Inf Syst Front (2017) 19:109–127
DOI 10.1007/s10796-015-9596-8

Extracting the patterns of truthfulness from political information systems in Serbia

Nenad Tomašev[1]

Published online: 26 September 2015

Abstract In modern information societies, there are information systems that track and log parts of the ongoing political discourse. Due to the sheer volume of the accumulated data, automated tools are required in order to enable citizens to better interpret political statements and promises, as well as evaluate their truthfulness. We propose an approach to use the established machine learning and data mining techniques for analyzing annotated political statements and promises available via the Serbian Truth-o-meter (Istinomer) system in order to extract and interpret the hidden patterns of truthfulness and deceit. We perform standard textual processing and topic extraction and associate topical truthfulness profiles with the promise makers, for pattern discovery and prediction. Prevailing trends in Serbian political discourse emerge as strong association rules where truthfulness is set as the target variable. The evaluated set of standard content-based prediction models exhibit a bias towards the negative outcomes, due to an overall low truthfulness rate in the data. Our results demonstrate that it is possible to use data mining within political information systems for generating insights into the workings of governments.

Keywords Data mining · Text mining · Information systems · Politics · Truthfulness · Association rules

✉ Nenad Tomašev
nenad.tomasev@gmail.com

[1] Artificial Intelligence Laboratory, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

1 Introduction

Governments across the world are using various systems for tracking public opinion and citizen activity and some concerns have recently been raised about the significant decrease of privacy in the modern age (Danna 2002; Cate et al. 2012; Custers et al. 2013; Sanches et al. 2013; Vaidya 2012; Mostafa and El-Masry 2013). Transparency is important in order to prevent corruption and the abuse of power. The public requires access to information that tracks government policies and/or historical political discourse in the media. This is especially true for countries undergoing a political transition towards modern democracy, such as Serbia.

A high rate of corruption in the post-communist European countries (Vachudova 2009) has lead to an increase in apathy and a general loss of interest among the voters (Greenberg 2010). Previous studies have shown that the overall truthfulness of political discourse varies greatly across different countries and that the rate of promises being kept is far from uniform (PÃtry 2009). Lying has been identified as a commonly used tool (Cliffe et al. 2000). An in-depth analysis of the political discourse has become an important topic (Fairclough and Fairclough 2013). However, modern data analytic techniques have not yet become widely used when addressing these issues.

We propose to use a semi-automated data mining approach for uncovering patterns in online textual archives of political promises. As a use case, we analyze the annotated set of promises made by politicians in Serbia over the span of several years. The analysis was performed on the publicly available Truth-o-meter data (Istinomer, http://istinomer.rs/). The proposed approach is language-independent and therefore easy to generalize for analyzing political promise archives in other regions of the world.

Text mining (Miner et al. 2012) is a form of data mining that aims at extracting high-quality information from text by detecting hidden patterns and trends (Zhong et al. 2012). This usually involves some sort of natural language processing (Jackson and Moulinier 2007). Text mining techniques are used throughout many different domains (Dörre et al. 1999), including life sciences (Dai et al. 2010; Uramoto et al. 2004), media, marketing and healthcare applications.

However, up until now, text mining has not been widely used in political discourse analysis. As political debates tend to be quite complex, the analysis is often done manually by experts and analysts. Text mining has been used for assessing public opinion on certain topics (Balasubramanyan et al. 2010) and predicting election outcome (Murray et al. 2009; Campbell 2008). These automated tools are used mostly by governments and political parties in order to adapt their public relations strategies (Howard 2005) and are usually not available to the general public for monitoring the performance of appointed representatives.

