

NOMAD: Linguistic Resources and Tools Aimed at Policy Formulation and Validation

George Kiomourtzis, George Giannakopoulos, Georgios Petasis, Pythagoras Karampiperis, Vangelis Karkaletsis
{gkiom, ggianna, petasis, pythk, vangelis}@iit.demokritos.gr
SKEL, NCSR Demokritos, Greece

Abstract

The NOMAD project (Policy Formulation and Validation through non Moderated Crowd-sourcing) is a project that supports policy making, by providing rich, actionable information related to how citizens perceive different policies. NOMAD automatically analyzes citizen contributions to the informal web (e.g. forums, social networks, blogs, newsgroups and wikis) using a variety of tools. These tools comprise text retrieval, topic classification, argument detection and sentiment analysis, as well as argument summarization. NOMAD provides decision-makers with a full arsenal of solutions starting from describing a domain and a policy to applying content search and acquisition, categorization and visualization. These solutions work in a collaborative manner in the policy-making arena. NOMAD, thus, embeds editing, analysis and visualization technologies into a concrete framework, applicable in a variety of policy-making and decision support settings. In this paper we provide an overview of the linguistic tools and resources of NOMAD.

Keywords: crowdsourcing, e-government, information extraction

1. Introduction

Argumentation is a branch of philosophy that studies the act or process of forming reasons and of drawing conclusions in the context of a discussion, dialogue, or conversation. Being an important element of human communication, its use is very frequent in texts, as a means to convey meaning to the reader. As a result, argumentation has attracted significant research focus from many disciplines, ranging from philosophy to artificial intelligence. Central to argumentation is the notion of argument, which according to (Besnard and Hunter, 2008) is “a set of assumptions (i.e. information from which conclusions can be drawn), together with a conclusion that can be obtained by one or more reasoning steps (i.e. steps of deduction)”. The conclusion of the argument is often called the claim, or equivalently the consequent or the conclusion of the argument, while the assumptions are called the support, or equivalently the premises of the argument, which provide the reason (or equivalently the justification) for the claim of the argument. The process of extracting conclusions/claims along with their supporting premises, both of which compose an argument, is known as argument extraction (Goudas et al., 2014) and constitutes an emerging research field.

Argument extraction may be proved an invaluable resource in the domain of policy making and politics. Within the political domain it could help politicians identify the peoples’ view about their political plans, laws, etc. in order to design more efficiently their policies. Additionally, it could help the voters in deciding which policies and political parties suit them better. Social media is a domain that contains a massive volume of information on every possible subject, from religion to health and products, and it is a prosperous place for exchanging opinions. Its nature is based on debating, so there already is plenty of useful information that waits to be identified and extracted.

Collaboration and crowd-sourcing are the realities of today’s public Internet. The so-called “Web 2.0” contains heterogeneous content that is inserted daily and spontaneously updated by its users. By exploiting publicly avail-

able data in the domain of policy making it would be possible to use this insight and information at multiple stages of the policy-life cycle (Charalabidis et al., 2012; Maragoudakis et al., 2011). The NOMAD project (Policy Formulation and Validation through non Moderated Crowd-sourcing) is a project that supports policy making, by providing rich, actionable information related to how citizens perceive different policies.

NOMAD builds on contributions on the informal web (e.g. forums, social networks, blogs, newsgroups and wikis), so as to gather useful feedback. The ability to leverage the vast amount of user-generated content for supporting policy makers in their decisions requires new tools that will be able to gather, analyze and visualize - from sources as diverse as blogs, online opinion polls and government reports - the opinions expressed on the informal Web. NOMAD changes the experience of policy making by providing decision-makers with fully automated solutions for content search, selection, acquisition, categorization and visualization that work in a collaborative form in the policy-making arena. To this end NOMAD embeds visualization, editing and analysis technologies, forming a complete tool applicable in a variety of policy-making and decision support settings.

The paper is structured as follows. We briefly describe the related work (Section 2) and we overview NOMAD architecture (Section 3). We then elaborate on the linguistic pipeline (Section 4) and especially on the argument extraction and sentiment analysis subprocesses (Section 5), before concluding the paper (Section 6).

