D2.2 Tracking technologies and social attitudes (Update)

Project acronym:	**SENSE4US**
Project full title:	**Tracking technologies and social attitudes**
Grant agreement no.:	**611242**
Responsible:	**Juri Papay**
Contributors:	**All Partners in the SENSE4US Consortium**
Document Reference:	**D2.2-UPDATE**
Dissemination Level:	**PU**
Version:	**FINAL**
Date:	**13/10/15**

Project co-funded by the European Union under the Seventh Framework Programme

History

Version	*Date*	*Modification reason*	*Modified by*
0.1	28/08/14	Initial draft	Juri Papay
0.2	08/09/14	Contributions from project partners	Juri Papay
0.3	29/09/14	Revision	Juri Papay
0.4	29/09/14	Draft reviewed by project partners	Hansard Society and GESIS
0.5	01/10/14	Final edits	Juri Papay
1.0	06/10/14	Submitted version	Juri Papay
1.1	16/01/15	Revised version according to the review comments	Juri Papay
1.2	10/03/15	Internal revision	Steve Taylor
1.3	28/08/15	Restructured document, focusing on the problems that the policy maker needs to resolve	Juri Papay
1.4	28/09/15	Internal QA by Steve Taylor	Juri Papay
1.5	12/10/15	Koblenz's input	Juri Papay
1.6	13/10/15	Introduction updated	Juri Papay

Table of contents

List of tables

List of abbreviations

Abbreviation	*Description*
ALOE	Assisted Linked Data Consumption Engine
API	Application Programming Interface
DoW	Description of Work
LDA	Latent Dirichlet Allocation
LOD	Linked Open Data
Lemon	Lexicon Model for Ontologies
RDF	Resource Definition Framework
LOD	Linked Open Data
OR	Operational research
OWL	Web Ontology Language
PM	Project Month – PM1 is the first month of the project, etc.
PSM	Problem Structuring Methods
RDF	Resource Definition Framework
REST	Representational state transfer
SOFIE	Self-Organizing Framework for Information Extraction
SPARQL	Protocol and RDF Query Language
SVR	Support Vector Regression
WP	Work package

Executive summary

This deliverable reviews the state of art and provides a critical assessment of tools and technologies that can help to achieve the objectives of the Sense4us project. We also discuss how the individual work packages and research tasks can benefit from the presented information and how the project contributes to the state of the art.

1 Introduction

Sense4us is a three year project launched in October 2013 and co-funded under the Seventh Framework Programme (FP7-ICT-2013-10). Sense4us aims to assist policy makers by providing tools to access a wide array of current data, take into account the views of citizens on policy issues in real time and help them to better understand the implications of proposed policies. Policy-makers at the EU, national (UK) and local (Germany) levels are engaged in the development of Sense4us to ensure the toolkit has the widest possible application.

This deliverable reviews the state of art and provides a critical assessment of tools and technologies that can help to achieve the objectives of the Sense4us project. In this document we also discuss how the individual work packages and research tasks can benefit from the presented information and how the project can contribute to the state of the art.

The Description of Work (DoW) specifies the requirements for Task 2.2 as follows:

> *Task 2.2 – Tracking other technologies and social attitudes (M03-33)*
>
> *In order to keep the project relevant and to grasp opportunities where they arise, research will be carried out and reports produced that take account of new technologies, social attitudes towards technologies and the impact of other research in the SENSE4US space. The outcome of this research will be fed into the decision making process of the project.*

This deliverable represents an updated version of D2.2 that was submitted for the review in October 2014. This update considers the comments and recommendations of reviewers that are detailed below:

a) "We recommend choosing only such information directly related to the project's objectives, showing how these affect the research within the work packages, and integrating such findings into next WP2 deliverables."
b) "Additionally, the resubmitted deliverable should outline the elements that go beyond the existing state of the art while also developing approaches that take into account the needs of the end-user groups."

To address comment (a) above, in this update we have surveyed the main directions of research, key concepts, technologies and social attitudes that are relevant for the Sense4us project. The outcome of this report has been used for the assessment of various alternative solutions that arise during the functional and architecture design and also integrated in the research of individual work packages. The Sense4us project covers numerous research aspects, and in order to narrow down the multitude of directions we have focused on the tasks that are the core objectives of the project and also reflect the requirements for individual work packages. The tasks that we have identified are as follows:

a) Extract topics and summarise information from a document (s) and from social media (WP4)
b) Find data related to individual topics (WP4)
c) Calculate sentiment for the information extracted from social media (WP5)
d) Policy modelling and simulation (WP6)

In the discussion of relevant technologies and tools we clearly stated how the given technology or tool fits with the objectives of the WP and also how the other WPs can benefit from the adoption.

To address point (b) above, we have devoted Section 7 to summarise the contributions to the state of art that the Sense4us project has produced so far. There is still a year left before the end of the project but we can already provide evidence confirming these contributions. The list of contributions include: Sense4us LOD Search Tool (WP4), SentiCircles (WP5), Sense4us

Policy Modelling Tool (WP6). In WP3 we have developed a technique to architecture design that is based on a systematic analysis of end-user requirements. The tools described in this deliverable have been downloaded, tested and their usability and usefulness assessed according to the end-user requirements.

The document is organised as follows. Section 2 describes the tools and techniques related to topic extraction and information summarisation. Section 3 outlines techniques for finding related information from open data sources. Section 4 focuses on the state of art in sentiment analysis. Section 5 describes the technologies and tools that allow to develop and simulate policy models. Section 6 deals with social attitudes in respect to the policy making process. Section 7 describes the contributions by Sense4us that go beyond the state of art.

2 Topic Extraction and Text Summarisation

In this report we discuss topic extraction and text summarisation in the same section since they are both closely related. In case of topic extraction the key phrases are extracted from the text, in case of summarisation the key sentences are identified.

Topic extraction and text summarisation provide the user with a quick overview of a document without having to read it. When faced with a task of extracting useful information from a large number of documents these techniques enable to get an idea about the key topics and to prioritise the documents according to their relevance. These techniques also help the user to get familiar with the problem domain and also to identify the dominant parameters and relationships between them that can be used for constructing simulation models.

One of the advantages of automatic document summarisation is that it can produce an output without human intervention and therefore it is unbiased. However this output must be checked for accuracy and consistency before it is published (Mani 2014). In general there are two techniques used for producing a summary: extraction and abstraction (Mani 2014). The extractive approach selects a subset of existing words, phrases, or sentences in the original text for generating a summary. The abstractive approach builds a semantic representation of the text and then uses natural language generation techniques for creating a summary. In this case the produced summary might contain words that are not explicitly present in the original text. Abstractive methods represent a promising direction of research, but the majority of summarisation tools still use extractive methods. From a survey of several publications (Mani 2014), (Lloret 2010), (Nenkova 2011), a number of key requirements related to document summarisation have been identified:

a) dealing with the thematic diversity of a large number of documents
b) combining the main themes with completeness
c) readability
d) conciseness
e) reflecting the logical structure of the main content
f) dividing the output into meaningful paragraphs
g) consistent referencing of information sources.

In the Sense4us project we have used the above requirements as a guidance for assessing the multitude of topic extraction and text summarisation tools.

In the context of the Sense4us project topic extraction is very important since it is the starting point for various tasks such as finding related information (WP4), sentiment analysis (WP5), and policy modelling and simulation (WP6). The topics extracted from a large corpus of text (e.g. policy documents) are used as seeds in the form of search terms for many other tools provided by the research WPs. In WP4, the extracted terms are used for searching the Linked Open Data (LOD) for identifying related information. WP5 works on extracting, summarising and processing the information from various social media sources, and the seed search terms can be used for social media searches. WP6 focuses on the development of parametric simulation models that allow to perform various “what if studies” for assessing the possible consequences of policy decisions. Both topic extraction and text summarisation are important steps in the process of model building since allow to identify the parameters that can be incorporated in the simulation models.

There are numerous methods known from the literature that allow topic extraction (Blei, D. M, et al 2003), the main issue is however to make sure that the extracted topics actually reflect the semantic content of the document. One of the approaches for ensuring "semantic coherence is called "automatic topic labelling". By labelling we understand a selection of words that the best describe the semantics of the given topic (Cano 2014). The most generic approach to automatic labelling has been to use as primitive labels the top-n words in a topic distribution learned by a topic model such as LDA (Griffiths 2004), (Blei, D. M, et al 2003).

Research however, indicates that selecting the top terms is not sufficient for interpreting the coherent meaning of a given topic (Mei, Q., et al 2007). More recent approaches have explored the use of external sources for example DBpedia and Wikipedia for supporting automatic labelling of topics by deriving so called "candidate labels" by using lexical (Lau, J.H., et al 2011), or graph-based (Hulpus 2013) algorithms. The paper by Mei *et al* proposed an unsupervised probabilistic algorithm for automatic assigning of labels in a topic model (Mei, Q., et al 2007). The proposed approach was defined as an optimisation problem involving the minimisation of the Kullback–Leiber divergence (Kullback-Leiber 2014) between the given topic and the candidate labels while maximising the mutual information between these two words. It was proposed by (Lau, J.H., et al 2011) to label topics by selecting the "top n" terms by using different ranking mechanisms including point-wise mutual information and conditional probabilities.

Another frequently used technique for automatic labelling utilises external data sources such as dictionaries, thesauruses or specialised topic ontologies. Methods relying on external sources for automatic labelling have been described by Magatti *et al* (Magatti,D., et al 2009). The authors derived candidate topic labels using the Latent Dirichlet Allocation (LDA) algorithm (Blei, D. M, et al 2003) and the hierarchy obtained from the Google Directory service. Then the initial candidate labels were extended with the Open Office English Thesaurus. In their paper, (Lau, J.H., et al 2011), described a case study where the label candidates for topics were extracted from Wikipedia articles.

Since topic extraction is a crucial component of Sense4us project we have invested considerable effort to make sure that topic extraction is accurate and reliably reflects the content of documents. The first prototype was based on the Latent Dirichlet Allocation (LDA) (Blei, D. M, et al 2003), however we have soon discovered that the output did not meet our criteria. LDA is a probabilistic method and one of the problems was that the topics and the generated labels could change between individual runs and were also not satisfied with the semantic coherence of topics and the document. An in depth literature survey had lead us to an alternative technique that does not use probabilistic algorithms but relies on a graph representation of text. The essence of this technique is to calculate the semantic or lexical similarity between the text unit vertices of the graph and identify the dominant vertices. This alternative approach lead us to the TextRank algorithm (Tarau 2004) that will be assessed in Section 2.3.

In the following sections we survey the topic extraction and summarisation tools and assess their suitability for the Sense4us project. We consider Semantria (Semantria 2014), AlchemyAPI (AlchemyAPI 2014), Yahoo Content Analysis API (YahooAPI 2014), TextRank (Tarau 2004) and LODifier (I. Augenstein 2012). The information about these tools is presented in a tabular format according to specific criteria that are relevant for the project.

2.1 Semantria

Category	Description
URL	https://semantria.com
Problem to be addressed	Topic extraction
Describe how the tool or software component addresses the given problem	Semantria allows topic extraction from a single or a collection of documents. Semantria relies on Wikipedia's ontology for constructing a concept matrix by using a deep learning algorithm.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	a) Simple API that allows the formulation of queries, tagging, grouping and structuring the collection of derived topics. b) Provides sentiment analysis for the data extracted from the social media. c) The API can be accessed directly from Excel via a plug-in. Allows customisation of output by editing the categories and tags in the configuration file.
Licensing	Commercial license
Effort required for the integration	Due to the commercial licence this tool is not considered for integration.
Comments	The commercial licence prevents the use of Semantria in Sense4us project.

Table 1 –Semantria evaluation

2.2 AlchemyAPI

Category	Description
URL	http://www.alchemyapi.com
Problem to be addressed	Topic extraction
Describe how the tool or software component addresses the given problem	AlchemyAPI provides topic extraction, image tagging and sentiment analysis of web pages, documents, tweets and photos. AlchemyAPI uses neural network algorithms for the linguistic and statistical processing of large volumes of text. Frequently used in business environment for extracting knowledge about the customers, competitors, company operations, marketing, sales and new product developments.
Suitability of the tool or software component to be integrated into Sense4us.	a) AlchemyAPI provides language support for Python, PHP, Ruby, Java, Perl, C/C++, C#(.NET)

	b) Can be used directly via REST interface, or through a developer kit. c) Provides content enrichment with Linked Data (LOD) Resources d) Free access to one thousand transactions daily.
Licensing	Commercial licence available for processing larger volumes of data.
Effort required for the integration	Using the REST interface does not require significant effort.
Comments	AlchemyAPI is used in WP5 for sentiment analysis for extracting the topics from a large volume of tweets. The non-paid service allows one thousand queries a day.

