requires specially written API client implementation for each Social Media. Thus the social media networks that are monitored need to be prioritised and focus has to be set on the important social media networks. Currently Prism is able to monitor Facebook, Twitter and Sina Weibo.

7.2   Social Media Metrics

Social media offers new opportunities in monitoring and measuring the outcomes of marketing efforts but also poses new challenges in offering the most relevant metrics. Many social media networks offer multiple metrics such as the number of likes and followers. While these metrics certainly tell something about one's influence in the social media network they do not tell anything about the return of investment because likes and followers themselves do not have any value. According to social media ana-

lyst Jeremiah Owyang using social media metrics that do not tell anything about the real business value has very little use for marketing executives (Owyang, J. 2011).

However digital marketing evangelist Avinash Kaushik has come up with good social media metrics that can be derived from most of the common social media networks and are all related to the real business value. These metrics are conversion rate, amplification rate, applause rate and economic value. Conversion rate is the number of comments per post which tells roughly how well one is interacting with his followers. Amplification rate means the shares and retweets per post which depicts how ones followers share his posts further. This is an important metric in social media where messages can receive huge audiences by being shared further and further. Applause rate means the number of likes per post. Applause rate tells how many people like and are interested in the content. The last metric introduced by Kaushik ties itself to return on investment by taking advantage of website traffic tracking tool such as Google Analytics. Google Analytics is able to differentiate traffic sources between social media networks and if used right even between individual messages in the networks. By using website traffic tracking tool it is possible to calculate the economic value of each visitor from a social media and the return on investment for the social media marketing efforts. (Kaushik, A. 2011)

At the time of release Prism creates line and bubble charts of the number of followers, likes, posts, comments and shares. All of the data is read once a day by using each social media network's own API. These charts contain all the data needed to create the metrics introduced by Kaushik but they are hard for the user to see because Prism doesn't calculate the rates already. Calculating these simple metrics is in the pipeline and will be added to Prism in the future. Also taking advantage of Google Analytics and other website traffic tracking tools is in the pipeline and will be added to enable the calculation of return on investment and economic value per visitor from each social media network.

8 User Interface

This chapter discusses the user interface of Prism and discusses of the principles, decisions and technologies behind it. Prism uses web based user interface to convey infor-

mation to the user and receive input from the user. User uses web browser such as Internet Explorer of Mozilla Firefox to access the interface. The objective of the interface is to be simple enough to get an overview of the client's online reputation fast and easily but also offer more advanced and refined information for those who are interested. The same user interface is used by the clients and employers who usually have very different level of skill and knowledge about Internet and reputation management.

The user interface is based on common web technologies such as XHTML 1.0, CSS and JavaScript. Prism also takes advantage Yii web application framework which includes a built in support for jQuery and jQueryUI JavaScript libraries. These libraries extend the capabilities of JavaScript making the creation of novel user interfaces easier. Prism also uses the CSS framework of jQueryUI. Prism also uses JavaScript libraries such as Highcharts and Google Chart Tools. (Yii Software LLC)

The main purpose of the Prism user interface is to display the results of the mention analysis to the user as well as the original mentions and the metadata associated to them. The user interface has to offer different visualizations such as line charts and pie charts to display the results of the mention analysis in easy to read format. Analysis visualizations have to offer a way to associate the analysed information to the actual found mentions. The user interface has to also offer data filtering by various criteria enabling the user to focus on specific set of data. The user interface has to include all the functionality needed to administer the application including user, client and campaign management.

8.1 Charts

Charts are graphical presentations of data by using different kind of symbols. The purpose of charts is to visualize complex and large amounts of data in a form that is easy for the reader to understand in a short amount of time without causing misunderstandings. There are multiple types of charts for example line charts, pie charts and flow charts.

Prism contains different types of data that needs to be conveyed to the user in a way that is easy and fast to understand but still accurate enough to base strategic decisions on it. Line, pie and bar charts are used along with tables and lists to convey the infor-

mation. Most of the data in Prism is time series data that is best to present in a form of line graph. Pie graphs are good to be used to summarize the same data displayed in the line graph to give an overview of the whole period. Bar graphs are good to be used on analysis that has two series instead of one for example to display the number of positive, neutral and negative mentions in each topic as seen in Figure 14.
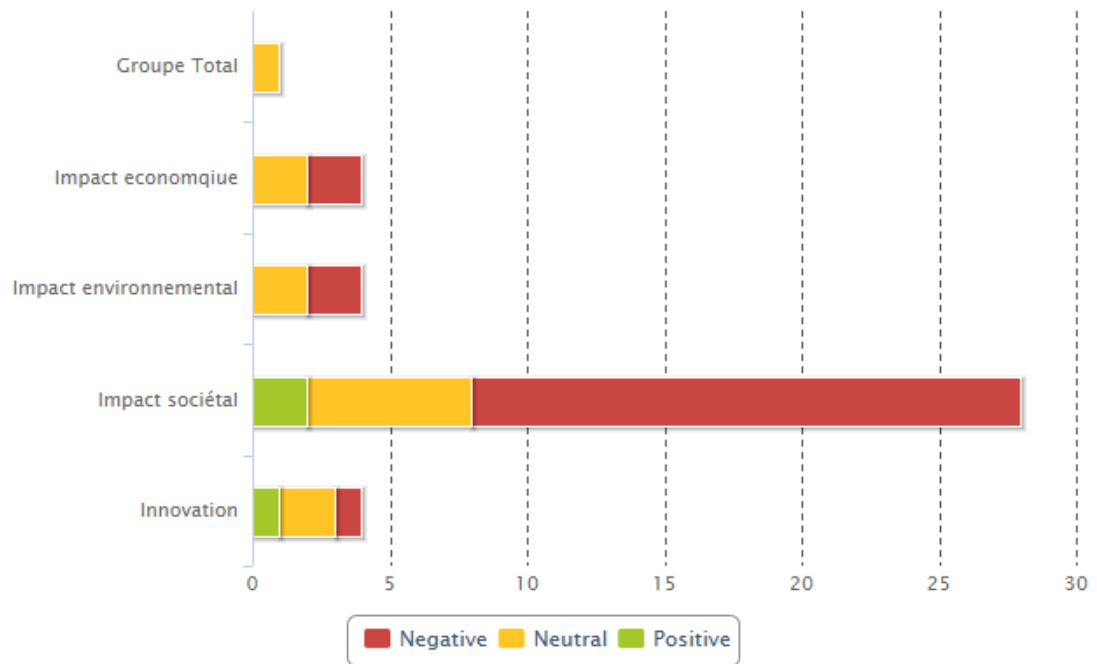