This paper is structured as follows. We begin by summarizing the major contributions of our work in Section 1.1 and reviewing related research in Section 2. Serbian Truth-o-meter data is described in detail in Section 4, as well as the information retrieval and text processing techniques that were used to produce the final data representation. Section 5 presents exploratory visualizations of some important trends in the data and sets ground for more complex analysis and introduces the reader to the topic. In Section 6, a way for detecting typical promises in the corpus is proposed. Section 7 compares topical truthfulness profiles of different politicians and discusses most similar politician pairs. Finally, Section 8 discusses possibilities for content-based automated truthfulness prediction, as well as a set of association rules mined from the examined corpus. Section 9 summarizes the results and Section 10 concludes the paper by setting directions for improvements and plans for future work.

1.1 Contributions

This paper is among the first to apply modern analytic techniques for truthfulness prediction in political information systems and the first one that addresses this issue in Serbian Truth-o-meter data in particular. This is important due to the well-known intricate dynamics of Balkan politics.

Political promise statements and comments were represented in a language-independent way via n-grams and extended by automatically extracted topics. Topical truthfulness tendencies were used for discovering truthfulness patterns in the data along with the original textual representations.

Topical promise profiles were generated for all politicians present in the database and utilized for prediction, as well as cross-politician comparisons. The initial results suggest that politicians in similar roles tend to make very similar promises. This might come as a slight surprise, as the data covers several government mandates where the ruling parties often had opposing political programmes.

Automatic extraction of association rules was used for pattern discovery in the data, with truthfulness as the target variable. These patterns show that Serbian politicians are more/less likely to be truthful on certain sets of topics than others. For example, if there is a promise made in a name of a political party regarding building new infrastructure during the election campaign - it is most likely not going to be fulfilled.

Finally, we have evaluated several standard predictive models for on-line promise truthfulness prediction. It was determined to be a challenging task in the absence of metadata and the predictions were skewed towards the negative outcomes. However, not all promise makers were found to be equally difficult to predict, which gives rise to optimism about future research directions and better and more effective prediction pipelines.

2 Related work

Public opinion mining and topic popularity analysis is usually based on data acquired from news websites (Scharl and Weichselbraun 2008), blogs (Adamic and Glance 2005) and other social media (Helbing and Balietti 2011; Stieglitz and Dang-Xuan 2012), like Twitter (Balasubramanyan et al. 2010) and Facebook. Relevant information can be extracted both from posts and comments, as well as likes and/or dislikes, if available. An alternative is to infer topic trends by mining search queries (Weber et al. 2012).

Many trends can be observed by simply analyzing text via standard bag-of-words or n-gram models, but more fine-grained approaches have lately become quite popular by employing different sorts of sentiment analysis and opinion mining (Maragoudakis et al. 2011; Liu 2007). It is possible to determine the attitude of an article or a comment by analyzing the sentiment expressed in individual sentences towards the mentioned terms and concepts (Pang and Lee 2008). Some a-priori knowledge of the positive and negative words and constructs of the language is required and is usually loaded from specially prepared databases and dictionaries (Baccianella et al. 2010).

Social network analysis can be incorporated into the textual analysis of informal online discourse for determining the political beliefs of forum participants (Malouf and Mullen 2008).

Large-scale online sentiment analysis in the context of tracking political trends is often done with intent of predicting or influencing the results of the upcoming elections (Murray and Scime 2010; Grosskreutz et al. 2010). News sources are a valuable source of information for tracking informal political discourse as well and are often analyzed alongside the associated blogs (Gamon et al. 2008).

Roll call voting data has been incorporated in various legislative studies (Clinton et al. 2004; Jackman 2001). Special attention has been given to incorporating recording bias into the models in order to ensure the correctness of the analysis (Carruba et al. 2006).

Modern information systems can also be quite useful for the general public, as they enable voters to directly express their opinions on certain public policy proposals (Charalabidis et al. 2012; Loukis and Charalabidis 2012; Charalabidis and Koussouris 2012) and interact with the government via e-government systems (Rana et al. 2013; Weerakkody et al. 2013; Janssen et al. 2012). These crowdsourcing legislative models represent a step towards a more open and direct democracy. Some information systems are capable of detecting the early spread of political misinformation, whether random or intentionally induced (Ratkiewicz et al. 2011).