Table 2 –AlchemyAPI evaluation

2.3 TextRank

TextRank (Tarau 2004) is a graph based technique that allows both topic extraction and summarisation. For key phrase extraction, TextRank constructs a graph with vertices representing sentences (or topics) and the edges are based on semantic or lexical similarity between the text unit vertices. The edges are typically undirected and can be weighted to reflect a degree of similarity. Once the graph is constructed, it is used to form a stochastic matrix, combined with a damping factor and the ranking over vertices is obtained by finding the eigenvector corresponding to eigenvalue 1 (i.e., the stationary distribution of the random walk on the graph). TextRank uses the idea that sentences "recommend" other similar sentences to the reader. Thus, if one sentence is very similar to many others, it will likely be a sentence of great importance. The importance of this sentence also stems from the importance of the sentences "recommending" it. Thus, to get ranked highly and placed in a summary, a sentence must be similar to many sentences that are in turn also similar to many other sentences. This makes intuitive sense and allows the algorithms to be applied to any arbitrary new text. The methods are domain-independent and easily portable. One could imagine the features indicating important sentences in the news domain might vary considerably from the biomedical domain. However, the unsupervised "recommendation"-based approach applies to any domain.

Category	Description
URL	https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf
Problem to be addressed	Topic extraction
Describe how the tool or software component addresses the given problem	TextRank is based on unsupervised learning that uses the structure of the text for determining the topics and the key phrases. The advantages:

	a) No training data required b) Textrank can work with any type of text. c) The algorithm is easily portable to new domains and languages d) Unlike LDA (which is based on entropy), the output from Textrank is deterministic. Users did not like LDA because it produced different results on the same text when run more than once. e) Textrank produces n-grams (n-word length extractions from the text) to describe the text. This is much more intuitive than the "bag of words" approach of LDA.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	a) suitable for topic extractions, the topics can consist of several words b) TextRank produces a high quality output that is consistent with the manual topic extraction and summarisation
Licensing	Redistribution and use in source and binary forms, with or without modification is permitted provided that the copyright notice is included.
Effort required for the integration	Small effort required, TextRank is implemented in Java, the source code is available (Nathan 2009).
Comments	**This tool has been tested by us, and we have begun integrating TextRank into Sense4us.** According to the literature this tool ranks high among the summarisation tools (Tarau 2004).

Table 3 – TextRank evaluation

2.4 LODifier

Category	Description
URL	http://link.springer.com/chapter/10.1007%2F978-3-642-30284-8_21 https://github.com/vimalkumarpatel/cs586
Problem to be addressed	Topic extraction
Describe how the tool or software component addresses the given problem	LODifier converts unstructured natural language into Linked Data (I. Augenstein 2012). LODifier uses deep semantic analysis, word-sense disambiguation, Semantic Web vocabularies for extracting entities and relationships between them. The output is translated into RDF representation where the individual nodes are linked to DBpedia and WordNet. This tool is especially useful for mapping between the terms found in e.g. social media analysis to URIs found in Linked Open Data, so can provide a link between the work of WP5 (social media analysis) and

	WP4 Linked Open Data search.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	The key features of LODifier are: abstracting away from linguistic surface variations, using a semantic graph for representing explicit structural information and linking up the concepts and relations to the LOD cloud (Augenstein, I et al 2012). The key operations performed by LODifier: a) Tokenisation of unstructured text b) Named entity recognition (Wikifier), produces DBPedia URIs c) Deep semantic analysis d) Lemmatisation e) Word sense disambiguation
Licensing	Free for academic use, no specific licensing information provided. Depends on a library that has a viral license preventing commercial exploitation[1].
Effort required for the integration	Available as a jarfile, can be run as a command line (https://github.com/vimalkumarpatel/cs586)
Comments	This tool covers several aspects of WP4 i.e. topic extraction and linking to LOD sources.

Table 4 – LODifier evaluation

2.5 Recommendations for the project

Since the early stages of the project we have been aware of the crucial role played by topic extraction and summarisation in relation to other activities. We have evaluated the tools according to several criteria focusing not only on the quality of produced results but also on the suitability for integration. In this respect we had to consider the complexity of the code or algorithm, maintainability and also the Licensing restrictions. In this respect Semantria due to its commercial licence was not considered further.

The next tool, AlchemyAPI, looked more promising for reasons that it is an online service with a REST interface and also it is suitable for working with large volumes of tweets. Although there is a commercial license but the online service allows a thousand queries a day free of charge that should be sufficient for the purposes of the project and to evaluate its potential use beyond the end of the project. The main issues were however the access and understanding of the code and permission for modifications. AlchemyAPI uses a neural network and also requires training data. The internal working of this neural network and the statistical processing are not known to us, we cannot influence the quality of results that produces in case the company decides to update the online service or the code release.

LODifier looks as a promising tool especially from the perspective of WP4 since it can translate a text into Linked Data (I. Augenstein 2012), however we found little information

[1] http://svn.ask.it.usyd.edu.au/trac/candc/wiki/Licence

regarding the quality of results produced by this tool[2]. A further issue was that only the jarfiles were published and we had no access to the source code.

TextRank is an algorithm that is publicly available and it can be implemented in various programming languages. The algorithm itself falls in the category of unsupervised learning hence it requires no training data unlike AlchemyAPI. This is a universal algorithm that can work according to the literature with texts in any languages (Tarau 2004). We have tested TextRank for English and German texts and checked the results against the human produced list topics and summary. The results were much better than the ones produced by the LDA technique that we had used at the beginning of the project. In addition, the end users preferred the output of TextRank because it contains phrases from the documents that make it much easier to understand than the output of the LDA topic analysis, which are unordered bags of words. The positive results obtained from our tests have also been confirmed by the literature documenting extensive evaluation of TextRank against the other techniques (Tarau 2004). We have begun integrating a Java implementation of TextRank into the toolkit.

[2] In case of TextRank there was plenty of evidence detailing the quality of produced results against the other techniques.

3 Finding data related to individual topics

There are several data resources considered by the Sense4us project, these are: Open Data, social media, academic research, policy-makers' own data and policy documents. The overall objective of "finding related information" (mainly WP4) is to provide the user with additional information that is relevant for a given subject area. In the context of the Sense4us project, Linked Open Data (LOD) is the major source of information.

The original concept of Linked Open Data was introduced by (Berners-Lee 2006). This concept was derived from the Semantic Web (Berners-Lee, T. et al 2001), which proposed an enhancement of the Web and the resources published on the Web with more expressive semantics. The authors suggested that this semantic augmentation could be achieved by establishing and describing hyperlinks between Web resources. Heath and Bizer (Heath 2011) in their paper summarised the benefits of Linked Data as follows:

> "*Linked Data provides a unifying data model a standardized data access mechanism, hyperlink-based data discovery and self-descriptive data*".

Finding the relevant data on the internet involves several stages: the first is identifying the nature of the data that we are searching for (e.g. determination of the search terms), the second is querying various data sources, the third stage is integrating the data and the fourth is visualising the data in a way that is meaningful for the user. Each of these stages involves several issues that need to be addressed by Sense4us. The seed for the data search comes from topic extraction and text summarisation (see Section 2), but LOD searches may also reveal additional information and associated search terms. As a possible strategy we can start with a general search for example in DBpedia. Then check the results that the search produces and identify new terms for in depth searches.

The information that is published on the web mainly text and finding data that can support policy making is not an easy task. The reason is that the data is often hidden in databases and the data is not yet fully addressed by the search engines. In principle publishing data on the web is as easy as publishing HTML pages. However there are general recommendations described in Tim Berners-Lee's paper that allow the data to be found (Berners-Lee 2006):

a) use URIs as names for things;

b) use HTTP URIs so that people can look up those names;

c) provide useful information, using the standards (RDF, SPARQL); and

d) include links to other URIs.

Finding the relevant data is the main objective of WP4 that focuses on Linked Open Data (LOD) search. This work is highly relevant for WP2 and WP6. WP2 represents the interests of end users who try to collect data for supporting various aspects of policy documents. The LOD searches can be iterated and repeated using information discovered in one search as the input to another, and so the information can be built up by the end user, increasing their understanding of the policy subject area and its environment, as well as providing input for building the causal model. For WP6, the search produces input for constructing policy models.

In Sense4us project we can identify four main themes in the context of finding relevant data:

a) integration of heterogeneous data sources

b) data linking
c) information discovery
d) visualisation of linked data

The term "*data integration*" refers to the problem of combining data from different sources and providing the user with a unified view of this data (Lenzerini 2002). The main issues in this respect are the differences in structures, formats and also differences in semantics when same concept is represented by different URIs. With the growing volumes of data available on the Web, it has been recognised that the traditional data integration approaches, which assumed that the data stored in relational databases, are inadequate for the challenges of the Web. This led to the emergence of so called "*dataspaces"* which represent an abstraction for data integration. The generic vision of a dataspace is based on the following principles (Franklin, M., et al 2005):

a) providing support for a wide variety of data formats
b) offering various levels of service depending on the type of query and parameters of data sources
c) integration of data sources.

The above principles and the knowledge accumulated in "dataspaces" research provided a conceptual framework and an additional input for WP4 for identifying the requirements and formulating them according to the objectives of the Sense4us project.

A good survey of dataspace solutions can be found in (Hedeler, C., et al 2010). In their paper, (Cafarella, M., et al 2011), described two Google projects: WebTables and Google Deep Web Crawler, which investigated the problem of structuring information published on the web. The WebTables project compiled a large collection of databases by crawling the Web and trying to combine the tables that are embedded in HTML pages. The Google Deep Web Crawler attempted to extract data from the so called Deep Web that is made up by a large number of Web forms. These two projects tried to harvest the vast amount of data on the Web that cannot be found by the traditional web crawlers.

The objective of "*data linking"* is to establish hyperlinks between related data items that are stored in different data sources. Once the data is linked, it can be accessed by using Semantic Web browsers in a similar way to the traditional web browsers. The difference is however, that in the case of the semantic search the users follow RDF links for navigating between data sources. The RDF links can also be followed automatically by specialised robots and semantic search engines (Heath 2011).

In relation to data linking there are several initiatives that Sense4us can benefit from, these are Microformats (Microformats 2014) and Linked Data (LinkedData 2014). The aim of Microformats (Microformats 2014) is to extend the web with structured data. The idea is to define a set of simple data formats that can be embedded into HTML via class attributes. The data items that are included in HTML pages via Microformats do not have their own identifier. This makes linking the data across documents and Web sites a major issue. The Linked Data initiative promotes a set of best practices for publishing and connecting structured data on the Web. Linked Data is typically published either as static RDF files on a web server, embedded into HTML files or via database access (relational databases or RDF triple stores).

An essential requirement for "*information discovery*" is that the URIs of resources can be looked up by HTTP clients. Linked data can be consumed in different ways. It can be crawled and integrated by specialised web-spiders for providing a consistent view of the data (see LDSpider (Isele 2011)). Data can also be consumed by de-referencing only the URIs that are currently in use by the given application. Linked Data applications using this method are typically Linked Data Browsers. Techniques for on-the-fly dereferencing are presented in Hartig et al., (Hartig, O., et al 2009).

Visualisation of linked data is one of the critical requirements of making sense of the collected information (Halevy 2012). Visualisation is usually based on the hierarchical graph structures that allow not only navigation but also better understanding and summarising of large amounts of data (Wills 2009), (Ell 2011). The LESS framework enables end-to-end integration of linked data from multiple sources based on parameterised data and queries (Auer 2010). LESS also uses a template mechanism which allows flexible integration of different formats, such as text, plots, graphs, maps, etc. The visualisation used in the Fusion project (Araujo 2010) allows mappings between the source ontology of linked data and the application ontology for the processing and visual presentation of information. Various demos of linked data visualization have been implemented in the TWC portal which demonstrates examples of linked governmental data (Ding 2011). The presentation layer of the TWC LOGD portal (Tetherless World Constellation (TWC) and Linking Open Government Data (LOGD) uses mash-ups based on the Google visualization API. In the Fusion Tables project Google also provides a framework for data integration and visualization (Gonzalez 2010).

In the following sections we survey assess data tools that represent interest in the context of the Sense4us project. The surveyed tools are the following: Silk (Volz, J. et al 2009), LIMES (Ngomo, A. et al 2011), SERIMI (Araújo 2011) and RiMOM (Li, J. et al 2009). Details about these tools is presented in a tabular format according to specific criteria that are relevant for the project. In the assessment we have considered several criteria such as functionality, license, access to source code etc.