Figure 14.   Bar graphs illustrating the count of different sentiment in each topic

HTML does not have a native support for any kind of charts but there are many technologies and applications that create charts that can be displayed on websites. The traditional approach is to use the server software to create images to display the charts. An open source code library called GD can be used to create images and is also available for PHP (Boutell, The PHP Group b). The downside of using images is that they are static and cannot contain more information than any other image. Technologies such as JavaScript and Flash offer ways to create dynamic charts that the user can interact with. They can for example show data values on a chart when the user hovers the cursor over a data point in the chart. JavaScript and Flash offer animations, transparency, tooltips and other features that make charts not only to look better but also easier to use. JavaScript and Flash cannot draw charts natively but there are many code libraries which use these technologies to draw the charts. These libraries have a lot of similarities but also differences can be found. Some of the libraries are free while the others being commercial products.

Multiple chart code libraries were evaluated to be used in Prism with the following requirements. The code library has to offer line, pie and bar charts. Chart appearance must be fully customizable including colours, positioning, size and symbols used. Charts must work on a browser that does not have flash plugin installed. Charts must have a clear API that allows chart control and data feed by JavaScript. Table 4 includes the four best known chart code libraries that run in a browser environment. As seen in the table, they all fulfil the set requirements but have very different costs. FusionCharts being probably the most expensive one because they did not disclose the price but told that it is a monthly fee and Google Chart Tools being totally free of charge. (amCharts, FusionCharts, Google Inc b, Highsoft)

Table 4.   Comparison of chart code libraries (amCharts, FusionCharts, Google Inc b, Highsoft)

Code Library	Line	Pie	Bar	Fully Customizeable	Requires Flash	JavaScript API	Cost
amCharts	Yes	Yes	Yes	Yes	No	Yes	899 €
Highcharts	Yes	Yes	Yes	Yes	No	Yes	$ 360
FusionCharts	Yes	Yes	Yes	Yes	No	Yes	Monthly fee
Google Chart Tools	Yes	Yes	Yes	Yes	No	Yes	Free

From all the evaluated chart libraries Highcharts was chosen as the library to be used in Prism. Highcharts relies on JavaScript generated HTML to display charts. It works reliably and fast while offering simple but powerful API to control and modify the behaviour of the charts. Even though all of the four libraries fulfil the requirements, many differences were found in the chart libraries. Highcharts for example offers a very simple way to feed data on it. Highcharts can be generated by using standard simple JSON string as data source, whereas for example Google Chart Tools use proprietary DataTable Class as a data source. DataTable Class is also JSON but it is much more complex than the string used by Highcharts. (amCharts, FusionCharts, Google Inc b, Highsoft)

JSON that is given to Highcharts to create a line chart:
[{"name":"NetBooster Agency","data":[[1323126000000,479],[1323212400000,477]]},
{"name":"NetBooster Finland","data":[[1323126000000,288],[1323212400000,288]]},
{"name":"NetBooster Asia","data":[[1323126000000,540],[1323212400000,540]]},
{"name":"NetBooster Group","data":[[1323126000000,72],[1323212400000,72]]}]

JSON containing the same data given to Google Chart Tools to create a line chart:

{"cols":[{"id":"account","label":"Account","type":"string"},{"id":"time","label":"Time","type":"date"},{"id":"page likes","label":"Page likes","type":"number"}],
"rows":[
{"c":[{"v":"NetBooster Agency"},{"v":new Date (2011,11,06)},{"v":"479"}]},
{"c":[{"v":"NetBooster Agency"},{"v":new Date (2011,11,07)},{"v":"477"}]},
{"c":[{"v":"NetBooster Finland"},{"v":new Date (2011,11,06)},{"v":"288"}]},
{"c":[{"v":"NetBooster Finland"},{"v":new Date (2011,11,07)},{"v":"288"}]},
{"c":[{"v":"NetBooster Asia"},{"v":new Date (2011,11,06)},{"v":"540"}]},
{"c":[{"v":"NetBooster Asia"},{"v":new Date (2011,11,07)},{"v":"540"}]},
{"c":[{"v":"NetBooster Group"},{"v":new Date (2011,11,06)},{"v":"72"}]},
{"c":[{"v":"NetBooster Group"},{"v":new Date (2011,11,07)},{"v":"72"}]}
]}