Ways of dealing with factual inconsistencies and the dissemination of misinformation are becoming a necessity. The public seems to be overwhelmed with media information and is often unable to reliably distinguish between facts and fabrications. This is why some media groups have begun working on specialized fact-checking services, introducing truth-o-meters and databases of statements and promises. PolitiFact (http://www.politifact.com/) is one such service. The Serbian analogue is the Istinomer website (http://istinomer.rs/) that we will be analyzing in this paper.

PolitiFact and Istinomer are not the only initiatives for establishing ground-truth factual information about political discourse. Some notable examples include http://www.truthfulpolitics.com/ and http://www.factcheck.org/.

3 Overview of the proposed approach

The proposed approach for detecting patterns of truthfulness in political promise data involves multiple steps. Figure 1 gives a brief outline of the major components and goals in the analysis. All of these steps will be discussed in full detail in the following Sections.

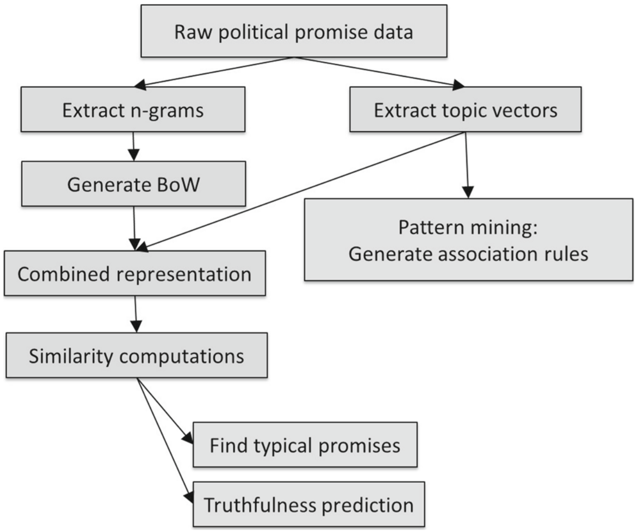


Fig. 1 A high-level overview of the proposed approach for analyzing political promise data

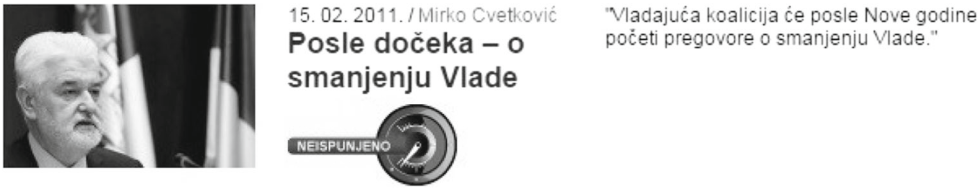
The approach relies on two separate feature extraction components. The syntactic component provides the basic textual representation used in the analysis. This syntactic representation is semantically enriched by topic extraction in order to enable topic-level analysis of the data. These representations are then used either separately or combined in subsequent steps in the analysis. Topic-level representation is better suited for pattern mining while the enriched syntactic representation can be used for similarity-based prediction and prototype selection or clustering.

For different languages, the process can be either fully automated or semi-automatic. This depends on the overall availability of natural language processing tools and services for each language in particular.

4 Data

Information retrieval approaches implicitly assume that the fetched data had not previously been falsified and that the overall noise levels are acceptably low. However, in heated political debates, opposing politicians often make entirely opposite claims and there is a high level of inaccurate or false information being spread. The discourse data is spread over diverse heterogenous sources and knowledge integration remains a difficult technical challenge.

In our analysis, we have focused on using the expert annotated political promise/statement data. In Serbia, this service is provided by Istinomer, similar to the PolitiFact Truth-o-Meter in the US.