3.1 Silk

Category	Description
URL	http://silk-framework.com/
Problem to be addressed	Finding related data
Describe how the tool or software component addresses the given problem	Silk is an open source software that allows to integrate data coming from different sources (silk 2015). The main components of Silk (Volz, J. et al 2009) are: a)link discovery engine that computes links between data sources b) evaluating and fine tuning the data links c) protocol for maintaining data links between continuously changing data
Suitability of the tool or software component	In addition to the above features further

to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	benefit is the Link Specification Language (Silk-LSL) for describing the heuristics for deciding whether a semantic relationship exists between two entities. Silk discovers or suggests connections between two RDF data sets based on string similarity.
Licensing	Apache Software License
Effort required for the integration	Silk is written in Python and there is also an implementation in Scala. The code can run from the command line that can be launched from Java. The integration does not require big effort.
Comments	Silk is already used in the project, in WP4 for automated link detection.

Table 5 – Silk evaluation

Given two RDF data sets, Silk analyses the resources in both data sets and produces as output a set of RDF statements connecting resources of both data sets. Such process can be guided by a configuration file in which an expert user defines the data sets and the the type of resources to be connected, the kind of typed links to create and the heuristics that Silk should consider to determine whether two resources should or should not be connected (A. J. Robert Isele 2010). Alternatively, Silk can learn such linkage rules automatically, after a set of reference links are labeled by a human user (C. B. Robert Isele 2013). Together with LIMES (LIMES 2015) and KnoFuss (KnoFuss 2015), Silk is one of the state-of-the-art technologies for data interlinking including learning-based link discovery. As advantages we can mention that Silk supports various types of links (i.e. it is not limited to owl:sameAs links), it is open source and extensible (C. B. Robert Isele 2013).

3.2 CROWDKI (Crowd-powered Knowledge Integration) *(DEVELOPED IN SENSE4US)*

Category	Description
URL	https://github.com/criscod/CROWDKI
Problem to be addressed	Finding related data
Describe how the tool or software component addresses the given problem	CROWDKI is a crowd-powered approach to knowledge integration, which aims at supporting data publishers in designing new interlinking processes, as well as validating and enhancing automatically computed links.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	The interlinking tools for example Silk, Limes produce reasonable results when they work with similar data sets. However they cope with less success when they deal with heterogeneous web of data and consequently the output may contain many errors that can be resolved only by human intelligence. We suggest to use CROWDKI as a refinement tool

	that can weed out the errors from the output of search tools.
Licensing	CC BY-SA (Creative Commons Attribution ShareAlike) license
Effort required for the integration	Source code written in java and the integration does not require significant effort.
Comments	

Table 6 – CROWDKI evaluation

CROWDKI (Crowd-powered Knowledge Integration) (CROWDKI 2015) is the approach developed by UNIKO (UNIKO 2015) to crowdsource the validation of links between RDF resources (see D4.2.1 for further details). CROWDKI is able to retrieve data coming from different data sets by following and consuming such links.

Both Silk and CROWDKI have been extensively tested and they have been identified as tools suitable for the Sense4us project. These tools are used for retrieving data and creating links between data sets. The purpose of CROWDKI is to reduce the number of incorrect e.g. *owl:sameAs* links generated by Silk. This is important, because if incorrect *owl:sameAs* links are published, the semantic search in Sense4Us will retrieve incorrect information.

3.3 LIMES

Category	Description
URL	http://aksw.org/Projects/LIMES.html
Problem to be addressed	Finding related data
Describe how the tool or software component addresses the given problem	LIMES is a link discovery software for large volumes of Web Data (LIMES 2015). a) uses time-efficient approximation techniques for calculating the similarity between instances. This is achieved by significantly reducing the number of comparing during the mapping process. b) implements supervised and unsupervised machine-learning algorithms for finding accurate link specifications c) The queries are described in LIMES Specification Language (LSL)
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	The main advantage of the tool is the speed of processing large knowledge bases. The tool uses the triangle inequality for calculating the distances, this allows to cut the number of links to be evaluated. The major drawback of LIMES is that it can be used only for metric spaces (where all distances are defined).
Licensing	LGPLv2.1 for non-commercial purposes

Effort required for the integration	The software is available only as a jarfile, that can be easily integrated in the java code.
Comments	No source code is available only the class files have been distributed. The "triangle inequality" technique needs to be considered especially if the speed LOD search becomes an issue.

Table 7 – LIMES evaluation

3.4 SERIMI

Category	Description
URL	https://github.com/samuraraujo/SERIMI-RDF-Interlinking
Problem to be addressed	Finding related data
Describe how the tool or software component addresses the given problem	SERIMI (Araújo 2011) is a tool for automatic RDF data interlinking.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	For linking two datasets SERIMI uses the instance matching algorithm consists of a selection and disambiguation phases. The main benefit is that for the interlinking no prior knowledge of the data, domain or schema of these datasets is required. The instance matching using partial string matching is not implemented yet and the metrics for evaluating the quality of matching is also to be refined.
Licensing	LGPL license
Effort required for the integration	SERIMI is implemented in JRuby version, as a command line it can be launched from Java.
Comments	There is no information on the scalability of the tool for large datasets.

Table 8 – SERIMI evaluation

3.5 RiMOM

Category	Description
URL	http://keg.cs.tsinghua.edu.cn/project/RiMOM/
Problem to be addressed	Finding related data
Describe how the tool or software component addresses the given problem	RiMOM (RiMOM 2015) is based on ontology mapping. Key features: a) automatic combination of multiple strategies for improving the matching effectiveness. b) calculation of similarity characteristics of

	different ontologies c) alignment across ontologies for archiving semantic interoperability
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	The RiMOM tool in the context of Sense4us would make sense if we searched a data domain that can be described by one or several ontologies.
Licensing	MIT license
Effort required for the integration	RiMOM is available as a jarfile.
Comments	We can see only a limited use of RiMOM in Sense4us since the data domain that we search are no ontologies developed.

Table 9 – RiMOM evaluation

3.6 Sense4us - LOD Search Tool *(DEVELOPED IN SENSE4US)*

In this section we describe the Sense4us Linked Open Data (LOD) Search Tool developed by University of Koblenz-Landau.

The Linked Open Data cloud is the result of movement from multiple data providers to the data freely available for everyone (LOD 2015). The data sets published within the LOD cloud are not only open, but also connected. The combination of data from governmental, geographic, live science, or linguistic topics makes the LOD cloud a valuable source for information. All data sets in the LOD cloud are described by the Resource Description Framework (RDF) (RDF 2015), a data model that allows data providers to offer data in a structured, machine-readable form that enables automated processing.

There are multiple systems that focus on exploratory search of LOD in general (A. G. Nuzzolese 2013), (H. S. Waitelonis 2010). There are also tools that are more specific, they allow searching and retrieving subgraphs from LOD data sets that describe relatedness of entities (P. Heim 2010), (L. Fang 2011).

Aemoo (A. G. Nuzzolese 2013), (Musetti 2012) is a tool that connects multiple data sources. This tool uses DBpedia as a source to identify connections between items from different data sources, but Aemoo offers no ranking.

LED (Noia 2010) is a tool suitable for exploratory search of the Web, it uses Dbpedia to explore keyword semantics. Yovisto (J. S. Waitelonis 2010) uses different versions of DBpedia for exploratory search of videos. All three tools focus solely on Dbpedia as LOD data set. Discovery Hub (N. Marie 2004) offers exploratory search on different data sets on-the-fly, but calculates ranking with an iterative approach.

For tools that find structures like paths or sub-graphs to describe the relation between entities, all of them (P. Heim 2010), (L. Fang 2011) utilize a single knowledge base (Yahoo! (P. Heim 2010) or DBpedia (G. Kasneci 2009), (L. Fang 2011) which has to be local for calculations of the relatedness and rankings.

The LOD Search Tool aims to capture information about two or more entities and identify unknown concepts and related stakeholders. The tool achieves this by accessing the Linked Open Data cloud, retrieving so called connectivity structures from data sets that contain information about the given entities, and ranking them based on information-theoretical approach. Connectivity structures are sub-graphs of one or more combined data sets that connect the given entities with each other. The LOD search tool constructs the sub-graphs by retrieving all paths between each combination of entities and ranking them. More detailed information is given in (Pirró 2015). The tool can handle any available data set within the LOD cloud that is accessible via SPARQL endpoint. A SPARQL endpoint offers extended usability to the data set by providing an extensive query language that the tool uses.

The difference between Sense4us LOD search tool and other systems is the ability to access any LOD data set to retrieve sub-graph structures between given entities and rank them according to their relatedness without extensive pre-processing of data or iterative calculations for rankings (see Table 10).

Category	Description
URL	http://www.sense4us.eu/index.php/work-overview#WP4
Problem to be addressed	Finding related data
Describe how the tool or software component addresses the given problem	The LOD Search Tool extracts information about multiple entities and identifies the most important concepts and stakeholders.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	The tool can work with any data set that provides SPARQL access. It allows to discover so called "*unknown concepts*" that are not obvious from the given contexts but nevertheless they represent important factors. The tool generates connectivity structures and ranks them according to their relevance.
Licensing	Covered by the consortium agreement
Effort required for the integration	The tool is written in java, packaged as a jar file. The prototype version of the tool is already integrated in the Sense4us platform.
Comments	The preliminary results produced by the tool are very promising. One of the main tasks is to present the information to the user in a way that is easy to understand.

Table 10 – Sense4us - LOD Search Tool evaluation

The system has contributed to the state of the art by (Pirró 2015). It contributes to the Sense4us project by providing a process to access LOD data sets and retrieve information about the relatedness of entities reaching over possibly multiple data sets to. The given explanation in form of a sub-graph can be further utilised, either by presenting the result to the user to give an extended insight into a topic or for additional processing like automated stakeholder extraction.

3.7 Recommendations for the project

One of the essential requirements for finding useful data is the credibility of the source. The value of data obtained from these sources becomes apparent for modelling (WP6) where often numerical data is required in order to calculate the correlation between parameters.

From the tools presented in this section the Silk tool and CROWDKI (Crowd-Powered Knowledge Integration) represent a particular interest for the Sense4us project. Silk allows to integrate data coming from different data sources and also establishes links between the data. As it happens with all data linking tools there might be errors that cannot be resolved automatically and for weeding out these errors a human intelligence is required. This is the point when CROWDKI comes into the picture by using crowd sourcing techniques for processing the links manually and assessing the relevance and correctness of links in different contexts.

A further recommendation is to develop metrics that allows to compare and assess the quality of information found by different LOD search tools. University of Koblenz has also produced a tool that contains some advance features for example: aggregation of SPARQL endpoints, entity linking and ranking of connections, crowd sourcing etc. All these features aim at improving the quality and relevance of links.

4 Sentiment Analysis

The objective of sentiment analysis in the context of the Sense4us project is to extract evidence from social media that can support the policy making process and to get feedback on certain topics. This involves determining public opinion in regard to specific policies, data mining of social media, along with the development of mechanisms and tools for harnessing this rich content.

Recent studies have shown a strong correlation between the online opinion of individuals and their offline attitude towards political issues in the real world (Hussain 2011). Hence tracking public opinion on various social media platforms can provide important insights for policy makers. By collecting information we can get a picture of the citizens' attitudes towards policies as well as positive and negative arguments. According to The IBM Center for The Business of Government "*Next Four Years: Citizen Participation*" more and more people are turning to social media to express their opinion on various topics and react on events (Citizen Participation 2012). Today, some social networking sites have managed to attract millions of active users, and a significant portion of all internet traffic (Traffic 2014). Social media has become a rich source of content and an easy and quick way to reach out to millions of people. While social networks are known to be highly accepted among younger generations, they are also gaining popularity with older generations (Madden 2014).

Twitter sentiment analysis has attracted much attention due to the rapid growth in Twitter's popularity as a platform for people to express their opinions and attitudes towards a great variety of topics. Approaches to Twitter sentiment analysis tend to focus on the identification of sentiment of individual tweets (tweet-level sentiment detection).

The information found in social media has certainly got the potential to improve the quality of evidence required for drafting a policy document (Leavy 2013). There are however, several concerns about the quality of data and especially using this source of information in the context of policy making or for assessing the public reaction to various policies. One of the concerns is the lack of information about the participants voicing their opinion on the social media (Wheeler 2012). The recent data collected by (Fernandez, M. et al 2012) show that only a small percentage of users generate most of the content, in real terms about 6% generate more than 36% of all collected tweets.