After choosing the code library to be used to display charts it is necessary to imple-ment the required behaviour of the charts and configure their appearance and other options. In Prism behaviour includes updating of the charts when filters have been changed and changes to the chart settings according to the data displayed. Charts in Prism also include custom data point clicking behaviour. When data points are clicked all the mentions and their meta data that make up the data point are shown. As seen in Figures 15 and 16, most of the features of the charts can be changed including col-ours, legend information, axis information, line styles, data point dot styles, etc.
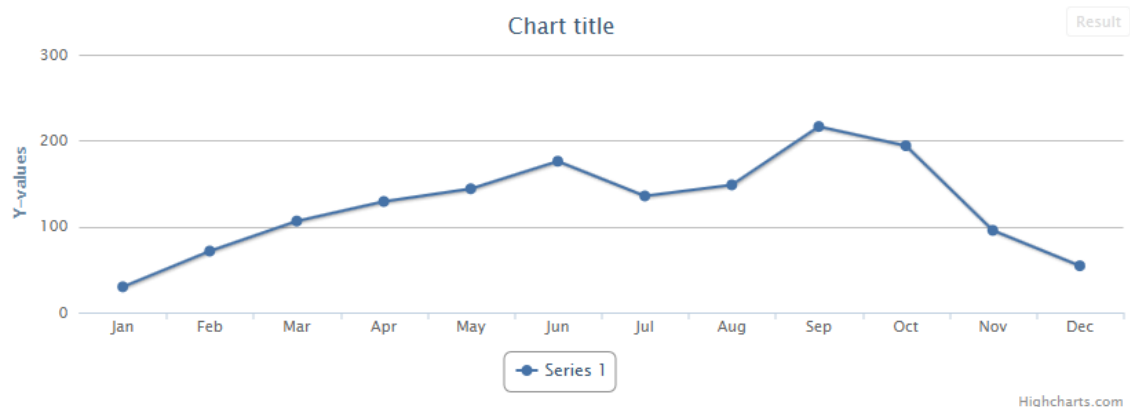
Figure 15.   Line chart with default configuration
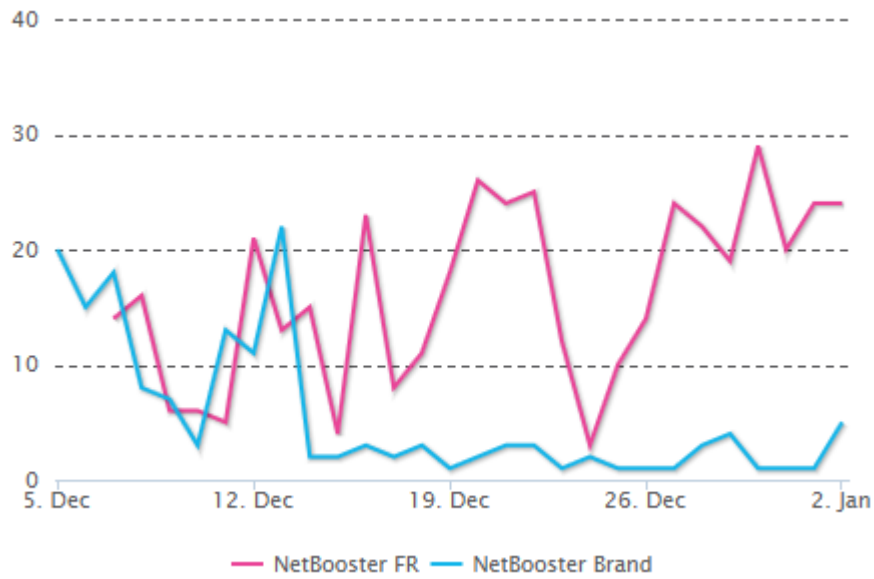
Figure 16.   Line chart with custom confugration

It is better to display different kind of data with different kind of charts which requires unique chart configurations for each data type. As seen in Figure 17, it is better to display sentiment data as stacked percentage line graph because it makes it easier to see the relation of each sentiment compared to each other instead of highlighting the total number of mentions. Colours such as red, yellow and green are also helpful to distinct the negative, neutral and positive sentiments from each other.
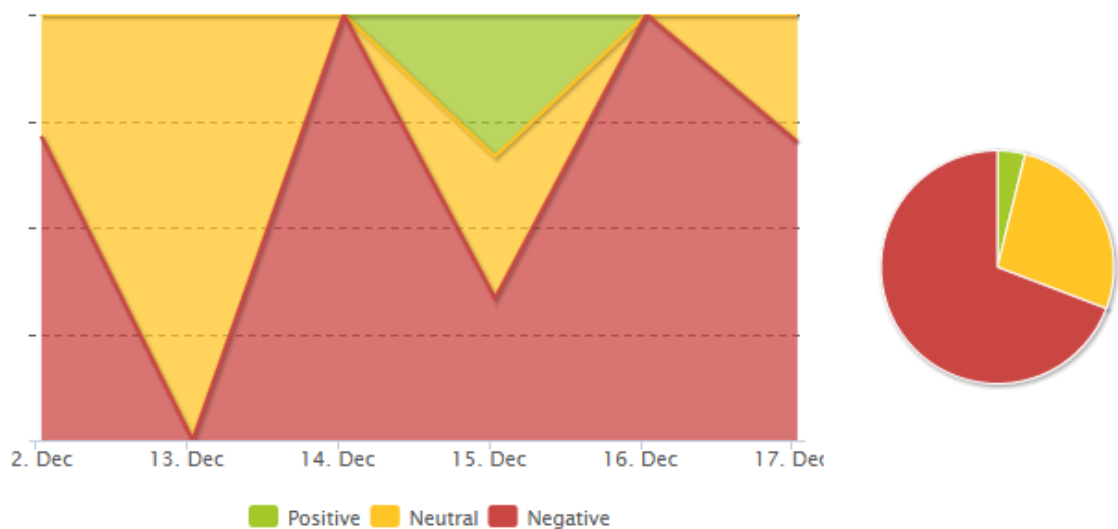
Figure 17.   Stacked percentage line chart and pie chart illustrating the results of sentiment analysis