2. Related Work

There exists a variety of ongoing projects and works that focus on how to use publicly available opinions and texts to determine social trends and analyze political positions. For example, the WeGov toolbox (Wandhofer et al., 2012), result of the homonymous EU project (see more at <http://wegov-project.eu>) is an online tool, with the following main functions: it enables the policy maker to search for discussions, topics and opinions from different Social

Media; it supports the analysis and summarization of discussions, to determine the discussion topic and important posts; and it finally helps policy makers communicate the extracted information through the Social Media. The system also predicts posts that are expected to generate higher attention, to allow the policy maker to focus on important topics. It also takes into account user behavior and interaction to classify users. Essentially the system does not detect opinion, but rather opinion importance (based also on user modeling).

In a clearly political application of sentiment analysis, the work of Diakopoulos and Shamma (Diakopoulos and Shamma, 2010) connects the temporal dynamics of sentiment on Twitter to a debate. To this end, the authors used an annotation process based on Amazon Mechanical Turk (supported by several quality filters) to annotate tweets related to the debate. They also aligned the times of the debate to tweets. The work studied the overall sentiment of the tweets, whether users favored a specific speaker/candidate and the temporal evolution of the sentiment. Part of the study was also focused on controversy. We stress that the method presented, connecting analysis to real-time social sentiment dynamics is not automatic: it is manual.

A study by Tumasjan et al. (Tumasjan et al., 2010) verifies that microblogging (Twitter) is used extensively for political deliberation, even though it is dominated (in terms of messages per user) by a small number of heavy users. To achieve sentiment analysis the authors used the LIWC2007 linguistic analysis tool, which analyses several “cognitions and emotions”, such as: future orientation, past orientation, positive emotions, negative emotions, sadness, anxiety, anger, tentativeness, certainty, work, achievement, and money. In this work, the authors study whether the convergence between the profiles of politicians (political proximity) can be deduced from analyzed microblogging data and argue that it is possible.

In a related work (Boero et al., 2012), the PADGETS Analytics and PADGETS Simulation Model are discussed, constituting components of a decision support system for policy intelligence. The components described in the framework are meant provide the analytic power to extract actionable knowledge related to specific campaigns and targets, while taking into account social dynamics and privacy issues. In the work, also a number of issues related to the limitations of using social media for the extraction of knowledge are discussed.

NOMAD comes to combine Web 2.0 crawling, with such powerful technologies as argument extraction, opinion mining and argument summarization, to provide a fully automated software suite, optimized for use in policy making settings. In the following paragraphs we go deeper into the architecture that allows us to implement this combination.

3. Architecture

The overall NOMAD system architecture, illustrated in Figure 1, relies on the separation of the whole working system into three main layers: the presentation layer, the storage layer and the processing layer. The presentation layer provides the user interface to support the policy mod-

eling process, by allowing the users to author their needs in the form of a domain (the policy domain), described as a set of terms, and a structured set of statements and labels describing policies themselves. These statements include policies (i.e., what one means to achieve), norms (i.e., how one means to achieve it) and arguments (i.e., arguments in favor of or against the norms). The storage layer provides an abstraction for persistence of data and meta-data used within the system and allows their use and re-use by different parties, to facilitate collaboration. In this paper we focus mostly on the processing layer which builds upon a variety of linguistic methods to process data from the public Web and support large scale analytics for policy modeling. In the next paragraphs we elaborate on the components present within this layer.

The NOMAD Service Orchestrator is the heart of the NOMAD pipeline. It is executed continuously and invokes the underlying components in a strictly defined sequence, taking into account the opportunities of parallelization for certain processes. The ordering allows processing data in different levels and depths of analysis, based on previous components’ results. The individual NOMAD components, exchange data via the semantically distinct though logically interlinked NOMAD repositories, which lie within the Storage Layer. The execution cycles are repeated, continuously updating the content and information available to the live system. The NOMAD Crawler and the Content Cleaner components are the first two components in the pipeline. The NOMAD Crawler is responsible for discovering and retrieving raw content — relevant to the expressed user need — from a set of covered Web 2.0 sources, as well as source-level demographic information for this content. The Content Cleaner, in turn, analyses the content retrieved from the NOMAD Crawler and extracts the clean textual information contained.