From the policy maker's perspective the main objective is assess the public reaction on various policy decisions. A succinct summary of social media information enables the policy maker to discover the public's reaction to various aspects of the policy, and unexpected reactions may be especially useful if they uncover issues that have not been previously considered. The policy maker may also gain a picture about the reaction on similar policy documents from the past an incorporate it in the current policy document.

It is also important to mention that a large portion of tweets are generated by professional "opinion forming" organisations such as news agencies and think tanks rather than individual citizens. This finding, although it is only a preliminary study, raises many questions regarding the use of information coming from the social media for assessing public opinion. From this we may conclude that in the policy drafting process we should be fully aware of the limitations of information coming from the social media especially regarding its quality, coverage and representativeness of the public opinion.

Most existing approaches to sentiment analysis focus on classifying the individual words into subjective (positive or negative) or objective (neutral). They can be categorised by supervised approaches and lexicon-based techniques. Supervised methods are based on training classifiers from various combinations of features such as word "*n-grams*" (Pak 2010), Part-Of-Speech (POS) tags (Agarwal, A., et al 2011) and tweets syntax features such as hash-tags, re-tweets, punctuations, etc. (Kouloumpis 2011). These methods can achieve good accuracy, however, training data is usually difficult to obtain especially for continuously evolving subject domains for example in Twitter (Liu, Sentiment analysis and subjectivity 2010). Supervised learning approaches require training data for sentiment classifier learning. In Twitter, training data are typically obtained by either assuming that tweets' polarities (positive, negative, neutral) can be inferred using emoticons or by taking consensus from the results returned by the sentiment detection websites. Moreover, supervised approaches are domain-dependent and require re-training with the arrival of new data. Given the great variety of topics that constantly emerge from Twitter, these limitations affect the applicability of such approaches (A. Go 2009), (Paroubek 2010), (H. Saif 2010).

On the other hand, lexicon-based approaches do not require training data (Bollen 2011). Instead, they use lexicons of words weighted with their sentiment orientations to determine the overall sentiment of a given text. These approaches have shown to work effectively on conventional text. However, traditional lexicons tend to be ill-suited for Twitter data, which often contains a large number of malformed words and colloquial expressions. Moreover, many lexicon-based approaches also make use of the lexical structure of a sentence to determine its sentiment, which becomes problematic in Twitter, where ungrammatical sentences are very common due to the 140-character length limit. Aiming to overcome these limitations, several sentiment analysis models and lexicons were developed to specifically work on Twitter data. Nevertheless, similarly to other lexicon-based approaches, Twitter-based approaches face two main limitations. Firstly, they confined with the fixed set of words that appear in the sentiment lexicons they employ. Words that do not appear in the lexicon are often not considered when analysing sentiment, which may create a problem when dealing with Twitter data, where new expressions and jargons constantly emerge. Secondly and more importantly, Twitter-based methods and the like offer fixed, context-independent, word-sentiment orientations and strengths. Although training algorithms have been proposed to optimize the terms' sentiment scores in the sentiment lexicons, it requires frequent retraining from human-coded data, which is labour-intensive and domain dependent (M. Thelwall 2012). Examples of methods based on sentiment dictionaries are SentiWordNet (sentiwordnet 2014), MPQA subjectivity lexicon (Wilson 2005) and SentiStrength (sentistrength 2014).

Another approach to sentiment analysis is based on ontologies and natural language processing techniques capturing the conceptual representations of words (Saif, H. et al 2012). The results produced by conceptual semantic approaches are better than the output produced by syntactical approaches, however they are often limited by the size and scope of knowledge bases (Cambria 2013).

Sentiment analysis was also one of the key components of two previous projects WeGov (WeGov 2014) and ROBUST (ROBUST 2014). In the WeGov project (WeGov 2014) several generic discussion analysis methods were implemented. These methods allowed the analysis

of the type of content and social features and also identified the behavioural roles of participants. In the ROBUST project (ROBUST 2014) the behavioural models were extended by applying these models to larger communities. In both WeGov and ROBUST projects, the analysis was focused on capturing and understanding the state of discussion at any given point in time.

Finding public opinion is closely connected to WP2, WP4 and WP6. In WP2 the main stake holder is the policy maker who tries to get some information about the public reaction to a policy's objectives or opinions about the policy subject area in general. The emphasis here is on the representation of large volume of information sourced from the social media. WP4 provides a list of topics that can serve as seeds for searching social media. Constructing a suitable query that can produce relevant information is one of the critical factors that need to be investigated in depth, and also finding means of assessing the relevance of results from the search. The modelling effort in WP6 focuses mainly on the integration of key parameters representing a given problem domain, however we can envisage a model that would allow to simulate the public opinion and give forecasts for different policy decisions.

In the following sections we provide a survey of tools that represent interest for the Sense4us project, these are SentiStrenght (SentiStrenght 2015), SentiWordNet (sentiwordnet 2014) and Sentiment Circles (Saif, H. et al 2012).

4.1 SentiStrength

SentiStrength (SentiStrenght 2015) is a lexicon-based method for sentiment detection on the social web. This method overcomes the common problem of ill-formed language on Twitter and the like, by applying several lexical rules, such as the existence of emoticons, intensifiers, negation and booster words (e.g., absolutely, extremely). One limitation of SentiStrength is its static sentiment values of terms, regardless of their contexts. Although SentiStrength comes with an algorithm to update the sentiment strength assigned to terms in a lexicon, this algorithm requires training from manually annotated corpora. Another common problem with SentiStrength is the full dependence on the presence of words or syntactical features that explicitly reflect sentiment. In many cases however, the sentiment of a word is implicitly associated with the semantics of its context (e.g., "great" is positive in the context "smile" and negative in the context "problem").

Category	Description
URL	http://sentistrength.wlv.ac.uk/
Problem to be addressed	Calculate sentiment
Describe how the tool or software component addresses the given problem	The sentiment is calculated based on an emotion detection algorithm that uses a neural net. The default sentiment word strength list is a collection of 298 positive terms and 465 negative terms.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	a) various scoring options are available: binary (positive/negative), trinary (positive/negative/neutral) and single scale (+/-4)

	b) originally developed for English but can be configured for other languages and contexts c) adaptability – the sentiment word strength list can be modified by a machine learning algorithm to optimise the sentiment word strengths d) automatic spelling correction, emoticon, slang and idiom lookup tables
Licensing	MIT license
Effort required for the integration	The integration requires small effort since the tool comes as a jar file.
Comments	

Table 11 – SentiStrength evaluation

4.2 SentiWordNet

Category	Description
URL	http://sentiwordnet.isti.cnr.it/
Problem to be addressed	Calculate sentiment
Describe how the tool or software component addresses the given problem	SentiWordnet (sentiwordnet 2014) is an automatically generated lexical resource that assigns three scores indicating “positivity”, “negativity”, and “neutrality” to each word in Wordnet.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	For calculating the sentiment the tool uses: semi-supervised learning and a random-walk algorithm for refining the sentiment scores. Since there are three scores for each word we obtain a finer grain sentiment. This can be used for validating sentiment calculations.
Licensing	SentiWordNet is distributed under an Attribution-ShareAlike 3.0 Unsupported (CC BY-SA 3.0) license.
Effort required for the integration	The output of this tool is a text file of words with three sentiment scores. This can be used for checking the sentiment results. This requires little effort.
Comments	SentiWordnet has web based graphical interface.

Table 12 – SentiWordNet evaluation

4.3 SentiCircles *(DEVELOPED IN SENSE4US)*

SentiCircles is a novel approach developed by the Sense4us project. This is a novel approach to sentiment calculations and contributes to the state of art with the following:

a) SentiCircles aim to represent the sentiment orientation of words with respect to their contextual semantics. The main notion behind this is that the sentiment of a term is not static, as in traditional lexicon-based approaches, but rather depends on the context in which the term is used, i.e., it depends on its contextual semantics (Saif, H. et al 2012).

b) Graphical representation of sentiment with a circle. The two upper quadrants of the circle have positive sentiment with upper left quadrant representing stronger positive sentiment. Similarly, terms in the two lower quadrants have negative sentiment values. Moreover, a small region called the "*Neutral Region*" can be defined. This region is located very close to X-axis in the "*Positive*" and the "*Negative*" quadrants only, where terms lie in this region have very weak sentiment.

c) SentiCircles can be used in the following two sentiment analysis tasks: a) *Entity-level Sentiment Analysis* which aim to detect the sentiment of a given named entity, b) *Post-level Sentiment Analysis* which allows to detect the overall sentiment of a given post.

Category	Description
URL	http://2014.eswc-conferences.org/sites/default/files/papers/paper_93.pdf
Problem to be addressed	Calculate sentiment
Describe how the tool or software component addresses the given problem	The tool represents new lexicon-based approach for sentiment analysis on Twitter. The main idea is to use a dynamic representation of words that captures their contextual semantics (i.e., semantics inferred from the co-occurrence patterns of words in text) in order to tune their pre-assigned sentiment strength and polarity in a given sentiment lexicon. The words that co-occur in a given context tend to have certain relation or semantic influence, that is captured by the SentiCircle approach (Saif, H. et al 2012).
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	The tool uses: a)Contextual semantics that is more accurate that the static semantics b) suitable for the sentiment analysis of tweets, three methods are used: median, pivot and hybrid. c)the tool has been extensively tested on numerous datasets: OMD, HCR and STS-Gold (al. 2013) and produced outstanding results
Licensing	MIT licence
Effort required for the integration	Minimal effort, the tool is packaged as a jar file only a wrapper is required.

Comments	Tool developed by the project partner – KMI, and has been integrated into the prototype.

Table 13 – Sentiment Circles evaluation

SentiCircles is dynamic technique it also uses a lexicon as the above mentioned static methods such as SentiStrength and SentiWordNet, however it captures the contextual semantics of words in order to update their sentiment orientation and strength, and also allows conceptual semantics to be added to enrich the sentiment analysis task.

In respect to CentiCircles we need to mention that for the Sense4us project KMI has developed a web-based platform that uses AlchemyAPI for entity (for example persons, locations and organisations) extraction and visualisation. We have selected AlchemyAPI due to high performance in comparison with other entity extraction tools. The input for the platform is a collection of tweets. The output is a list of entities and their associated concepts extracted from the tweet collection. Information of these entities is then added to the SentiCircles generation process.

4.4 Recommendations for the project

There are several recommendations that aim to improve the quality of information sourced from the social media, these are:

a) Extract or elicit information about the contributor's identify, location affiliation etc.
b) Introduce advanced filtering mechanisms that allow to obtain high quality formation and remove for example commercial adverts, the tweets of large media companies etc.

One of the key issues in sentiment analysis is the identity and affiliation of the entity that contributes to social media discussions. This information would enable to assign weights to different opinions according to the affiliation. At present this information is very sparse therefore we suggest to develop methods that allow to elicit at least some information. Even more important would be to find out more about the opinion of the given entity on various topics. One recommendation is to provide analysis of social media user types, so that the commercial organisations can be highlighted or filtered out, for example.

It is also important to mention that a significant portion of tweets are generated by professional "opinion forming" organisations such as news agencies and think tanks rather than individual citizens. This raises many questions regarding the suitability and limitations of information coming from the social media for assessing public opinion especially regarding quality of information, coverage and representativeness.

5 Policy Modelling and Simulation

The main objective of modelling is to examine the effect of policy options on different stakeholders and their reaction to this effect, e.g. to answer questions of the form: "what if we provided more charging stations in our town? Would this increase the take up of electric cars?". The model construction itself is an iterative process that can be described by several stages such as: of specification of scenarios, identification of key parameters and relationships between the parameters, specifying the range of parameters, validating the output of the model against the initial assumptions and information collected using tools from other WPs. The modelling work in (WP6) interacts with all work packages. WP2 uses the model for evaluating the effects of various policy decisions and modifying the draft policy document accordingly. WP4 provides the raw data and the list of parameters to be used for model construction. WP5 supplies data representing citizens' sentiment as a reaction to various policy decisions.

The literature describes decision making as a process of identifying and choosing alternatives based on the considerations and preferences of the decision maker. The objective of this process is to reduce the number of alternatives so that a favourable outcome can be selected. This is an iterative process which requires the continuous checking of initial assumptions, parameters and outcomes (Harris 2012). Decision support tools and techniques enable the policy maker to predict the consequences of a policy before they are enacted. With these tools the user can carry out "*what if*" studies that allow a given policy to be analysed from the perspective of different stakeholders and for the likely effects on society and groups of individuals (Jiwani 2010), (Mitchell 2009).