In addition to Highcharts, Prism also uses Google Chart Tools to display charts. Google Chart Tools are used because they offer chart types that are not available in Highcharts and because Google Chart Tools is free software. Google Chart Tools are based on Flash and they do not work on browsers without a Flash plugin. Because not all users are able to view the charts, Prism uses Google Chart Tools only to show information that is also shown in charts generated by Highcharts. Google Chart Tools generated charts show the same information as the ones generated by Highcharts but in different format. Google Chart Tools has a chart called motion chart which includes bubble chart, bar chart and line chart. As seen in Figure 18, motion chart is actually an animated video that shows a snapshot of the data at a time.
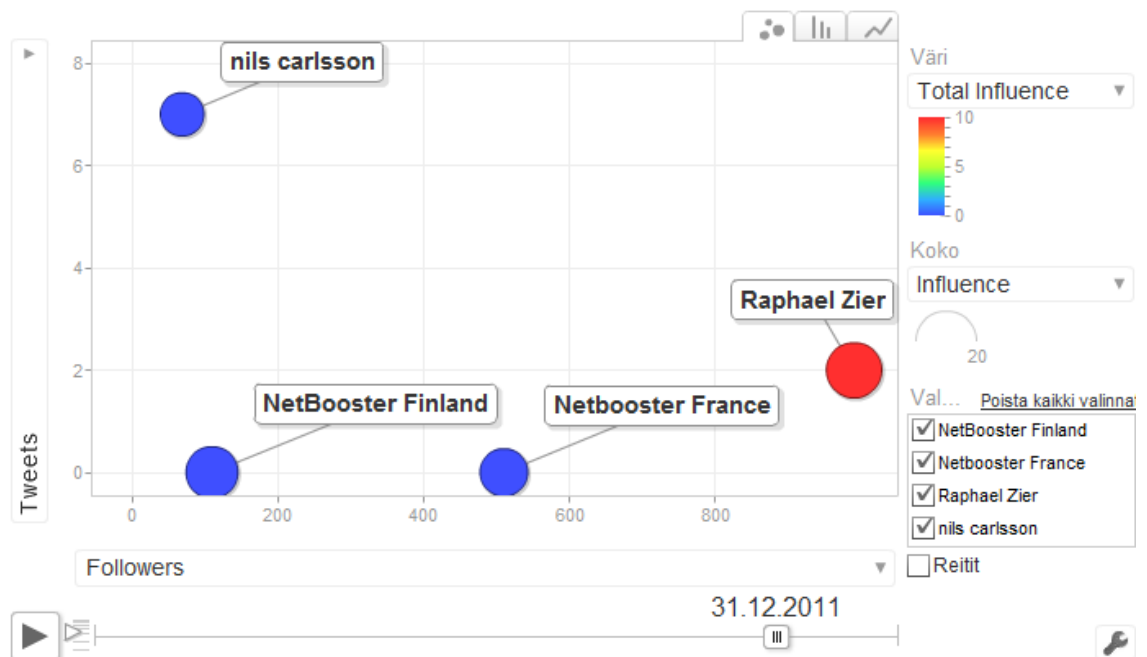
Figure 18.   Twitter statistics in Google Motion Chart

Charts are important in conveying complicated statistical information in easy to read format. There are many ways to show charts in browser environment but client side code libraries are an easy way to reach good results. There are multiple commercial and free chart code libraries available offering different set of features. Prism uses various chart types with different settings and two different code libraries to display statistical information.

8.2   Dashboard

Dashboard is the most important view of the Prism user interface. It is the first view that a client sees when he logs into Prism. As seen in Figure 19, dashboard includes a number of graphs that display the general trend of the online reputation of the client.

Figure 19.   Dashboard

By hovering the cursor over the graphs and clicking them the user is able to get more information about a specific metric and the dataset related to it. Currently the dashboard includes charts and a list of topics found recently. There are plans to include more textual information about the current situation of the online reputation.

8.3   Filtering

Charts offer an easy way to read large numbers of information fast and understand the meaning of this information to your business. Because charts display a lot of data at a same time it is impossible to zoom into some subset of this data and analyse this subset separately from the rest of the data displayed in the chart.

In Prism this problem is solved with filters. Filters are used to show only a subset of the data in all of the charts visible to the user. As seen in Figure 20 there are survey and date filters to set the date range filtering and select between different surveys. Each survey includes different campaigns which include different keywords thus creating totally separate data sets.

Figure 20.   Survey and date filters

As seen in Figure 21, there are many kind of filters available including categories, topics, campaigns, keywords, sentiment, websites, languages and even text search in the mention title and description.

Figure 21.   Filters to display only French websites

All types of filters can be combined and applied simultaneously. Filters use user inter-face elements such as checkboxes and autocomplete text fields to make filtering fast and intuitive. Chosen filters are saved to session and are immediately applied to all the charts and lists visible to the user. Because filters are saved in the session they are persistent across page loads. As seen in Figures 21 and 22 it is also possible to save, load filters and delete saved filters as well as reset current filters.