As depicted in Figure 1, the different components do not communicate with each other in a direct, peer-to-peer fashion. Instead, their coordination is realized via the Service Orchestrator. The components’ responsibility is to inform the Service Orchestrator on the finalization of their respective processes. We stress that NOMAD works with descriptions in free, natural language, provided by the Web users. To exploit the explicit and implied information that the models provide, it is essential to build a mechanism that bridges the worlds of knowledge representation and web search. In the next paragraphs, we briefly describe the linguistic tools integrated within NOMAD that bridge these worlds.

4. NOMAD Linguistic Pipeline

The NOMAD Linguistic Pipeline encapsulates all the components necessary for processing linguistically the input text derived from the acquired Web 2.0 content. In the following paragraphs we summarize the technical specifications of the components that the Linguistic Pipeline implements.

Thematic Classifier The Thematic Classifier analyses the texts given at its input and classifies them in one or more of the predetermined categories supported by the

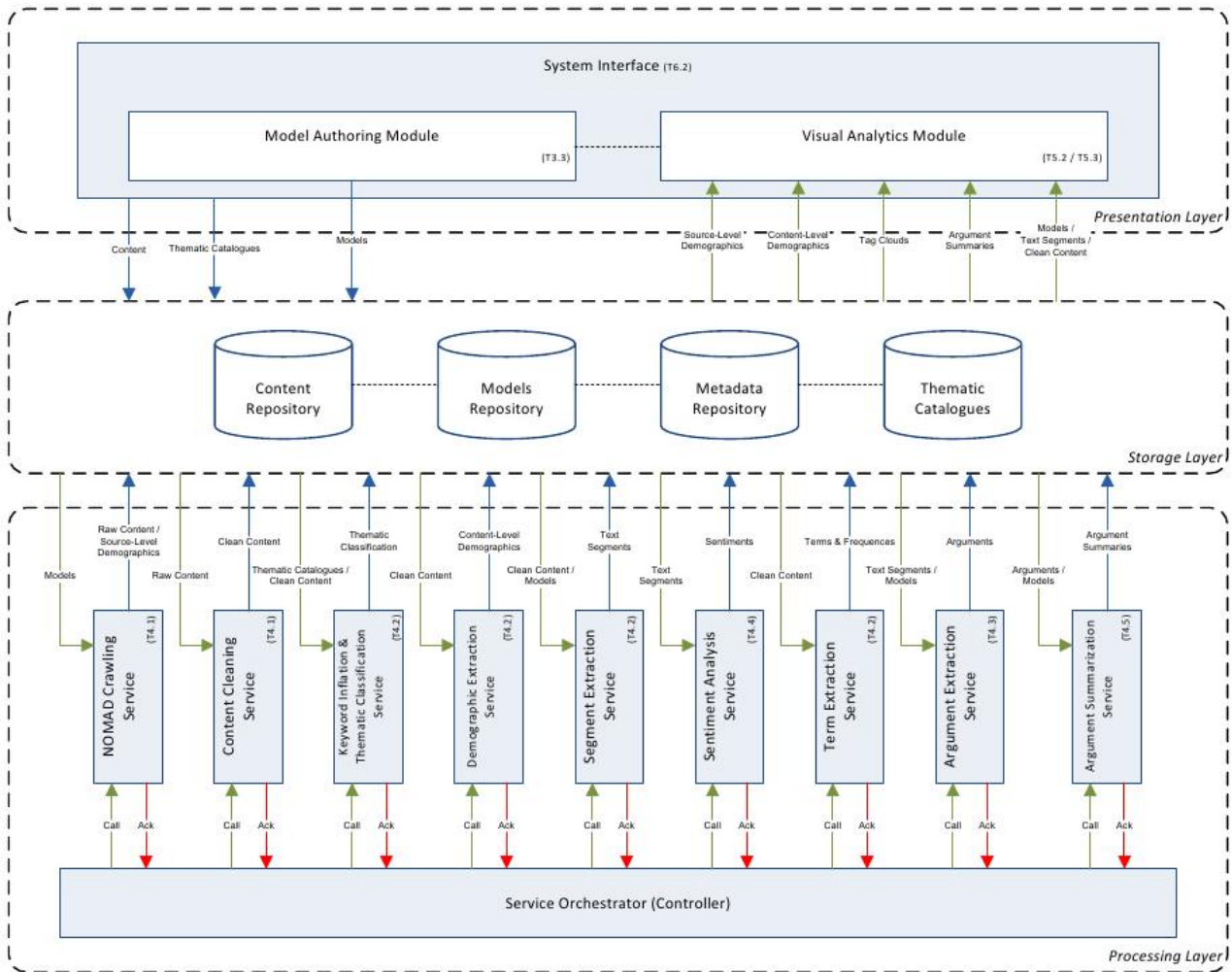


Figure 1: NOMAD Architecture

system, based on the domain model. The module relies on a predetermined thematic catalog, where the categories of interest are defined. Each category is conceptualized by a set of initial terms. Content that is found to contain these terms or terms semantically similar to these is considered to belong in this category.

Linguistic Demographics Extractor The Linguistic Demographics Extractor component extracts demographic information from the content discovered by the NOMAD crawling modules. Depending on the source of the content, the module either harvests information provided by the user profiles in social networking platforms, where certain characteristics may be declared explicitly, or uses content analysis (similar to (Rao et al., 2010)) in order to guess demographic information (e.g. age, gender etc.). The module supports stylometry-based approaches to estimate the demographic characteristics of authors. The problem with such approaches is that in such approaches a good corpus is required to train models of writing per age category/gender/education level. In other cases additional knowledge, e.g. originating from a user profile,

is required.

Segment Extractor The segment extraction service operates on the clean content retrieved by the NOMAD crawlers. It iterates every yet unprocessed document, and tries to find fragments of text that match a certain domain or policy model. The algorithm matches tokens on the underlying entity level, in order to fetch as many segments possible for the domain. A segment is the cumulative result of the above token matching process, per sentence of each document. So, if a specific domain entity is matched in a document's consequent sentences, the segment will be the whole sentence fragment.