One of the key components of a decision support system is *modelling* which incorporates several activities. In relation to modelling the following activities can be identified (Druckenmiller et al 2009), (Mintzberg 1994):

a) describing the scenarios
b) identifying the key variables, ranges of variables and uncertainties
c) describing the stakeholders and preferences
d) specifying the assumptions of the model
e) specifying the mathematical model or a simulation tool
f) interpreting of the output
g) providing feedback to the modelling process.

We have used the above list as a guidance for evaluating the functionality of various modelling tools.

5.1 Modelling techniques

In the literature we can find various approaches to modelling, an overview of various techniques can be found in (Mingers 2004). In the context of Sense4us project the following techniques have been considered:

a) Operational research and Problem Structuring Methods (PSM)
b) Multi-Criteria Decision Analysis
c) Multi-attribute utility theory
d) Portfolio decision analysis

e) Game theory.

Operational Research (OR) is one of the well-established methodologies used for modelling and decision support. There are numerous tools developed in OR that can be used for the evaluation of alternative scenarios, see details in (Weimer and Vining 2005), (Gass 2011), (OR 2014). There are also advanced techniques that aim to extend the traditional OR methods by taking into account qualitative data in the modelling process. These techniques are referred to as Problem Structuring Methods (PSM) (Mingers 2004). The traditional OR mainly works with well-structured problems, the PSM on the other hand tackles problems which are unstructured and contain many uncertainties. Typical features of unstructured problems are multiple actors, multiple perspectives, conflicting interests, important intangibles and large number of uncertainties (Mingers 2004). A detailed account of PSM is beyond the scope of this report, therefore we just mention the typical examples of PSM techniques (J. Rosenhead 1989):

- Strategic options development and analysis (SODA) (Eden 2000)
- Soft systems methodology (SSM) (Connell 2001)
- Strategic choice approach (SCA) (Friend 2001)
- Drama theory (Rosenhead 1996)
- Viable systems model (VSM) (Harnden 1990)
- Decision conferencing (Phillips 1989)

The information available for modelling can be either quantitative or qualitative. The quantitative information can be integrated into mathematical expressions while the qualitative information is essential for the conceptualisation and formulation of the model (Sterman 2000). Needless to say it is harder to deal with qualitative data since there are generally no accepted rules for obtaining, analysis and interpreting this type of information. The problems involved with qualitative information can be summarised by the following quotation from (Luna-Reyes et al 2003):

> *"Although there is general agreement about the importance of qualitative data during the development of a system dynamics model, there is not a clear description about how or when to use it. The lack of an integrated set of procedures to obtain and analyze qualitative information creates, among several possible problems, a gap between the problem modeled and the model of the problem."*

Another approach to modelling is based on Multi-Criteria Decision Analysis (MCDA). This method is widely used in complex decision problems for evaluating various alternatives leading to the achievement of multiple objectives (Keeney R.L 1993). One of the assumptions of this technique is that the alternatives are defined at the beginning of the modelling process and they do not change. However, in practice these alternatives are often difficult to specify and they can also change as the modelling work progresses and more information becomes available (Franco L. A. 2011).

A recently developed method based Multi-attribute Utility Theory proposes an alternative approach for integrating both quantitative and qualitative data into the same model (Danielson, M., et al 2006). The main idea of Portfolio Decision Analysis (PDA) is based on the selection of a portfolio of alternatives (instead of a single one only) which satisfies certain resource constraints, for example costs, budget etc. (Fasth 2012).

The application of Game Theory is one of the promising directions of policy modelling and decicion making support and this is one of the key elements of the policy modelling and simulation module of the Sense4us system. Game theory in general terms can be described as the "*the study of mathematical models of conflict and cooperation between intelligent rational decision-makers*" (Myerson 1991). This theory is used in numerous fields such as economics, politics, management, defence, biology etc. The objective is to design game strategies that can produce the desired outcome. The essential steps in Game theory based simulations are the design of the game matrix that captures all possible outcomes and calculating, identifying the game pattern and calculating the Nash equilibrium (Wikpedia 2015). The Nash equilibrium in essence represents an outcome of a non-cooperative game in which each player can follow their own strategy regardless of the state or progress of the game. In Game theory a multitude of patterns have been developed for modelling negotiations, auctions, environmental legislation, etc. The concept of the Nash equilibrium can be used for analysing the outcome of the interaction of several policy makers. This case can be representative if the interests of more departments or interests groups is taken into account the policy drafting process. Each stakeholder group is a player in the game and can act independently, and has different win/lose criteria. The participants might also have conflicting interests, in this case these conflicts can be mitigated by repeated interactions and modifications of interests and introduction of cooperation opportunities.

5.2 Modelling Tools

There are numerous design tools that allow to construct conceptual and simulation models for assessing the consequences of various policy decisions. These tools can be classified as follows (tools 2014):

- Information Control - gathering, storage, retrieval, and organisation of data, information and knowledge
- Paradigm Models - paradigms, frameworks or perspectives for conceptualising the model
- Simulation Models - models that provide answers to "*what if*" questions;
- Ways of Choosing - reducing the number of alternatives
- Representation Aids - visualisation of data or problem space

We have used this classification as a roadmap for getting oriented in the multitude of modelling tools.

In the context of the case study (“green energy”) investigated by the Sense4us project we can mention an award winning tool “*2050 carbon calculator*” developed by the *UK Department for Energy and Climate Change.* The tool allows users to simulate the outcomes of different energy policy scenarios, and the impact of their choices on the climate (calculator 2014). This climate calculator produces emission reduction pathways and demonstrates the impact of real scientific data. The tool also allows a user to check whether the data included in policy documents supports the long term objectives on emissions reduction. A further example of a similar tool, extended for the entire world is called the *Global Calculator* which allows a user to estimate the impact of the world’s energy, land and food systems on the climate (Global 2014).

In the following sections we consider several modelling tools that represent interest for the Sense4us project. There are several reasons why we opted for these tools:

a) they all support causal mapping that allows to model vaguely defined problems that are typical for policy mode
b) these tools can work with qualitative information
c) provide sophisticated graphical interfaces allowing to define the key factors and the relationships between them

Causal maps can be developed by individual decision-makers to model the structural systemic elements of their situation and show how change is propagated through the system. "*Causal maps provide a visual, mental imagery based simulation of the system's behavior for system analysis and social communication*" (Laukkanen 2015).

The tools that we have selected are: CMAP3 (CMAP3 2015), Decision Explorer (DExp 2015), STELLA (STELLA 2015) and Gambit (Gambit 2015). The list also includes the tool developed by Stockholm University in charge of WP6. This tool introduces several novel features and contributes to the state of art (see Section 5.2.5).

5.2.1 CMAP3

Category	Description
Problem to be addressed	Policy modelling and simulation
URL	http://www2.uef.fi/fi/cmap3
Describe how the tool or software component addresses the given problem	Provides a framework for processing and editing causal maps (identifying and synthesizing concepts and causal links) from multiple users' input.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	CMAP3 (CMAP3 2015) – Comparative and composite causal mapping .The tool allows quantitative system dynamics simulations, system performance analysis and prediction. Support for composite causal mapping methods: a) low or semi structured approaches based on primary data acquired by interviewing or text-writing and coded/standardized by inductive/ interpretive methods b) templates using a list of concepts c) argument maps based on data
Licensing	No specific Licensing information provided, the author mentions that the tool is free for academic research.
Effort required for the integration	The categories and concepts used by the tool need to be customized for addressing the issues of policy analysis and decision support, this requires considerable effort. There is no graphical tool for visualizing the conceptual

	map.
Comments	No source code available, only binary executable.

Table 14 – CMAP3 evaluation

5.2.2 Decision Explorer

Category	Description
Problem to be addressed	Policy modelling and simulation
URL	http://www.banxia.com/dexplore/
Describe how the tool or software component addresses the given problem	Decision Explorer (DExp 2015) can model so called "soft" issues" where the problem is described by qualitative information rather than exact numerical data. This tool can also be used for brainstorming for capturing the key concepts and to outlining the scope of the problem.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	Decision Explorer is a commercial tool that allows to handle thoughts and ideas that surround complex or uncertain situations. The main components of the tool are: a) Graphical concept map ("3D molecular model". b) Memo cards for storing additional information c) Extensive graph manipulation functions
Licensing	Commercial licence from Banxia Software Limited.
Effort required for the integration	Not considered for integration since it is not an open source product.
Comments	No information provided whether the tool provides support for simulation and "what if" studies. Basically it is a sophisticated graphical tool that allows to draw concept maps.

Table 15 – Decision Explorer evaluation

5.2.3 STELLA

Category	Description
Problem to be addressed	Policy modelling and simulation
URL	http://www.iseesystems.com/softwares/stella-pro/v1.aspx
Describe how the tool or software component addresses the given problem	Stella is sophisticated interactive tool for constructing and simulating system models.

Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	STELLA (STELLA 2015) is a concept mapping tool that allows exploring ideas, generating insight and conduct "what if" studies. The product features include: a) Live Analytics – allows dynamically explore model behaviour by changing the parameters of the model. b) Explore Results – for each entity in a tabular or graph format c) Sophisticated graphical tool, Bezier Connectors for avoiding overlapping lines. d) Animation of simulations e) Monte Carlo simulations and statistical calculations
Licensing	Commercial licence only
Effort required for the integration	Not considered due to the commercial licence.
Comments	The reason for reviewing STELLA was that it represents the top of the range tool and it contains numerous features and approaches from which we can learn in Sense4us.

Table 16 – STELLA evaluation

5.2.4 Gambit

Category	Description
Problem to be addressed	Policy modelling and simulation
URL	http://www.gambit-project.org
Describe how the tool or software component addresses the given problem	Gambit (Gambit 2015) is an open-source library of tools in game theory. Game theory plays an important role in the simulation phase of causal maps by driving the interaction between the components of the model.
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	Gambit allows to construct, analyse and explore various game models. Gambit provides a Python API, a command line interface and also a simple graphical interface for game design. Limitations of Gambit: a) for finite games only b) provide support non-cooperative game theory, i.e. the rules of the game are written down explicitly, and in which players choose their actions independently c) analysing large games can be time

	consuming d) Gambit is not the best option for game theoretic model verification. Because the competitive decision-making situation is different from a policy making. The players, moves and rules differ according the mental model defined by the user for the policy problem.
Licensing	Free/Open Source software, released under the terms of the GNU General Public License, Version 2. GPL is a viral licence we cannot use it in the project.
Effort required for the integration	A wrapper is required for the Python application.
Comments	The current Sense4us prototype is written in Java, code written in Python can be invoked on a command line.

Table 17 – Gambit evaluation

5.2.5 Sense4us - policy modelling and simulation tool *(DEVELOPED IN THE PROJECT)*

Category	Description
URL	http://www.sense4us.eu/
Problem to be addressed	Policy modelling and simulation
Describe how the tool or software component addresses the given problem	Causal maps of policy models
Suitability of the tool or software component to be integrated into Sense4us. State the benefits that can be obtained by integrating the tool.	a) Reflects the systemic nature of complex policy problems for which a regulation/policy needs to be based on a view of the system as a whole. b) Provides a visual problem model that can bring together different policy actors, clearly communicate their thoughts and can be used as a contextual framework that highlights knowledge gaps, guides information searching and models the search results. c) The prototype provides a fully computerised implementation of the model building, scenario triggering, scenario simulation and game theoretic computations.
Licensing	Tool is developed by the project partner – Stockholm.
Effort required for the integration	A simple wrapper is required.
Comments	The prototype version of the tool is already

	available, developed in WP6. This tool has been extensively tested by the project partners.

Table 18 – Sense4us policy modelling and simulation tool evaluation

The modelling approach developed by the Sense4us project is based on "Causal mapping and situation formulation", (Acar, W., 1983). This is a problem structuring method with the following modifications and enhancements:

a) defining the causal links based on causal inferences extracted from verbal description of the problem and quantifying change transfer using time-series from trusted data sources, instead of merely estimations by the decision-maker. The output is a mathematical model that identifies influences and trends based on reliable historical data to produce forecasts.
b) linking to game theory concepts to perform competitive analysis for the involved actors. The gaming approach uses real decision-makers/human agents combined with the simulation model.
c) creating scenarios of change in terms alternative futures and alternative courses of action (policy options).
d) policy models are constructed by using three main categories, these are: *actors*, *variables*, and *change transmission channels* (*links*)
e) graphical representation of complex problem situations using causal maps
f) multicriteria decision analysis framework that provides an in-depth performance evaluation of policy options, a method to elicit priorities and preferences from stakeholders

The key concept of causal maps as they are used in Sense4us can be described as follows:

A qualitative analysis of the causal mapping model can show the opportunities for policy interventions to achieve targeted changes in impact variables defined as policy objectives. The policy options represent different combined and controlled changes in the system inputs to produce the targeted outcomes. By quantifying these changes, scenarios of change can be defined as a combination of specified percentage changes occurring at a specific time point or at successive time points. Quantifying the policy goals in the form of a goal vector defined for each actor allows the analysis of these scenarios with respect to goal achievement. In addition, structural analysis of the causal map, can support scenario analysis (e.g., reachability analysis shows the ability of a scenario of change triggered at an independent variable to achieve a particular goal if the goal variable is reachable from this variable).