15. 02. 2011. / Mirko Cvetković

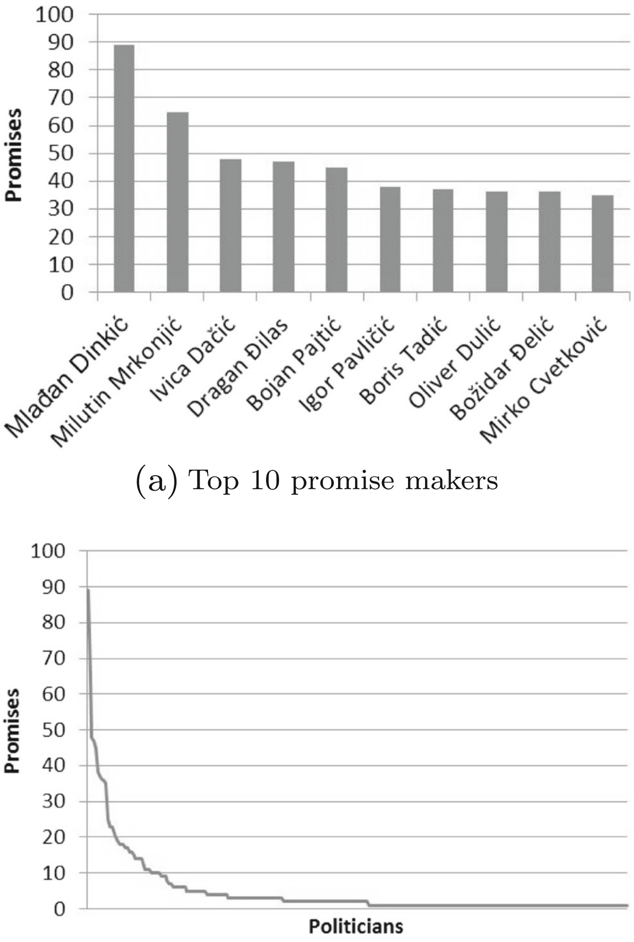
Posle dočeka – o smanjenju Vlade

"Vladajuća koalicija će posle Nove godine početi pregovore o smanjenju Vlade."

Fig. 2 An example from the Istinomer website: a promise given by Mirko Cvetković stating that the size of the government is going to be reduced in January 2011. This promise was not fulfilled

4.1 The Istinomer project

Istinomer is a non-government project lead by a group of news reporters, political analysts and actively participating citizens that help in populating the database of political statements, promises, comments and insults. The readers may participate by proposing certain actions and initiatives. The goal of the Istinomer team is, as stated on the webpage, to enforce responsibility when addressing the public and to keep the people from forgetting the controversial events and affairs.



(a) Top 10 promise makers

(b) The total distribution of promises (sorted)

Fig. 3 The number of promises given by politicians and processed by Istinomer, aggregated by politician

The content offered via the Istinomer web portal is divided into several sections. The major sections include: promises, truthfulness, consistency, insults, investigations, look-back, notes, topics, polls and a list of tracked political figures. The truthfulness section covers the statements that are obviously false at the time when they are given, when a simple check is sufficient to uncover the real truth. The promises section deals with the announcements of future plans, promises that are given to the voters at a certain time in order to gain their trust and support. An example is given in Fig. 2.

Even though analyzing the factual truthfulness data is interesting in its own right, this paper focuses on analyzing the truthfulness of political promises instead. Factual truthfulness can more easily be checked with little or no computer aid, so there is less need for using automated analytic support systems. However, truthfulness of given promises can not be reliably estimated at the time when they are given. This is where automated profiling can potentially be useful.

The following information is provided within each promise page: title, upload date, date of the initial statement, politician name, short statement (usually a sentence or two, sometimes more), as well as a one-line contextual clarification and a comment/analysis by the Istinomer team. Public comments are allowed, but rarely used, so most articles contain no user comments.

We have analyzed both the short political statements themselves and the more detailed Istinomer comments (pruned of certain words, see Section 4.2). The latter are similar in content to news articles reporting on promises and statements in real-time.