Figure 22.   Menu to save, load and delete filters

Prism offers many types of filters for advanced users to find the information they are looking for. Saved filters and default filters enable consultants to set up filters that of-fer easy access to specific data for the client.

8.4   Viewing Mentions

In addition to viewing charts generated by the mention data gathered and analysed by Prism it is also possible browse through the mention data itself. This can be done by navigating to the mentions page at the main navigation menu or by clicking a data point in any chart. Clicking a data point in a chart also applies filters needed to display only the mentions related to the clicked data point.

Figure 23.   Mention view

As seen in Figures 10 and 23 it is possible to see the picture, address and the category of the website and title, short description, campaign, keyword, publish date and topics of the mention. It is also possible to send the mention as an email by clicking the email icon or view the website as a PDF how it was when the mention was found. Consultants can also change the sentiment of the mention and delete it.

8.5 Administration

Prism has a number of settings that need to be set for each client in order to make Prism's automatic mention finding and analysing to work. Primarily keywords, campaigns, surveys, topics, categories and client's information need to be set and updated. This data is saved to Prism by using html forms that include different kind of user interface elements including text boxes, checkboxes and file upload fields. These forms are also used to view the information saved. As seen in Figure 24, some of this information can also be viewed in table format which includes a link to a form to perform viewing and editing. Administration user interfaces have simplistic graphical design that follows the design on the client's views. As seen in Figure 24, all administration views include a breadcrumb in the top of the screen to see what object an administrator is editing. Breadcrumb also offers links to the current objects ancestor objects.

Figure 24.   Administration form of a Survey

Administration views also include two more complicated features than basic edit forms. One of them is a feature that estimates the costs of a survey. After setting up a survey and all campaigns and keywords related to it Prism is able to estimate how many API requests are made each month and calculate the total price of those API requests. This task is initiated by pressing a button in a html form of a survey and the results will be emailed to the user who pressed the button. Estimating the survey costs can take several hours because Prism has to analyse all the mentions found in the past month and discard all the irrelevant mentions.

	A	B	C	D
1	Fr_Auto	Fr_Avis conso	Fr_Buzz / webmarketing	Fr_communiqués
2	www.forum-auto.com	fr.answers.yahoo.com	www.e-marketing.fr	communique-presse.index-net.org
3	fr.autoblog.com	60millions-mag.com	blog.x-prime.com	actu.sinuscom.org
4	www.voitures.net	assurland.com	cyrille-chaudoit.blogspot.com	actualites-web.com
5	www.autotitre.com	avisconso.com	fabiendenais.typepad.fr	articlesenligne.com
6	www.carbu.fr	avis-consommateur.fr	lavieenrosedejulio.com	aujourd-hui.com
7	www.autonews.fr	avistore.com	lesniffer.typepad.com	autopromopro.com
8	automobile.fr.msn.com	bestofchat.com	marketing-alternatif.com	big-presse.com
9	www.oliomobile.org	ciao.fr	nicknolteweb.blogspot.com	bloc.com
10	blogautomobile.fr	conso.net	nouveau-marketing.com	businessportal.com
11	blog.usinenouvelle.com	cyberacteurs.org	pierre-philippe.blogspot.com	businessportal24.com
12	www.aramisauto.com	domtomfr.com	prland.blogs.com	businesswire.com
13	classic-auto.fr	experatoo.com	renaud.perrin.free.fr	categorynet.com
14	zorgblogauto.canalblog.com	femininbio.com	roycod.com	communique24.com
15	blog.jpblogauto.com	forumactif.com	suchablog.com	communique-de-presse-gratuit.com
16	gazoline.net	forumfr.com	superfiction.net	communiques.indexweb.info
17	blogues.cyberpresse.ca	forum-voyage.fr	tendancemedia.free.fr	communiques-de-presse.fr
18	321auto.com	hurrycam.fr	viacomit.free.fr	communiquez.biz
19	auto.aol.fr	leguide.com	vincentabry.com	faire-le-web.fr
20	auto.orange.fr	lesarnaques.com	viralbuzzandco.blogspot.com	fr.pressbot.net
21	auto.yahoo.fr	mamzelleemie.com	vlancom.blogspot.com	generalite.com
22	autodeclics.com	opinion-assurances.fr	voxinablog.com	lesnewsdunet.com

Figure 25.   Part of an excel file used to import categories to a survey

Another complicated feature is the possibility to export and import Excel files that include website categories and all the websites included in them. This makes it easier and faster to copy website categories with lot of websites between surveys. Excel files, as seen in Figure 25 use a simple format with the category name on the first row of the spreadsheet and websites included in this category in the consequent rows of the column. Excel files are written and read by using PHPExcel library (Baker, M. 2011). When importing an excel file it is first analysed and the results are displayed to the user, as seen in Figure 26, who then confirms that the file is interpreted correctly. After confirmation the data is saved to the database.

Figure 26.   Data read from the Excel files waiting for confirmation

Administration views enable administrators and consultants to perform all the tasks required to set up Prism for new clients and to improve the quality of the surveys for example by administering the categories and found mentions. Using the administration features requires very little education and expertise and is easy to learn by doing.

9   Discussion and Conclusions

The objective of this bachelor's thesis was to document the decisions and steps taken on building an online reputation management system that replaces an older system with the same basic features.

The starting situation of the project offered a lot of opportunities as the first version of Prism outlined the basic features of the tool as well as the biggest problems that needed to be addressed. On the other hand it was a challenge because the new version had to be an improvement to the old one.