5.3 Recommendations for the project

One of the main issues of modelling is to establish the relationship between individual parameters. A first step in this direction is to use simple statistical methods that produce a single number such as Pearson's (Pearson 2015), Spearman's (Spearman 2015) or Chi-square (Chi 2015) coefficients for calculating the level of correlation between various parameters. The functions that describe the relationships between the parameters can be complex and often difficult or impossible to derive. However for the initial version of the model it would be sufficient to find out if there is any proportionality between the parameters. The only restriction is that the above mentioned statistical methods can only be used if numerical data

is available. The correlation coefficient ranges between -1 and +1, if the coefficient is close to zero, the two parameters are independent and have no effect on each other.

A recommendation is to quantify the causal links by numerical values as much as possible and monitor how they change during the model simulating process, but this will be limited by the information available. The model bulding tool should be able to indicate to the user which links have more reliable link coefficients, and this can be as simple as allowing the user to colour code links – green links could be the ones with user has the greatest confidence in, for example. This would allow to capture some extreme values that indicate fragility of the model. For example if one of the numerical values on the causal link is not changing smoothly and abruptly changes its value it indicates the model is not well balanced and hence not realistic. If we looked at the above mentioned tools, CMAP3 and STELLA do not quantify causal links. Stella is a quantitative systems dynamics simulation tool that provides strong visualisation capabilities. It uses a set of flow diagrams templates and is highly dependent on data and on functions for defining the relations between the model variables. This tool is more suitable for researchers than for policy makers that have to deal with a large number of qualitative factors.

A further recommendation for the tool is to provide strong support for qualitative data and a mechanism for dealing with uncertainties. This is especially beneficial for the policy drafting process for which complexity, insufficient data / knowledge and uncertainties are inherent features. Modelling in this case requires simplifications and assumptions on the key factors and relationship between them. As a result we obtain an exploratory model that allows to explore the implications of varying assumptions and hypotheses. In this case the user is motivated to collect as much data as possible in norder to limit the number of experiments that are required to answer a policy question.

6 Social attitudes

6.1 Definition

In this section we give a short overview of the social attitudes of the public to policies and to the policy making process in general. We also discuss the ways that the public can engage and contribute to policy making by participating in debates, providing feedback and suggestions to decision makers. The term, "*social attitude*", is a wide concept and perhaps can be captured by the following quotation:

> *"Social attitude is a person or groups reaction to other people, races, cultures, ideas, or traits. Social attitude measures the person or groups like or dislike toward a certain subject."* (Answers 2014)

The heterogeneity of individuals and groups within society will naturally result in different reactions to policies. Some policies may change the behaviour and preferences of social groups or individuals.

6.2 Engaging the public in policy making

The participation of the public in politics is regularly measured and evaluated. According to statistics (Politics 2014) the overall interest of public in policy making process is low. In terms of policy making it is important to consider the behavioural characteristics of social groups and individuals. Regarding behaviour, the following observations can be made (Politics 2014):

- people display '*bounded rationality'* since they do not have access to the data and have not got the time for complex calculations and reasoning;
- people's preferences change over time – it also often happens that short term preferences are not in line with long term goals;
- people exhibit reciprocity and value fairness.

For engaging the public in policy related debates according to DEFRA's study the following so called "*4E*" requirements should be addressed (Collier 2010):

- Encouraging - giving the right signals, incentives and disincentives ensuring that the target audience responds;
- Enabling - putting in place the capacity, infrastructure, services, skills, guidance, information and support needed;
- Engaging - getting people involved; and
- Exemplifying - leading by example and sharing responsibility.

6.3 Public perceptions of the policy making process

The latest figures from the 2014 Hansard Society report into political engagement highlight the following (Audit 2014):

- 50% of the British public say they are *'very'* or *'fairly'* interested in politics;
- 50% of the British public say they feel they know *'a great deal*' or *'a fair amount*' about politics;
- 49% of the British public say they are certain to vote in the event of an immediate general election; and
- 33% of the British public think that the system of governing in this country works well.

Despite the relatively high level of interest in politics, the policy making process itself to a large extent has been disconnected from the public. In other words there is little if any input from the public into the policy making process. It is also difficult to measure the feedback from the public regarding the policies that have already been implemented. As one of the benefits of the rapid development of social media, it is expected that the public will have a greater say in the policy making process. However, there are still barriers that prevent policy makers from becoming engaged with the citizens in the policy making process. These aspects can be summarised as follows (Collier 2010):

a) citizens feel that they are not well enough informed about the way decisions are made;
b) they also think that they do not have much influence over decision making;
c) most of the information which the public receive and react to is delivered via traditional media such as television and news;
d) the news items are interpreted by professional journalists and often published with editorial bias; and
e) the supporting evidence and data are not presented or are difficult to access. This makes the validation of policies difficult if not impossible.

6.4 Recommendations for the project

In terms of assessing the social attitude we'd like to recommend to consider not only the information collected from social media but also taking into account how the social attitude had changed in the past and over a longer time period. As with many thing in the society we might discovery cycles that allow to not only to interpret the current social attitude but also make a forecast for the future.

For assessing the impact of policies and the likely reaction of public it is important learn from the past. By analysing the social attitudes and reactions of the public in the past towards similar policies we can assess the likely impact of alternatives. To illustrate this point we can mention an example of university fees. It was well known from the past that the public would not welcome any university fees, however by introducing favourable loan schemes and conditions the social attitude has changed in a positive direction towards this policy (Ormston 2015). Further reason for this shift in the positive direction was massive publicity campaign and numerous

7 Sense4us contributions to the state of art

In this section we describe the contributions to the state of art that the Sense4us project has produced so far. There is still a year left before the end of the project but we can already provide some evidence confirming these contributions. This information will be elaborated further in the final deliverables of the project.

7.1 LOD Search

In terms of the LOD search represented by WP4 there a several elements that contribute to the state of art. The LOD Search Tool extracts information about two or more entities and identify unknown concepts and related stakeholders. This is achieved by retrieving the connectivity structures from data sets that contain information about the given entities, and ranking them based on information-theoretical approach. Connectivity structures are sub-graphs of one or more combined data sets that connect the given entities with each other. The LOD search tool constructs the sub-graphs by retrieving all paths between each combination of entities and ranking them. The difference between Sense4us LOD search tool and other systems is the ability to access any LOD data set to retrieve sub-graph structures between given entities and rank them according to their relatedness without extensive pre-processing of data or iterative calculations for rankings.

CROWDKI (Crowd-Powered Knowledge Integration) represent an interesting approach since it applies various crowd sourcing techniques for weeding out the errors in the results produced by the LOD search. In essence it is a refinement tool that uses human intelligence for identifying the information that is the most relevant for the given problem domain.

7.2 Sentiment analysis

The development of the SentiCircles (Saif, H. et al 2012) approach to sentiment analysis represent a novel idea and a contribution to the state of art. As opposed to traditional lexicon-based approaches SentiCircles does not consider the sentiment of terms static, but depending on the context in which the terms are used. To capture the words' contextual semantics, we follow the distributional hypothesis that words that occur in similar contexts tend to have similar meanings (Turney 2010). Therefore, the contextual semantics of a term *m* in our approach is computed from its co-occurrence with other terms. Note that a context is normally given by a collection of posts. In Sense4us project, a collection of posts representing social media discussions about a particular policy of interest.

7.3 System architecture

Although the design of a software architecture is not strictly related to the main themes of this deliverable nevertheless in respect to the contributions of the project we'd like to mention that we have developed and followed a process allowed to develop the software architecture in a simpler and more systematic way. The details of this process are described in deliverable D3.2 *"First System Architecture Description"*. Our approach to architecture design can be summarised as follows:

a) for the architecture design we do not rely solely on the end user requirements but use multiple information sources

b) the architecture is designed “up-front”, prior to any implementation. This is in contrast with the agile method
c) We develop the architecture iteratively by over-laying sub-architectures for each scenario
d) We formulate the use case scenarios along the objectives of the project rather than trying to aggregate individual functional (user) requirements. Our experience shows that the objectives represent a wider scope and provide themes that allows an easier integration of individual functional (user) requirements into use case scenarios

7.4 Policy modelling and simulation

The Sense4us projects has already produced a prototype for a policy-oriented modelling and simulation tool. This tool allows users, through a web-based interface, to build a systems model of a public policy problem situation using a graphical representation of the involved actors, the key variables, control flows and causal dependencies. A quantitative dynamic simulation model of the structured problem is used to simulate the system behaviour and responses to changing external factors and policy interventions over time. The tool supports the design of policy options and integrated impact assessment in terms of social, economic and environmental impacts. The main contribution of this research so far can be summarised as follows:

a) Defining the mechanism for policy modelling and simulation based on the information from sentiment analysis and LOD search.
b) Creating customisable and reusable models we use standardised categories of the model elements for example executive actors, policy instruments, external factors, policy impacts etc.
c) We use the technology of causal map graphical representation that allows to summarise the gathered information about a social, socioeconomic or sociotechnical system.

The current work involves some promising directions that may contribute to the state of art. The outcome of this research will be critically appreciated towards the end of the project, the main directions involve the following:

a) Integration of text analysis algorithms for causal inference extraction from textual data using Natural Language Processing.
b) Integration to Multi-criteria decision analysis (MCDA) models - building criteria models and data formats for policy appraisal based on the problem model.
c) Consideration of more complex forms of the cause-effect relationships (influences or causal links), including: time-/ value- dependent change transfer coefficients or differential equations.
d) The simulation as a serious game. Adding the formal structural elements of games — e.g., fun, play, rules, a goal, winning, challenges, competition, in addition to the feature of processing or debriefing using artificial intelligence techniques.

8 Summary

In this deliverable we have surveyed the main directions of research, the key concepts and tools that we consider to be relevant for the Sense4us project. We have also identified the contribution that the project makes to the state of art. The project itself covers a wide range of topics that involve open data sources, harvesting information from social media, visualisation of semantic data, and the analysis of social and demographics aspects in policy making. Producing a comprehensive survey of all these areas is beyond the scope of the presented report. Therefore we have focused only on the themes that cover the key components of the Sense4us system such as topic extraction, text summarisation, finding related information, policy impact simulation, finding related public opinion and social attitudes. In each of these subject areas we have provided numerous references to the literature that capture the main directions of research, the key concepts and tools that represent interest for the Sense4us project. The intention of this report was to highlight the opportunities, take account of new technologies and provide pointers to alternative solutions that can be used in the project.

The findings of the report for individual subject areas can be summarised as follows:

a) Topic extraction and document summarisation - in the light of the rapidly increasing volume of digital publishing, is becoming one of the key research areas of computer science. This research promises the automatic production of digests that are concise, comprehensive and also reflect the key points of input documents. Recently there has been a significant progress in this direction, however there are only a few studies that provide a critical assessment of the quality of output against the digests produced by humans.
   Regarding the topic extraction techniques we came to a conclusion that LDA has not been popular with users because it is firstly difficult to understand – it produces “bags of words” that describe topics, and many times they are difficult to decipher. Secondly, it is based on entropy, which means if the LDA is run on the same text more than once, the results will differ slightly, reducing confidence in its results.