Argument Extractor and Sentiment Analysis The Argument Extractor detects and extracts arguments from the analyzed sources, while the Sentiment Analysis component classifies the arguments based on the polarity of the sentiment they express. These two components of the system are elaborated in the following paragraphs.

Tag Cloud Generator The Tag Cloud Generator service operates on the cleaned textual information for each

content entry and applies the following process for identifying terms present in that text; first, stop-words are discarded. Next, the remaining words are stemmed and lemmatized using an appropriate lemmatizer for English and Greek languages and the occurrences of each distinct stem are counted. As far as the German language is concerned, we are also evaluating different solutions for lemmatization and plan to embed such a component in the near future.

Argument Summarizer The Argument Summarizer exploits the arguments detected by the Argument Extractor and selects one or more representative arguments to form a non-redundant extractive summary. The Argument Summarizer examines the segments that were found to be related to the argumentation of a given policy model and combines them in order to produce a concise, abbreviated report that preserves the main points of all the arguments. This component adapts existing clustering and automatic summarization approaches (e.g., (Giannakopoulos et al., 2010; Giannakopoulos et al., 2014)) to render the report. The approaches are elaborated below:

Centroid-based approach In this approach (see e.g., (Radev et al., 2004)) the texts of a given cluster are represented in the vector space. Each dimension of the vector maps to a (stemmed/lemmatized) word that appears in the argument of the cluster. The presence of a word in a text (argument) is indicated by the value 1 in the corresponding vector dimension, while the absence by the value 0. Given the set of vectors representing the cluster texts, we choose as representative the argument text that is mapped to a vector closest to the centroid to the vector set.

Representative graph-based approach In this approach the texts of a given cluster are represented as n-gram graphs (NGGs). A representative graph G is extracted through the merging of all the argument NGGs (update operator). We select as representative argument the one that maps to an NGG which is maximally similar to the representative graph G (as also applied in NewSum (Giannakopoulos et al., 2014)).

The two approaches have low computational cost and can be applied at runtime, if required. They are also rather language independent (with the exception of the stemming component). On the other hand, they are slightly different in that, while the centroid aims for the argument that is closer to the “average” argument, the NGG approach aims for the argument that maximally covers all the others. We will experiment with both approaches and see which is best for the NOMAD setting. The argument summarization is work-in-progress and, thus, extended evaluation has not yet been accomplished.