The new version of Prism was released in January 2012 after being under development for 7 months. Sales of the new version of Prism have been started and existing clients have started using it too. The release included all the major features that were required. After the release it was decided to keep on developing Prism further by adding more features such as video monitoring, key performance indicators Google+ integration and localization. At the time of writing Prism is only used in NetBooster France but the aim is to offer it also in other countries where NetBooster and its subsidiaries are operating and work on this has already been started. All in all the future of Prism is looking bright and eventful.

This bachelor's thesis successfully documents the biggest steps on building an online reputation management system as well as identifies the major problems related as well as the solutions that can be used to overcome them. The major technical challenges in building an online reputation management tool are the finding and analysing of the mentions as well as the user interface used to display the discovered and analysed information. The fifth chapter of the thesis introduces the problems and solutions related to finding the mentions. The sixth chapter introduces all the different metrics used in Prism to provide aggregated data of the mentions found. Also the technologies and the third party services used to calculate these metrics are introduced. The eight chapter introduces the user interface elements used in Prism and evaluates various chart libraries in the market.

This bachelor's thesis can be used as a reference for designing and implementing an online reputation management system. Even though the present thesis provides solutions to some problems it does not provide a walkthrough or technical implementation details for building an online reputation management system.

References

(2011) Twitter Sentiment (Nov) [WWW document]
https://spreadsheets.google.com/ccc?key=0Ajqw0BPB53J7dFZmTGhCeGFvNzBGazli
QXRlYnBfeFE#gid=0 (Accessed Dec 4 2011)

Agarwal, A. (2009). How to Identify the Language of any Text (May 5) [WWW document] http://www.labnol.org/internet/identify-language-of-text/8441/ (Accessed Nov 28 2011)

Alexa Internet Inc. (a) n.d. Alexa [WWW document] http://www.alexa.com/ (Accessed Nov 24 2011)

Alexa Internet Inc. (b) n.d. Alexa Internet – Toolbar Download [WWW document] http://www.alexa.com/toolbar (Accessed Nov 24 2011)

Amazon Web Services LLC, n.d. Alexa Web Information Service [WWW document] http://aws.amazon.com/awis/ (Accessed Nov 24 2011)

amCharts, n.d. amCharts: JavaScript/HTML5 charts [WWW document] http://www.amcharts.com/ (Accessed Jan 4 2012)

Baker, M (2011) PHPExcel (Mar 4) [WWW document] http://phpexcel.codeplex.com/ (Accessed Jan 10 2012)

Boutell (2012) GD Graphics Library (Jan 3) [WWW document] http://www.boutell.com/gd/ (Accessed Jan 4 2012)

buypennystocks (2012) Stocks to Watch – NetBooster SA (ALNBT), (Jan 30) [WWW document] http://www.buypennystocks.com/penny-stocks-blog/stocks-to-watch-netbooster-sa-alnbt (Accessed Feb 11 2011)

Byelex Multimedia Products B.V., n.d. OpenDover – Overview [WWW document] http://developer.opendover.nl/docs (Accessed Dec 7 2011)

Chen, S. C., n.d. PHP Simple HTML DOM Parser [WWW document] http://simplehtmldom.sourceforge.net/ (Accessed Dec 17 2011)

Cianciullo, J., n.d. About – Social Mention [WWW document] http://socialmention.com/about/ (Accessed Dec 20 2011)

Ciancuillom, J., (2010) Social Mention API Documentation, (Apr 5) [WWW document] http://code.google.com/p/socialmention-api/wiki/APIDocumentation (Accessed Nov 20 2011)

Cision, n.d. Print Media Monitoring and Press Clipping Service, [WWW document] http://us.cision.com/media-monitoring/print-monitoring.asp (Accessed Nov 14 2011)

Clarabridge, n.d. Clarabridge API [WWW document] http://www.clarabridge.com/Product/HowClarabridgeWorks/Implement/Clarabridge API.aspx (Accessed Dec 7 2011)

Cosenza, V. (2011). World Map of Social Networks [WWW document] http://vincos.it/world-map-of-social-networks/ (Accessed Jan 21 2012)

Econsultancy (2011) Online Reputation and Buzz Monitoring 2012 Buyer's Guide (Dec)

FusionCharts, n.d.FusionCharts [WWW document] http://www.fusioncharts.com/ (Accessed Jan 4 2012)

Google Inc. (2007) Google Terms of Service, (Apr 16) [WWW document] http://www.google.com/accounts/TOS (Accessed Nov 20 2011)

Google Inc. (a) n.d. Google Language API Family (Deprecated) [WWW document] http://code.google.com/intl/en-EN/apis/language/ (Accessed Nov 28 2011)

Google Inc. (b) n.d. Google Chart Tools [WWW document] http://code.google.com/intl/en-EN/apis/chart/ (Accessed Jan 4 2012)

Google Inc. (c) n.d. Unusual Traffic from your computer network [WWW document] http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=86640 (Accessed Nov 20 2011)

Google Inc. (d) n.d. What are Google Alerts? [WWW document] https://www.google.com/support/alerts/bin/answer.py?hl=en&answer=175925&topic=28415&parent=28413&rd=2 (Accessed Nov 20 2011)

Google Inc. (e) n.d. What is custom search? [WWW document] http://code.google.com/intl/en-EN/apis/customsearch/ (Accessed Nov 20 2011)

Google Operating System Blog. (2007) Google Search Feeds and SOAP API (Not from Google), (Jan 17) [WWW document] http://googlesystem.blogspot.com/2007/01/google-search-feeds-and-soap-api-not.html (Accessed Nov 20 2011)