   To address these criticisms, we investigated a number of tools and settled on a freely-available Java implementation of TextRank. This produces more easy-to-understand results because it selects phrases from the document under scrutiny.

b) Finding information in LOD space that can aid reasoning and decision making is essential for drafting policies. If we looked at the information that is published on the web, it is mainly text based. Although there is an increasing amount of data, however it is mainly hidden in databases and difficult to find since the data is not indexed by the search engines. The Linked Open Data initiative is trying to improve this situation by integrating data sources and making them public. However, it must be said that until data search becomes an integral part of standard web browsers the wealth of data will likely to remain hidden.
   The lack of mapping between datasets is a barrier to searching them. What is needed is a method of mapping the same concept across multiple datasets. This is being

addressed in WP4, but our survey provides additional solutions that could be considered.

c) Finding information in social media has got the potential to gauge the mood of the society in relation to various events and policies. However, it must also be said that this information can be distorted since according to recent studies most of the traffic on certain topics is generated by a small fraction of participants by (Fernandez, M. et al 2012). One of the problems with harvesting social media is the lack information about the profile of participants which makes any sound statistical studies about the real public opinion difficult to conduct.
Many users of social media are not citizens – for example a high proportion of posters on Twitter are commercial organisations. A mechanism is required to identify different types of social media user, so that commercial postings can be identified and (if the user wishes) filtered out. To address this, KMI are looking into characterisation of social media users into categories, (e.g. commercial organisation) so that they may be identified and the appropriate action taken.
Some social media searches produce few tweets. It would be desirable to automatically augment the search term given by the user with synonyms, to increase the chances of more results being returned.
Location-based social media searches are important to end users, because of the national and local interests in policy issues. Geo-location data is present in approximately 10-20% of tweets, and in general is less important than the home location of a user making a tweet, because the user may be outside their home domain when they made the tweet, and the user's home location represents their main constituency of interest.

d) Policy impact simulation aims to develop simulation models that allow the investigation of various "*what if*" studies by changing the numerical input values of the model. The main issue in this respect is interpreting the policy document(s) in terms of a simulation model. This is a complex task and it is unlikely that it can be achieved automatically. Although some rudimentary model can probably be extracted directly from the text, further conceptualisation, refinement and validation of this model most certainly will require human intervention.

e) Social attitudes to politics are regularly monitored and measured by numerous organisations. The reason is to measure the level of public engagement and also to find out whether there is acceptance of a policy. Although some citizens are interested in politics they have little if any involvement in the process of policy making itself. There are only a few opportunities (if any) to express opinions or to provide input to the policy making process. The consequence is that there is a lack of understanding why certain policies have been accepted and what was the justification for their introduction.

9 References

A. G. Nuzzolese, V. Presutti, A. Gangemi, A. Musetti, and P. Ciancarini. 2013. “Aemoo - exploring knowledge on the Web.” *Proc. 5th Annu. ACM Web Sci. Conf. - WebSci ’13.* ACM. 272–275.

A. Go, R. Bhayani, and L. Huang. 2009. *Twitter sentiment classification using distant supervision.* CS224N Project Report,, Stanford.

Aday, S. 2010. *Blogs and bullets: New media in contentious politics.* http://www.usip.org/files/resources/pw65.pdf.

Agarwal, A., et al. 2011. “Sentiment analysis of twitter data.” *In proceedings of the acl 2011 workshop on languages in social media .* 30–38.

al., H. Saif et. 2013. “Evaluation datasets for twitter sentiment analysis a survey and a new dataset, the sts-gold.” *In Proceedings, 1st Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM) in conjunction with AI*IA Conference.* Turin.

2014. *AlchemyAPI.* http://www.alchemyapi.com/) .

Aletras, N. and Stevenson, M. 2013. “ Evaluating topic coherence using distributional semantics on Computational Semantics (IWCS 2013).” *In Proceedings of the 10th International Conference .* Potsdam: Association for Computational Linguistics. 13–22.

ALOE. 2014. http://aksw.org/projects/aloe.

Answers. 2014. http://www.answers.com/Q/What_is_the_definition_of_social_attitude.

Araujo, S. et al. 2010. “Fusion – visually exploring and eliciting relationships in linked data.” *The Semantic Web – ISWC 2010.* Springer Berlin / Heidelberg. 1–15.

Araújo, S., et al. 2011. “SERIMI - resource description similarity, RDF instance matching and interlinking.” *Proceedings of the 7th International Workshop on Ontology Matching.*

Assistant, Ultimate Research. 2014. http://www.ultimate-research-assistant.com/GenerateResearchReport.asp.

Audit. 2014. http://www.auditofpoliticalengagement.org/.

Auer, S. et al. 2010. “Less template-based syndication and presentation of linked data.” *The Semantic Web: Research and Applications.* Springer Berlin / Heidelberg. 211–224.

Augenstein, I et al. 2012. “LODifier: generating linked data from unstructured text.” *ESWC'12 Proceedings of the 9th international conference on The Semantic Web: research and applications.* Berlin: Springer-Verlag. 210-224 .

Beevolve. 2012. *An Exhaustive Study of Twitter Users Across the World.* http://www.beevolve.com/twitter-statistics.

Berners-Lee, T. 2006. http://www.w3.org/DesignIssues/LinkedData.html.

Berners-Lee, T. et al. 2001. “The Semantic Web.” *Scientific American* 284: 34-43.

Blei, D. M, et al . 2003. “Latent dirichlet allocation.” *Journal of Machine Learning Research*, 3: 993–1022.

Bollen, J. et al. 2011. “Twitter mood predicts the stock market.” *Journal of Computational Science* 2 (1): 1-8.

Cafarella, M., et al. 2011. “Structured Data on the Web.” *Communications of the ACM* 54 (2): 72-79.

calculator, Climate. 2014. https://www.gov.uk/2050-pathways-analysis.

Cambria, E. 2013. “An introduction to concept-level sentiment analysis.” *Advances in Soft Computing and Its Applications.* Springer. 478–483.

Cano. 2014. “Automatic Labelling of Topic Models Learned from Twitter by.” *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers).* Baltimore, Maryland, USA: Association for Computational Linguistics. 618–624.

Chi. 2015. *Chi.* https://en.wikipedia.org/wiki/Chi-squared_test.

2012. “Citizen Participation.” http://www.businessofgovernment.org/blog/business-government/next-four-years-citizen-participation.

CKAN. 2014. http://thedatahub.org/.

CMAP3. 2015. *CMAP3 & COMPARATIVE AND COMPOSITE CAUSAL MAPPING.* http://www2.uef.fi/fi/cmap3.

Collier, A., et al. 2010. “Understanding and influencing behaviours: a review of social research, economics and policy making in Defra.” February. http://archive.defra.gov.uk/evidence/series/documents/understand-influence-behaviour-discuss.pdf.

Connell, N. 2001. “Evaluating soft OR: Some reflections on an apparently unsuccessful implementation using a soft systems methodology (SSM) based approach.” *Journal of the Operational Research Society* 150–160.

CROWDKI. 2015. https://github.com/criscod/CROWDKI.

Danielson, M., et al. 2006. “Cross-disciplinary research in analytic decision support systems.” *Proceedings of the 28th International Conference Information Tech. Interfaces.* IEEE.

DCAT. 2014. http://www.w3.org/egov/wiki/Data_Catalog_Vocabulary.

de Vries, B. et al. 2005. “Understanding Design Through Design Support Tools.” *Design Research in the Netherlands 2005.* Eindhoven: Eindhoven University of Technology. 205-214.

2015. *DExp.* http://www.banxia.com/.

Ding, L., et al. 2011. “Twc logd: A portal for linked open government data ecosystems.” *Web Semantics: Science, Services and Agents on the World Wide Web.* Elsevier. 325–333.

Druckenmiller et al. 2009. “Agent-Based Collaborative Approach to Graphing Causal Maps for Situation Formulation.” *Journal of the Association for Information Systems* 10 (3).

Eden, C. 2000. “On evaluating the performance of GSS:Furthering the debate.” *European Journal of Operational Research* 218–222.

Ell, B., et al. 2011. “Labels in the web of data.” *The Semantic Web – ISWC 2011.* Springer Berlin/Heidelberg. 162–176.

Fasth, T., Larsson, A. 2012. “Portfolio Decision Analysis in Vague Domains.” *Proceedings of IEEE IEEM.* IEEE. 61-65.

Fernandez, M. et al. 2012. “Using Social Media To Inform Policy Making: To whom are we listening? .” In *European Conference on Social Media: ECSM*, 174-182. ePub .

Firth, J. R. 1962. *A synopsis of linguistic theory. Studies in Linguistic Analysis.* Oxford: Basdil Blackwell.

Franco L. A., Montibeller G. 2011. “Problem structuring for multicriteria decision analysis interventions.” In *Wiley encyclopedia of operations research and management science*, by Cochran et al. (ed). USA: Wiley.

Franklin, M., et al. 2005. “From databases to dataspaces: a new abstraction for information management.” *SIGMOD Record* 34 (4): 27-33.

Friend, J. 2001. “The strategic choice approach.” In *Rational Analysis for a Problematic World Revisited: Problem Structuring Methods for Complexity,Uncertainty and Conflict*, by J., Mingers, J. (Eds.) Rosenhead, 115–150. Chichester: Wiley.

Fung, A., Shkabatur, J. 2012. *iral Engagement: Fast, Cheap, and Broad, but Good for Democracy?* http://www.archonfung.net/docs/articles/2012/ViralEngagement5.pdf.

G. Kasneci, S. Elbassuoni, and G. Weikum. 2009. “MING: Mining Informative Entity Relationship Subgraphs.” *Proceedings of the 18th ACM Conference on Information and Knowledge Management.* 1653–1656.

Gambit. 2015. *Gambit.* http://www.gambit-project.org/.

Gass, S. I. 2011. “"George B. Dantzig". Profiles in Operations Research.” *International Series in Operations Research & Management Science* 217–240.

Global. 2014. *Global.* http://www.globalcalculator.org/.

Gonzalez, H. et al. 2010. “Google fusion tables: data management, integration and collaboration in the cloud.” *Proceedings of the 1st ACM symposium on Cloud computing.* ACM. 175–180.

Griffiths, T.L, Steyvers, M. 2004. “Finding scientific topics.” *PNAS, 101(suppl. 1)*, : 5228–5235.

H. Saif, Y. He, and H. Alani. 2010. “Semantic sentiment analysis of twitter.” Boston, MA: In Proceedings of the 11th international conference on The Semantic Web.

Haase, P., et al. 2011. “The Information Workbench as a Self-Service Platform for Linked Data Applications.” *Workshop on Consuming Linked Data (COLD 2011).* Bonn: ISWC 2011.

Halevy, A. Y. 2012. “Towards an ecosystem of structured data on the web.” *Proceedings of the 15th International Conference on Extending Database Technology.* EDBT ’12. 1-2.

Hansard. 2014. “Audit of Political Engagement.” http://www.hansardsociety.org.uk/wp-content/uploads/2014/04/Audit-of-Political-Engagement-11-2014.pdf.

Harnden, R. 1990. “The languaging of models: The understanding and communication of models with particular reference to Stafford Beer's cybernetic model of organization structure.” *Systems Practice* 3 (3): 289–302.

Harris, R. 2012. *Introduction to Decision Making.* VirtualSalt.

Hartig, O., et al . 2009. “Executing SPARQL Queries over the Web of Linked Data.” *International Semantic Web Conference.* 293-309.

Heath, T. and Bizer, C. 2011. *Linked Data: Evolving the Web into a Global Data Space.* Morgan & Claypool.

Hedeler, C., et al. 2010. “Dataspaces.” In *Search Computing*, by S., Brambilla, M. Ceri, 114-134. Springer-Verlag.

Hulpus, I., et al. 2013. “Unsupervised graph-based topic labelling using dbpedia.” *In Proceedings of the sixth ACM international conference on Web search and data mining, WSDM ’13.* New York: ACM. 465–474.

Hussain, M., and Howard, P. 2011. “The role of digital media.” *Journal of democracy* 22 (3): 35–48.

I. Augenstein, S. Padó, S. Rudolph. 2012. “LODifier: Generating Linked Data from Unstructured Text.” In *Lectures in Computer Science, The Semantic Web: Research and Applications*, 210-224. Springer. http://www.aifb.kit.edu/web/LODifier/en.

iResearch. 2014. http://iresearch-reporter.com/.

Isele, R., et al. 2011. “LDSpider: An open-source crawling framework for the Web of Linked Data.” *Proceedings of 9th International Semantic Web Conference (ISWC 2010).*

Jiwani, G.N. 2010. *Uncovering the Unknown of Government Policy Decision-making Process at Senior Levels: Multiple Case Study.* University of Washington.

John, Tony. 2012. *What is Semantic Search?* http://www.techulator.com/resources/5933-What-Semantic-Search.aspx.

Keeney R.L, Raiffa H. 1993. *Decisions with multiple objectives: preferences and value tradeoffs.* Cambridge University Press: Cambridge.

KnoFuss. 2015. http://technologies.kmi.open.ac.uk/knofuss/.

Kouloumpis, E. et al. 2011. “Twitter sentiment analysis: The good the bad and the omg!” *Proceedings of the ICWSM.* Barcelona.

Kullback-Leiber. 2014. http://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence.

L. Fang, A. A. Das Sarma, C. Yu, and P. Bohannon. 2011. “REX: Explaining Relationships Between Entity Pairs.” *Proc. VLDB Endow.* 241–252.