In the following section we overview the novel argument extractor and argument sentiment analysis components of

NOMAD, which empower the policy making decision support tools.

5. Argument Extractor and Sentiment Analysis

The Argument Extractor module is responsible for identifying arguments in documents retrieved by the NOMAD crawlers, and to associate the extracted arguments with policy arguments authored by the policy maker through the NOMAD authoring tool. Integrating several approaches able to extract arguments at various levels of detail (i.e. segments that represent argument claims/supports, or sentences containing arguments), the argument extraction service iterates over all unprocessed documents and extracts arguments that relate to the policy arguments, or arguments that do not relate to policy arguments, which are considered as possible new arguments that should be revised by the policy maker as possible additions to her/his policy. Extracted arguments are represented as segments, stored for future use by the Sentiment Analysis module.

In the following paragraphs we briefly describe the NOMAD corpus, which was created to allow training the Argument Extractor, while an evaluation of the argument extractor on the Greek sub-part of the NOMAD corpus, is presented in (Goudas et al., 2014). Then, we focus on the sentiment analysis module that applies a polarity indicator to each extracted argument.

5.1 The NOMAD Corpus

In the context of the Argument Extractor of NOMAD, a small corpus has been collected and manually annotated with domain named entities, argument components (claim, support segments), and polarity of the document authors towards the argument components. In addition, a draft policy model was created by the annotators, populated mainly by arguments in favour or against “draft policies” and ways to apply these policies. The main focus was on arguments, which were identified solely from the corpora, by listing arguments as found in the documents. The arguments on this “draft policy” were associated with the segments of argument components that were identified in texts.

The purpose of the creation of this corpus is twofold: A first objective is to offer a “gold standard”, which can be used to evaluate our technologies and approaches. A second objective is the exploitation of this corpus as training material, if such a need arises, for example in the cases of extracting a list of cue words or indicator markers, usually employed in discourse analysis, a task that includes the argument extraction.

The NOMAD corpus contains three sets of documents, one for each language that the NOMAD targets to process. Thus, there is a set of documents for the German, English, and Greek language. The thematic domain in each set reflects the domain of the selected pilot case for each language: The German set of documents contains documents related to open data, the English set of documents contains documents related to allergens, and the Greek set of documents contains documents related to renewable energy sources. The characteristics of the NOMAD corpus are shown in the following table (Table 1). From a preliminary analysis, we found that in the German and Greek

Language	Document Number	Argument Number
German	267	761
English	646	1256
Greek	125	1319

Table 1: Characteristics of the NOMAD (manually annotated) corpus.

subparts of the corpus, most of the arguments appear only once in documents, which is not the case for the English corpus subpart, where only 10 arguments appear only once in the corpus.

The corpus contains documents that were collected by the NOMAD Crawler module, originating from various sources, such as news, blogs, sites, etc. The corpus was constructed by manually filtering a larger corpus (through a special filtering tool that was constructed in the context of the project), automatically collected by the NOMAD Crawler, in order to keep only documents that contain arguments. Unfortunately, the corpus cannot be made public due to copyright restrictions of the original text owners and publishers. For more information please consult Deliverable D4.3.1 of the NOMAD project (Petasis et al., 2014).

5.2 Argument Sentiment Analysis

The Sentiment Analyzer discovers the polarity (positive, negative, neutral) of a specific statement. The component uses different linguistic analysis methods (lists of affective terms) and classification types (based on n-gram graphs (Aisopos et al., 2011; Giannakopoulos et al., 2008)) tools in order to complete its task.

The *wordlist-based method* is based on the standard Term-frequency (Hu and Liu, 2004) approach and is implemented by parsing the document through a tokenizer, and compare each extracted token to a predefined polarity labelled wordlist. The service examines the presence of such tokens in for each document passed and calculates its polarity as shown in Eq. 1:

$$\frac{PositiveCount - NegativeCount}{PositiveCount + NegativeCount} \quad (1)$$

The service uses SentiWordNet-derived lists (Esuli and Sebastiani, 2006; Baccianella et al., 2010) for analysing English text, SentiWS (Remus et al., 2010) for German text and custom term lists for Greek text. The latter comprise 284 terms of positive valence and 502 terms of negative valence. The lists were obtained using machine learning techniques, by analysing manually annotated content obtained from product reviews and assigning terms as positive and negative depending on their appearance in segments of positive and negative polarity, as determined by the human annotators, respectively.