Highsoft, n.d. Highcharts [WWW document] http://www.highcharts.com/ (Accessed Jan 4 2012)

Höhrmann, B., n.d. CutyCapt [WWW document] http://cutycapt.sourceforge.net/ (Accessed Dec 11 2011)

Kaushik, A. (2011). Best Social Media Metrics: Conversation, Amplification, Applause, Economic Value [WWW document] http://www.kaushik.net/avinash/best-social-media-metrics-conversation-amplification-applause-economic-value/ (Accessed Jan 21 2012)

LinuxManPages (2004) Convert [WWW document] http://linuxmanpages.com/man1/convert.1.php (Accessed Dec 11 2011)

Mashable (2008a) Top 10 Free Tools for Monitoring Your Brand's Reputation, [WWW document] http://mashable.com/2008/12/24/free-brand-monitoring-tools/ (Accessed Jan 22 2012)

Mashable (2008b) Top 10 Reputation Tracking Tools Worth Paying For, [WWW document] http://mashable.com/2008/12/29/brand-reputation-monitoring-tools/ (Accessed Jan 22 2012)

Maurice De Kunder (2011) The Size of the World Wide Web (The Internet) [WWW document] http://www.worldwidewebsize.com/ (Accessed Dec 20 2011)

Microsoft, n.d. Bing API [WWW document] http://msdn.microsoft.com/en-us/library/dd900818.aspx (Accessed Nov 20 2011)

NetBooster (2011a) About NetBooster Group [WWW document] http://www.netbooster.com/about-netbooster (Accessed Feb 11 2011)

NetBooster (2011b) The History of NetBooster Group [WWW document] http://www.netbooster.com/history (Accessed Feb 11 2011)

Netcraft Ltd. (2011) Web Server Survey [WWW document] http://news.netcraft.com/archives/2011/11/07/november-2011-web-server-survey.html (Accessed Nov 20 2011)

Nokia Corporation, n.d. Qt – Cross-platform application and UI framework [WWW document] http://qt.nokia.com/ (Accessed Dec 11 2011)

OpenAmplify., n.d. OpenAmplify – Social Integration [WWW document] http://www.openamplify.com/social_integration (Accessed Dec 7 2011)

Orchestr8 LLC (a) n.d. Company Information [WWW document] http://www.alchemyapi.com/company/ (Accessed Nov 28 2011)

Orchestr8 LLC (b) n.d. Concept Tagging [WWW document] http://www.alchemyapi.com/api/concept/ (Accessed Dec 7 2011)

Orchestr8 LLC (c) n.d. Language Detection : Web API [WWW document] http://www.alchemyapi.com/api/lang/urls.html (Accessed Nov 29 2011)

Orchestr8 LLC (d) n.d. Sentiment Analysis [WWW document] http://www.alchemyapi.com/api/sentiment/ (Accessed Dec 4 2011)

Orchestr8 LLC,(e) n.d. Sentiment Analysis : Web API [WWW document] http://www.alchemyapi.com/api/sentiment/urls.html (Accessed Nov 29 2011)

Owyang, J. (2011). Number of Fans and Followers is NOT a Business Metric –What You Do With Them Is (Nov 17) [WWW document] http://www.web-strategist.com/blog/2011/11/17/number-of-fans-and-followers-is-not-a-business-metric-what-you-do-with-them-is/ (Accessed Jan 21 2012)

ProgrammableWeb (2011) Web Services Directory [WWW document] http://www.programmableweb.com/apis/directory (Accessed Dec 20 2011)

Saplo AB, n.d. Text Analysis API [WWW document] http://saplo.com/api (Accessed Dec 7 2011)

Semantic Engines LLC, n.d. Sentiment Analysis API [WWW document]
http://semanticengines.com/sentimentapi.aspx (Accessed Dec 7 2011)

SimilarGroup Ltd. (2011) SimilarGroup API [WWW document]
http://similargroup.com/api.php (Accessed Nov 20 2011)

The PHP Group (a) n.d. Curl - Introduction [WWW document]
http://www.php.net/manual/en/intro.curl.php (Accessed Dec 17 2011)

The PHP Group (b) n.d. GD - Introduction [WWW document]
http://www.php.net/manual/en/intro.image.php (Accessed Jan 4 2012)

The PHP Group (c) n.d. ignore_user_abort [WWW document]
http://php.net/manual/en/function.ignore-user-abort.php (Accessed Dec 17 2011)

The PHP Group (d) n.d. register_shutdown_function [WWW document]
http://php.net/manual/en/function.register-shutdown-function.php (Accessed Dec
17 2011)

Third Door Media (2011) Google Begins Multi-Month User Experience Update, (Jun
28) [WWW document] http://searchengineland.com/google-begins-multi-month-
user-experience-update-83592 (Accessed Nov 20 2011)

Twitter Sentiment [WWW document]
http://twittersentiment.appspot.com/search?query=NetBooster (Accessed Dec 4
2011)

WebKit Team., n.d. The WebKit Open Source Project [WWW document]
http://www.webkit.org/ (Accessed Dec 11 2011)

Wikipedia (2010) Concept Mining (Sep 8) [WWW document]
http://en.wikipedia.org/wiki/Concept_mining (Accessed Nov 29 2011)

Wikipedia (2011a) Alexa Internet (Nov 22) [WWW document]
http://en.wikipedia.org/wiki/Alexa_Internet (Accessed Nov 24 2011)