Langegger, A. et al. 2008. “A semantic web middleware for virtual data integration on the web.” *Proceedings of the 5th European semantic web conference on The semantic web: research and applications.* Springer-Verlag. 493-507.

Lau, J.H., et al . 2011. “Automatic labelling of topic models.” *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1.* Stroudsburg, PA, USA: Association for Computational Linguistics. 1536–1545.

Laukkanen, Mingde Wang and Mauri. 2015. *Comparative Causal mapping: The CMAP3 method.* Ashgate Publishing, Ltd.

LDA. 2014. http://en.wikipedia.org/wiki/Latent_Dirichlet_allocation.

Leavy, J. 2013. *Social Media and Public Policy, What is the evidence? .* Alliance for Useful Evidence report.

Lenzerini, M. 2002. “Data integration: A theoretical perspective.” *21st ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems.* Madison: ACM. 233-246.

Li, J. et al. 2009. “RiMOM: A Dynamic Multistrategy Ontology Alignment Framework.” *IEEE Transactions on Knowledge and Data Engineering.* IEEE Computer Society. 1218-1232.

Liesiö, J., et al. 2000. “Preference Programming for Robust Portfolio Modeling.” *European Journal of Operational Research* 181 (3): 1488-1505.

LIMES. 2015. *LIMES.* http://aksw.org/Projects/LIMES.html.

Lin, C.-Y. 2004. “ROUGE: A package for automatic evaluation of summaries.” *Proceedings of ACL Text Summarization Branches Out Workshop.* 74–81.

LinkedData. 2014. http://linkeddata.org/.

Liu, B. 2010. “Sentiment analysis and subjectivity.” In *Handbook of Natural Language Processing*, by N. Indurkhya and F. J. Damerau (eds).

—. 2010. *Sentiment analysis and subjectivity. Handbook of natural language processing.*

—. 2010. *Sentiment analysis and subjectivity. Handbook of natural language processing.*

Lloret, E., Palomar, M. 2010. “Challenging Issues of Automatic Summarization: Relevance Detection and Quality-based Evaluation.” *Informatica* 34: 29–35.

LOD. 2015. *LOD.* http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData.

Lourenço, J., et al. 2012. “PROBE – A Multicriteria Decision Support System for Portfolio Robustness Evaluation.” *Decision Support Systems* 54 (1): 534-550.

Luna-Reyes et al. 2003. “Collecting and analyzing qualitative data for system dynamics: methods and models.” *System Dynamics Review* 19 (4): 271-296.

M. Thelwall, K. Buckley, and G. Paltoglou. 2012. “Sentiment strength detection for the social web.” *Journal of the American Society for Information Science and Technology* 63 (1): 163-173.

Madden, M. 2014. “Older Adults and Social Media, Pew Internet.” http://pewinternet.org/Reports/2010/Older-Adults-and-Social-Media.aspx.

Magatti,D., et al. 2009. “Automatic labeling of topics.” *In Proceedings of the 2009 Ninth International Conference on Intelligent Systems Design and Applications, ISDA ’09.* Washington: IEEE Computer Society. 1227–1232.

Makela, E. 2005. *Survey of Semantic Search Research.* http://www.seco.tkk.fi/publications/2005/makela-semantic-search-2005.pdf.

Mani, I. 2014. *Summarization Evaluation: An Overview.* http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings2/sum-mani.pdf.

McCrae, J., Spohr,D., Cimiano, P. 2011. “Linking lexical resources and ontologies on the semantic web with lemon.” In *The Semantic Web: Research and Applications*, 245–259. Heidelberg: Springer.

Mei, Q., et al. 2007. “Automatic labeling of multinomial topic models.” *In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining.* New York: ACM. 490–499.

Microformats. 2014. http://microformats.org.

Mingers, J., Rosenhead,J. 2004. “Problem structuring methods in action.” *European Journal of Operational Research* 530–554.

Mintzberg, H. 1994. *The Rise and Fall of Strategic Planning.* New York: Prentice-Hall.

Mitchell, B. 2009. *Policy-making process.* http://www.waterencyclopedia.com/Oc-Po/Policy-Making-Process.html.

Musetti, A. G. Nuzzolese, F. Draicchio, V. Presutti, E. Blomqvist, A. Gangemi, and P. Ciancarini. 2012. “Aemoo: Exploratory search based on knowledge patterns over the semantic web.” *Semant. Web Chall.*

Myerson, R.B. 1991. *Game Theory: Analysis of Conflict.* Harvard University Press.

N. Marie, A. B. Labs, F. Gandon, I. Sophia-antipolis, S. Antipolis, M. Ribière, and F. Rodio. 2004. “Discovery Hub?: on-the-fly linked data exploratory search.” *9th Int. Conf. Semant. Syst.* 17–24.

Nathan, Paco. 2009. *TextRank Java implementation.* https://github.com/ceteri/textrank.

Nenkova, A., Kathleen, K. 2011. “Automatic Summarization.” *Information Retrieval* 5 (2-3): 103–233.

Newsblaster. 2014. http://www1.cs.columbia.edu/~sable/research/hlt-blaster.pdf.

Ngomo, A. et al. 2011. “LIMES A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data.” *IJCAI.* 2312-2317.

Noia, R. Mirizzi and T. Di. 2010. “From Exploratory Search to Web Search and Back.” *Proceedings of the 3rd Workshop on Ph.D. Students in Information and Knowledge Management.* 39–46.

OR. 2014. *What is Operations Research?* https://www.informs.org/About-INFORMS/What-is-Operations-Research.

Ormston, R. and Paterson, L. 2015. *Higher Education, Investing in the future? Attitudes to University.* NatCen Social Research. http://www.bsa.natcen.ac.uk/media/38917/bsa32_highereducation.pdf.

P. Heim, S. Lohmann, and T. Stegemann. 2010. “Interactive relationship discovery via the semantic web.” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics).* Springer-Verlag. 303–317.

Pak, A., Paroubek, P. 2010. “Twitter as a corpus for sentiment analysis and opinion mining.” *Proceedings of LREC 2010.* Valletta.

Paroubek, A. Pak and P. 2010. “Twitter as a corpus for sentiment analysis and opinion mining.” Malta: Proceedings of LREC 2010.

Pearson. 2015. *Pearson.* https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient.

Phillips, L. 1989. “People-centered group decision support.” In *Knowledge-Based Management Support Systems*, by Doukidis et al, 208–224. Chichester: Ellis-Horwood.

Pirró, G. 2015. “Explaining and Suggesting Relatedness in Knowledge Graphs.” *Proceedings of the 14th International Semantic Web Conference (ISWC).* Bethlehem, Pennsylvania, USA: Springer-Verlag. 1–16.

2014. “Politics.” http://bsa-30.natcen.ac.uk/read-the-report/politics/introduction.aspx.

Puron-Cid, G., et al. 2012. “IT-Enabled Policy Analysis: New Technologies, Sophisticated Analysis and Open Data for Better Government Decisions.” *Proceedings of the 13th Annual International Conference on Digital Government Research.* 97-106.

Rachel Ormston, Lindsay Paterson. 2015. *Higher Education: Investing in the future? Attitudes to University.* British Social Attitudes, NatCen Social Research. Accessed 2015. http://www.bsa.natcen.ac.uk/media/38917/bsa32_highereducation.pdf.

Radev, D., et al. 2005. “Newsinessence: Summarizing Online News Topics.” *Communications of the ACM - The digital society* 95-98.

Radev, Gunes Erkan and Dragomir R. 2004. “exRank: Graph-based Lexical Centrality as Salience in Text Summarization.” *ournal of Artificial Intelligence Research*, 12: 457-479.

RDF. 2015. *RDF.* http://www.w3.org/TR/rdf-primer.

Researcher, NewsFeed. 2014. http://newsfeedresearcher.com/.

RiMOM. 2015. *RiMOM.* http://keg.cs.tsinghua.edu.cn/project/RiMOM/.

Robert Isele, Anja Jentzsch, Christian Bizer. 2010. “Silk Server - Adding missing Links while consuming Linked Data.” *COLD 2010.*

Robert Isele, Christian Bizer. 2013. “Active learning of expressive linkage rules using genetic programming.” *J. Web Sem.* 2-15.

ROBUST. 2014. http://www.robust-project.eu/.

Rosenhead. 1996. “What is the problem. An introduction to problem structuring methods.” *Interfaces* 26 (6): 117–131.

Rosenhead, J. 1989. *Rational Analysis for a Problematic.* Chichester: Wiley.

Saif, H. et al. 2012. “Semantic sentiment analysis of twitter.” *11th Int. Semantic Web Conf.* Boston: ISWC.

Salo, A., et al. 2011. “Improving Resource Allocation with Portfolio Decision Analysis.” *Analytics* 23-26.

Schulte, A., et al. 2011. “LDIF - Linked Data Integration Framework.” *Workshop on Consuming Linked Data (COLD 2011).* Bonn: ISWC 2011.

2014. *Semantria.* http://www.semantria.com.

2015. *SentiStrenght.* http://sentistrength.wlv.ac.uk/.

sentistrength. 2014. http://sentistrength.wlv.ac.uk/.

sentiwordnet. 2014. http://sentiwordnet.isti.cnr.it/.

2015. “silk.” http://silk-framework.com/.

silk. 2015. *silk.* http://silk-framework.com/.

Spearman. 2015. *Spearman.* https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient.

STELLA. 2015. http://www.iseesystems.com/softwares/stella-pro/v1.aspx.

Sterman, J. 2000. *Systems Thinking and Modeling for a Complex World.* Boston: Irwin/McGraw-Hill.

Suchanek F. M., Sozio, M., Weikum, G. 2009. “Sofie: a self-organizing framework for information extraction.” *Proceedings of the 18th international conference on World wide web* 631–640.

2014. *Support.* http://en.wikipedia.org/wiki/Support_vector_machine.

Tarau, Rada Mihalcea and Paul. 2004. “TextRank: Bringing Order into Texts.” *in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004).*

2014. *Texalytics.* http://textalytics.com/core/topics-info.

tools. 2014. http://www.ifm.eng.cam.ac.uk/research/dstools/#2b.

2014. “Traffic.” http://www.comscore.com/.

Tumasjan, A., et al. 2010. “Predicting elections with twitter: What 140 characters reveal about political sentiment.” *In Proceedings of the fourth international aaai conference on weblogs and social media.* 178-185.

Tummarello, G., et al. 2010. “Sig. ma: Live views on the Web of Data.” *Web Semantics: Science, Services and Agents on the World Wide Web* 8 (4): 355-364.

Turney, P.D. et al. 2010. “From frequency to meaning: Vector space models of semantics.” *Journal of artificial intelligence research* 31 (1): 141–188.

UNIKO. 2015. http://uniko.ac.at/wissenswertes/uniko_pedia/crowdfunding/.

van Ossenbruggen, J. et al. 2011. “Interactive Vocabulary Alignment.” *International Conference on Theory and Practice of Digital Libraries.* 296-307.

VoId. 2014. http://vocab.deri.ie/void/guide.

Volz, J. et al. 2009. “Discovering and Maintaining Links on the Web of Data.” *International Semantic Web Conference.* 650-665.

Waitelonis, H. Sack and J. 2010. “Exploratory semantic video search with yovisto.” *Proc. - 2010 IEEE 4th Int. Conf. Semant. Comput. ICSC 2010.* IEEEE. 446–447.

Waitelonis, J., Sack, H. 2010. “Exploratory semantic video search with yovisto.” *Proc. - 2010 IEEE 4th Int. Conf. Semant. Comput. ICSC 2010.* 446–447.

WeGov. 2014. http://www.wegov-project.eu/.

Weimer, D., and A. Vining. 2005. *Policy Analysis: Concepts and Practice .* Upper Saddle River, NJ: Pearson Prentice Hall.

Wheeler, B. 2012. *Why not let social media run the country?* http://www.bbc.co.uk/news/uk-politics-19555756.

Wikpedia. 2015. *Nash equilibrium.* https://en.wikipedia.org/wiki/Nash_equilibrium.

Wills, G. J. 2009. “Visualizing hierarchical data.” *Encyclopedia of Database Systems.* Springer US. 3425–3432.

Wilson, T. et al. 2005. “Recognizing contextual polarity in phrase-level sentiment analysis.” *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing.* Vancouver.

WordNet. 2014. http://wordnet.princeton.edu/.

2014. *YahooAPI.* http://developer.yahoo.com/search/content/V1/termExtraction.html.

Yardi, S., Boyd, D. 2010. “Dynamic debates: an analysis of group polarization over time on Twitter.” *Bulletin of Science, Technology & Society* 30 (5): 316-327.