4.2 Preprocessing and data representation

We have implemented a web-crawler in Java and downloaded the content of the webpages containing the reviews of the political promises on the Istinomer website. The data was crawled on 21.7.2013., so the data contains only the promises that were evaluated and published on the website up to that point. This dataset comprises 1380 promise statements and comments made by 273 different politicians.

The distribution of promises is highly skewed to the right, as shown in Fig. 3. Most statements in the corpus were given by a small number of very influential officials and many

Table 1 Fuzzy quantification of qualitative truthfulness estimates

Qualitative estimate	θ
"the work has not even started"	0
"unfulfilled"	0
"started but then stopped"	0.25
"working on it"	0.5
"almost fulfilled"	0.75
"fulfilled"	1

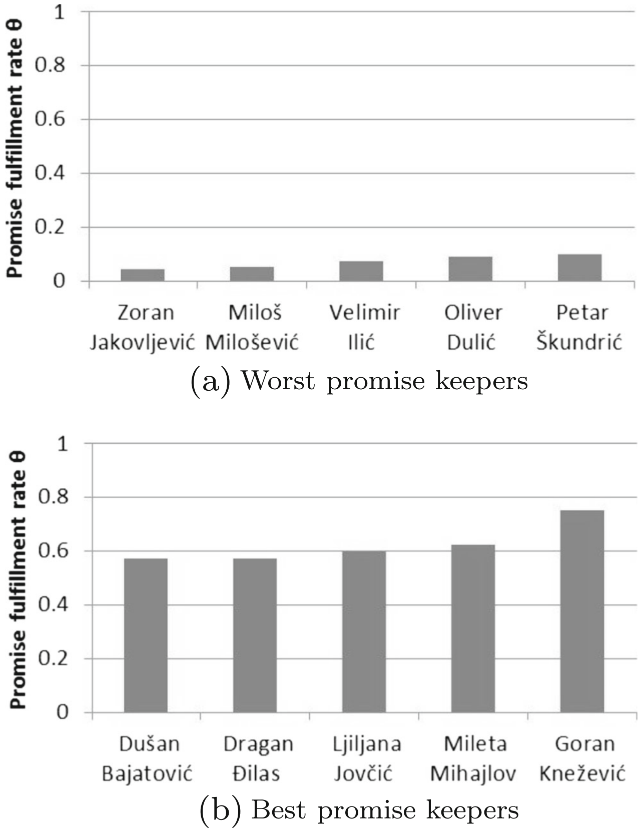
politicians have only 1 or 2 assigned promises. This follows the distribution of political power, as the more powerful politicians are able to make and fulfil more promises. Additionally, they are the usually in the focus of political studies, so this might also have been caused by a reporting bias.

The Istinomer promises data is not annotated with any tags or keywords that might help with topic detection. This means that all analysis must be based on the free text statements and comments. The data is available exclusively in Serbian.

Unlike for English, there is no stable and reliable text mining and natural language processing toolkit for Serbian that has been thoroughly tested in practice and used consistently in various analytic projects. These tools are currently being developed and improved by several research teams (Vitas et al. 2003). Attempts are being made at building an effective stemmer for Serbian (Vlado and Šipka 2008; Milošević 2012), but a quick analysis reveals that this particular rule-based implementation does not perform very well on complex texts and discussions.

A language-independent approach for handling statement and comment texts was applied, via character n-grams (Damashek 1995), instead of the more usual bag-of-words representation. For reasons stated above, no stemming or lemmatization was performed.

Character n-grams have been successfully used in the past in many text mining applications, including plagiarism detection (Stamatatos 2009), document categorization (Cavnar and Trenkle 1994), authorship attribution (Kešelj et al. 2003) and named entity recognition (Klein et al. 2003). Previous studies have shown that using $n = 4$ is a good choice for most European languages, so we have followed this recommendation and used 4-character long sequences in our own analysis of the Istinomer data.