As a second approach, we have applied and evaluated the use of n-gram graphs to represent different polarity classes, based on labelled instances, as has been done in other classification settings (including social media texts) (Aisopos et al., 2012; Giannakopoulos and Palpanas, 2010). We term this approach the *n-gram graph approach*.

We assume that each sentiment value (negative, neutral, positive) represents a class, and each class is represented

Method	Accuracy	Standard Deviation
Baseline	65.20%	
W	15.71%	3.23%
N	71.74%	3.89%
W - N	71.35%	4.10%

Table 2: Results for Greek documents

by a single, merged n-gram graph, created by a portion of the n-gram graphs of the class training instances. Then, each instance is represented in the similarity space of these n-gram graphs. In other words, each instance is represented by a vector (s_1, s_2, s_3, \dots) , where s_i is the similarity of the instance n-gram graph to the corresponding class-i graph. Afterwards, a machine learning algorithm (e.g., SVM) is used to learn to classify instances in this rich similarity space. The training instances used originate from the arguments annotated by the NOMAD human annotators.

In order to perform evaluation tests for the sentiment analysis module, we used the English and Greek portions of the annotation. We evaluated the approach using three different variations:

- Wordlists only (W) - we note that we applied stemming when using wordlists, because otherwise the accuracy of the method was extremely low ($< 10\%$)
- N-gram graphs only (N)
- Combined Wordlists and N-gram graphs (W-N)

To estimate the performance of the systems we used a 10-fold cross-validation approach using stratification (i.e. keeping the ratio of the class instances in the training and test sets constant). In the following paragraphs we illustrate the performance per language, also comparing with a baseline classifier (majority classifier, which always replies by selecting the most represented class in the dataset).

On the Greek corpus, the distribution of classes in the data was as follows:

- 364 negative
- 79 neutral
- 830 positive

On the English corpus, the distribution was as follows:

- 616 negative
- 27 neutral
- 473 positives

It is interesting that the distributions are different across languages. In the following table we provide the results in terms of accuracy.

From the above tables it is clear that there is no improvement over the n-gram graph method only, when combining with wordlists. A feature selection study using several criteria (e.g., Information Gain) showed that the wordlist feature in the 4-dimensional space offers no additional, useful information.

Method	Accuracy	Standard Deviation
Baseline	55.20%	
W	31.09%	2.67%
N	83.82%	1.80%
W - N	82.64%	3.31%

Table 3: Results for English documents

We note that an evaluation of the OpinionBuster (Petasis et al., 2013) commercial system on a part of our Greek corpus (814 instances) had an accuracy of 74.20% (604 out of 814 instances correctly classified). OpinionBuster employs compositional polarity classification, by combining sentiment lexica with linguistic patterns to determine polarity. Clearly the performance of the NOMAD analysis and OpinionBuster is equivalent within statistical error and significantly better than the baseline.

We also examined why there exists a significant difference (around 10% in accuracy) in the performance across languages. One possible explanation is the difference in the distribution of instances across classes. One second observation we made was that the instances per language were different, in that the average length varies a lot. In English instances there was an average of 22 to 26 words per instance over the classes. In Greek instances the words per item were from 5 to 6, i.e. significantly shorter. This may lead in a difficulty to detect longer linguistic patterns that can lead to a decision more safely. This shortcoming, if indeed one, may be mitigated by changing the annotation process. However, this might be a matter of future research. With the reference to the sentiment analysis component we conclude the description of the NOMAD components, which form the backbone of the NOMAD set of tools and integrated system. We stress that the argument-related contributions (detection, sentiment analysis and summarization) of NOMAD form a significant step ahead in the effort to support policy makers throughout the policy making lifecycle. They also outline and face some very realistic research problems asking for solutions.