Wikipedia (2011b) ASP.NET, (Dec 14) [WWW document]
http://en.wikipedia.org/wiki/Asp.net (Accessed Dec 14 2011)

Wikipedia (2011c) Comparison of web application frameworks, (Dec 14) [WWW
document]
http://en.wikipedia.org/wiki/Comparison_of_Web_application_frameworks (Ac-
cessed Dec 14 2011)

Wikipedia (2011d) cron, (Dec 15) [WWW document]
http://en.wikipedia.org/wiki/Cron (Accessed Dec 17 2011)

Wikipedia (2011e) Java (programming language), (Dec 14) [WWW document]
http://en.wikipedia.org/wiki/Java_%28programming_language%29 (Accessed Dec
14 2011)

Wikipedia (2011f) Language Identification (Nov 16) [WWW document]
http://en.wikipedia.org/wiki/Language_identification (Accessed Nov 28 2011)

Wikipedia (2011g) Named-entity Recognition (Nov 15) [WWW document]
http://en.wikipedia.org/wiki/Entity_extraction (Accessed Nov 29 2011)

Wikipedia (2011h) Natural language processing (Nov 10) [WWW document]
http://en.wikipedia.org/wiki/Natural_language_processing (Accessed Nov 20 2011)

Wikipedia (2011i) Online Reputation Management, (Oct 22) [WWW document]
http://en.wikipedia.org/wiki/Online_reputation_management (Accessed Nov 14
2011)

Wikipedia (2011j) Scaffold (programming), (Sep 22) [WWW document]
http://en.wikipedia.org/wiki/Scaffold_%28programming%29 (Accessed Dec 14
2011)

Wikipedia (2011k) Sentiment Analysis (Nov 29) [WWW document]
http://en.wikipedia.org/wiki/Sentiment_analysis (Accessed Nov 29 2011)

Wikipedia (2011l) Sina Weibo, (Nov 19) [WWW document]
http://en.wikipedia.org/wiki/Sina_Weibo (Accessed Nov 20 2011)

Wikipedia (2011m) Web Search Engine, (Nov 19) [WWW document]
http://en.wikipedia.org/wiki/Search_engine (Accessed Nov 20 2011)

Wikipedia (2011n) Zend Framework, (Nov 19) [WWW document]
http://en.wikipedia.org/wiki/Zend_Framework (Accessed Dec 14 2011)

Wikipedia (2011o)Text analytics (Nov 3) [WWW document]
http://en.wikipedia.org/wiki/Text_analytics (Accessed Nov 20 2011)

Wikipedia (2012p) Named-entity Recognition (Jan 3) [WWW document]
http://en.wikipedia.org/wiki/Chart (Accessed Jan 4 2012)

Yii Software LLC, n.d. Features of Yii [WWW document]
http://www.yiiframework.com/features/ (Accessed Dec 14 2011)

Prism version 2 functional requirements

Web monitoring

1. Client can see the amount of daily mentions for each campaign
2. Client can see the amount of daily mentions for each category
3. Client can see the weighted weekly index representing his brand's reach
4. Client can see the daily percentage of user generated content from all the content by campaign
5. Client can see the percentage of mentions for each campaign
6. Client can see the percentage of mentions for each category
7. Client can see the individual mentions that are included in the dataset by clicking a point in the graph
8. Client can see all the mentions and filter them by categories, campaigns, keywords, sentiments and topics
9. Client can see the publish date, title, description, content, address of the mention even if it was deleted from the original source
10. Client can see the daily amount and percentage of positive, negative and neutral weekly mentions
11. Client can see the weighted weekly index of positive, negative and neutral weekly mentions
12. Client can see the amount of positive, neutral and negative mentions for each topic and campaign
13. Client can see the amount of positive, neutral and negative mentions for each category and campaign
14. Client can see the amount of positive, neutral and negative mentions for each category and topic
15. Client can see a tag cloud containing the keywords most often related to mentions
16. Client can change the date range for all the shown data
17. Client can export all the shown data to excel
18. Consultant can send a mention as an email to the client when viewing a mention
19. Consultant can change the sentiment of a mention
20. Consultant can remove mentions manually.

Social media monitoring and engaging

21. Client can add and remove social media accounts
22. Client can connect his Google Analytics accounts to the application
23. Client can see the daily number of followers and fans for each of his social media account
24. Client can see the daily number of his followings
25. Client can see the daily number of comments, shares and retweets
26. Client can see the total number of users reached with his tweets and shares (number of followers + followers of retweeters)

27. Client can post messages to social media platforms
28. Client can see the daily amount of clicks and conversions for each social media platform and post

Administration

29. Consultant can add, edit and delete clients
30. Consultant can grant rights for clients to perform operations normally done by consultant
31. Administrator can add, edit and delete consultants and administrators
32. Consultant can add, edit and delete campaigns of each client
33. Consultant can add, edit and delete keywords of each campaign
34. Consultants can choose to automatically remove all the mentions where the keyword is mentioned only in a link anchor text for each campaign
35. Consultant can add, edit and delete topics (if not handled automatically)
36. Administrator can add, edit and delete categories (if not handled automatically)
37. Administrator can see usage and status reports of the application
38. Consultant can change the categorization of a website
39. Consultant can change the weight of a category (or a website if weight is calculated automatically)
40. Consultant can change the percentage of user generated content of a category
41. Client can login to the system
42. Consultant can exclude websites from keyword search
43. Consultant can keep or delete the mentions when deleting keywords or campaigns
44. Client can change the language of the user interface
45. Consultant can change the system logo shown to the client
46. Consultant can hide powered by NetBooster logo