(a) Worst promise keepers

(b) Best promise keepers

Fig. 5 The best and worst promise keepers, among those that have given at least 5 promises

Some comments had an explicit mention of whether the promises were fulfilled or not within the comment text, so all the occurrences of "fulfilled" ("ispunjeno") and "unfulfilled" ("neispunjeno") were removed from Istinomer comments prior to extracting the n-grams. Also, all of the special punctuation characters have been removed from the texts and all the whitespace characters have been merged

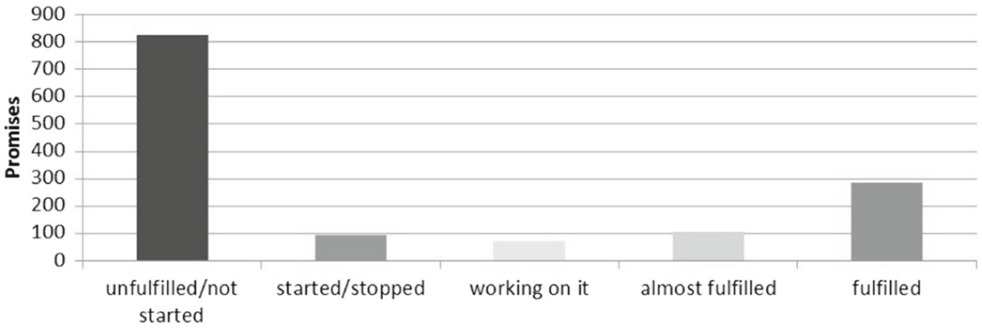
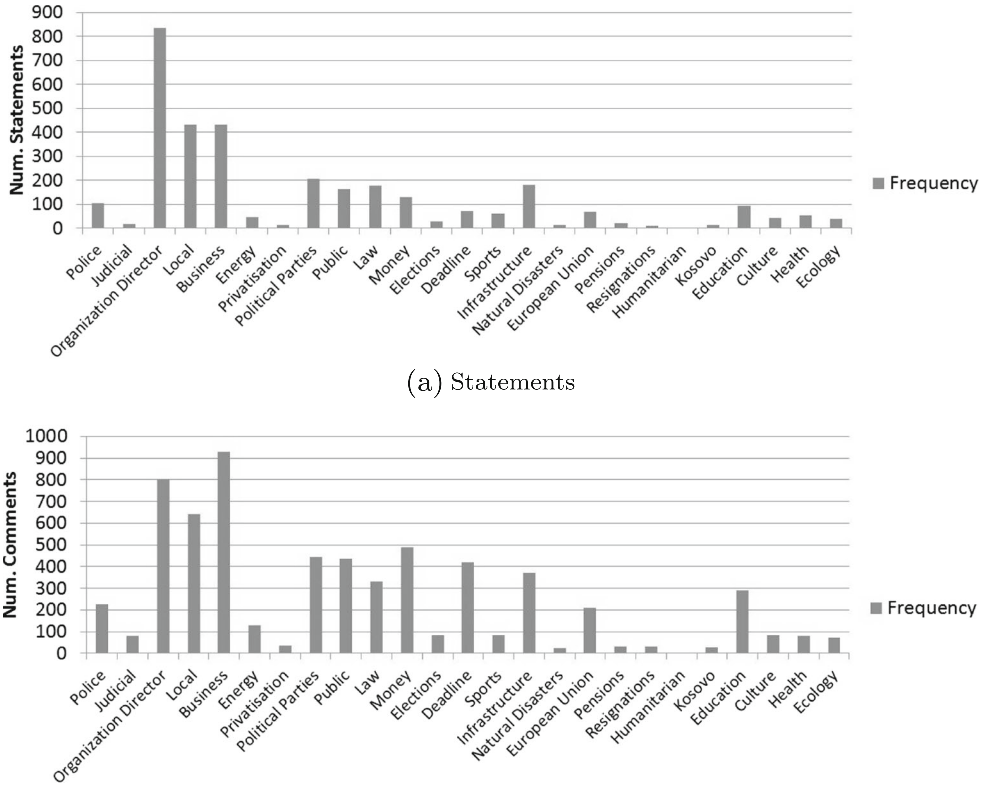