6. Conclusion

In this paper we have described the NOMAD architecture, focusing on its linguistic pipeline. The pipeline implements and combines a variety of Natural Language Processing methods, enabling the use of visual analytics to provide actionable feedback to policy makers. The methods cover several research domains, from POS tagging and tokenization, to argument extraction, sentiment analysis and summarization, and provide a significant use case for the joined power of language resources and tools. NOMAD highlights how language technologies can affect the way e-governance and policy making can use the rich soil of Web 2.0 to cultivate a unique, interactive relationship with the people, for the benefit of all.

7. Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no

288513. For more details, please see the NOMAD project's website at <http://www.nomad-project.eu>.

8. References

- Aisopos, F., Papadakis, G., and Varvarigou, T. (2011). Sentiment analysis of social media content using n-gram graphs. In *Proceedings of the 3rd ACM SIGMM international workshop on Social media*, page 914. ACM.
- Aisopos, F., Papadakis, G., Tserpes, K., and Varvarigou, T. (2012). Content vs. context for sentiment analysis: a comparative analysis over microblogs. In *Proceedings of the 23rd ACM conference on Hypertext and social media*, page 187196. ACM.
- Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, page 22002204.
- Besnard, P. and Hunter, A. (2008). *Elements of Argumentation*. MIT Press.
- Boero, R., Ferro, E., Osella, M., Charalabidis, Y., and Loukis, E. (2012). Policy intelligence in the era of social computing: towards a cross-policy decision support system. In *The Semantic Web: ESWC 2011 Workshops*, page 217228. Springer.
- Charalabidis, Y., Triantafillou, A., Karkaletsis, V., and Loukis, E. (2012). Public policy formulation through non moderated crowdsourcing in social media. In *Electronic Participation*, page 156169. Springer.
- Diakopoulos, N. A. and Shamma, D. A. (2010). Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 11951198. ACM.
- Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, page 417422.
- Giannakopoulos, G. and Palpanas, T. (2010). Content and type as orthogonal modeling features: a study on user interest awareness in entity subscription services. volume 3, page 296309.
- Giannakopoulos, G., Karkaletsis, V., Vouros, G., and Stamatoopoulos, P. (2008). Summarization system evaluation revisited: N-gram graphs. volume 5, page 5.
- Giannakopoulos, G., Vouros, G., and Karkaletsis, V. (2010). MUDOS-NG: Multi-document summaries using n-gram graphs (tech report).
- Giannakopoulos, G., Kiomourtzis, G., and Karkaletsis, V. (2014). *NewSum: "N-Gram Graph"-Based Summarization in the Real World*. IGI.
- Goudas, T., Louizos, C., Petasis, G., and Karkaletsis, V. (2014). Argument extraction from news, blogs, and social media. In *A. Likas and K. Blekas and D. Kalles (Eds.): SETN 2014, LNCS 8445*, pages 287–299. Springer International Publishing Switzerland.
- Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, page 168177. ACM.
- Maragoudakis, M., Loukis, E., and Charalabidis, Y.

- (2011). A review of opinion mining methods for analyzing citizens contributions in public policy debate. In *Electronic Participation*, page 298313. Springer.
- Petasis, G., Spiliotopoulos, D., Tsirakis, N., and Tsantilas, P. (2013). Argument extraction from news, blogs, and social media. In *Proceedings of PATHOS 2013*.
- Petasis, G., Kanoulas, E., Karkaletsis, V., and Tsochan-taridis, I. (2014). D4.3.1: Argument extraction module.
- Radev, D. R., Jing, H., Sty, M., and Tam, D. (2004). Centroid-based summarization of multiple documents. volume 40, page 919938.
- Rao, D., Yarowsky, D., Shreevats, A., and Gupta, M. (2010). Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, page 3744. ACM.
- Remus, R., Quasthoff, U., and Heyer, G. (2010). Sentiws-a publicly available german-language resource for senti-ment analysis. In *LREC*.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welp, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. vol-ume 10, page 178185.
- Wandhofer, T., Van Eeckhaute, C., Taylor, S., and Fernan-dez, M. (2012). Wegov analysis tools to connect policy makers with citizens online. In *Proceedings of the tGov Conference May 2012, Brunel University*, page 7.