Fig. 4 The distribution of truthfulness in Istinomer promise data

Fig. 6 The number of positive topic matches in Istinomer promise comments and statements



(a) Statements

(b) Comments

to a single whitespace delimiting separate words. All words were transformed to lower case.

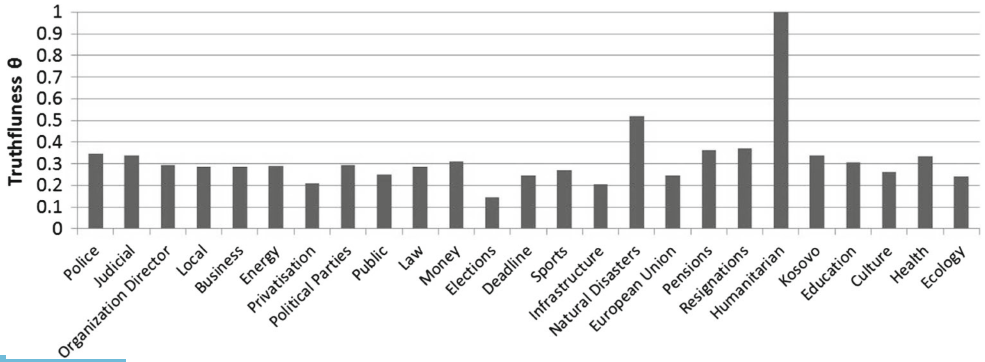
We have implemented a simple n-gram extractor in Java and have generated the n-gram representations both for the statements and the associated comments made by the Istinomer team. In the following analysis, we will examine the results obtained from the statement and comment data separately.

The short statement texts were merged with the one-sentence contextual clarifications, that contain some basic semantic information and nothing that would point directly towards the truthfulness of the statements. Here is an example of a statement context: “The Minister of Education about the construction of the student dormitory in Čačak.” (“Ministar prosvete o izgradnji studentskog doma u Čačku.”). This contextual clarification was given for the statement where the minister promised that the construction of the dormitory will begin on June 1st 2013, which has not been fulfilled.

4.3 Quantifying truthfulness

Each Istinomer article presents an evaluation of a certain political promise both in form of a free text comment and a truth-o-meter score. This score is qualitative and can take one of the following values: “the work has not even started” (“ni započeto”), “unfulfilled” (“neispunjeno”), “started but

Fig. 7 The average truthfulness of topics automatically extracted from Istinomer promise comments



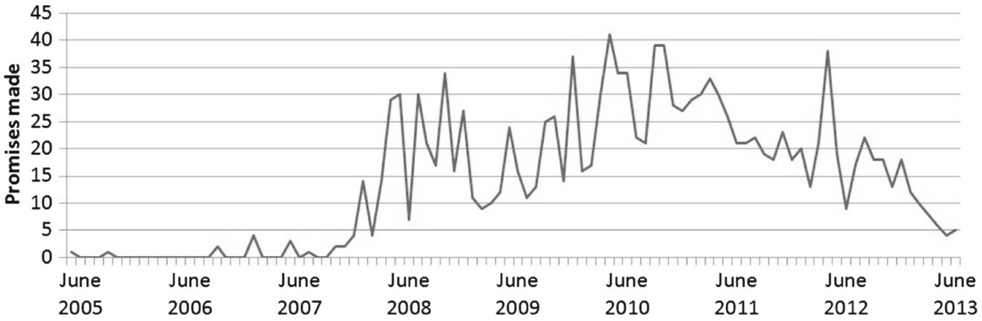


Fig. 8 The number of promises made by government officials in Serbia, over the past couple of years

then stopped" ("krenuli pa stali"), "working on it" ("radi se na tome"), "almost fulfilled" ("skoro ispunjeno"), "fulfilled" ("ispunjeno"). In order to be able to measure cumulative truthfulness over many statements, we have decided to assign each of these qualitative grades with a number $\theta \in [0, 1]$, approximately quantifying the truthfulness of the statement. The fuzzy approximation was done as shown in Table 1.

It should be noted that this fuzzy interpretation of the qualitative promise fulfilment tags is not unique and not necessarily optimal. Ideally, we would be working directly with the objective quantitative scores obtained by collaborative grading. In practice, such high-quality evaluations are rarely available and are not provided for Istinomer database.

The overall truthfulness of political promises in Serbia is very low, the average being merely 0.305 - which is far less than 50 %. It is also much less than what was reported in some similar studies for other European countries (PÃtry 2009), where the authors claimed that the political parties fulfill 67 % of their promises, on average. This discrepancy might indicate that there is an actual difference in truthfulness and that the overall truthfulness in Serbia is lower than in most developed democracies or, alternatively, that there was a certain data acquisition bias. It might be the case that the Istinomer team is focusing on unfulfilled promises for delivering their critique or that the previous studies have had a similar positive bias for interpreting truthfulness rates. The distribution of truthfulness in Istinomer promise data is shown in Fig. 4.

The least truthful and most truthful promise makers are shown in Fig. 5. The two least truthful promise makers, according to Istinomer data (by July 2013), are Zoran Jakovljević and Miloš Miloševic, the ex-mayor and city councilor of Valjevo and the ex-mayor of Šabac. They are closely followed by Velimir Ilić, a controversial figure and the founder of the 'New Serbia' political party. As for the promise fulfillers, Goran Knežević seems to have the highest truthfulness score, regardless of the aflatoxin controversy in 2013 (http://goo.gl/uhXg2Z). The second best promise fulfiller is Mileta Mihajlov, the ex-mayor of Zrenjanin.

4.4 Topic extraction

Different promises relate to different types of social and economic problems. In order to perform a detailed analysis of the political statements, it is necessary to assign some higher-order semantics to the processed statements. We propose to approach the problem by automatically detecting the topics from statement and comment texts.

Forming topic models from text is a difficult task and various types of approaches have been developed in the past, including clustering (Seo and Sycara 2004), keyword-based approaches (Wartena and Brussee 2008), singular value decomposition and independent component analysis (Hamamoto et al. 2005), as well as the frequently used latent Dirichlet allocation (LDA) (AlSumait et al. 2008). Many of these approaches are unsupervised and based on an assumption that the full set of topics is not known in advance. However, the political statements address certain topics of public interest that can easily be defined manually.

In our analysis of Istinomer data, we have manually defined 25 different topics, as follows: *Police, Judicial, Organization Director, Local, Business, Energy, Privatisation, Political Parties, Public, Law, Money, Elections, Deadline, Sports, Infrastructure, Natural Disasters, European Union, Pensions, Resignations, Humanitarian, Kosovo, Education, Culture, Health, Ecology*. Many of these topics

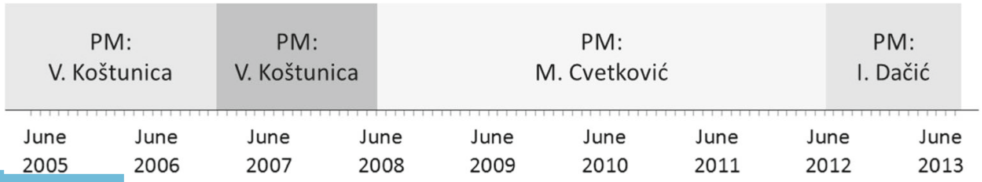


Fig. 9 Governments in Serbia, in the period analyzed in the Istinomer data

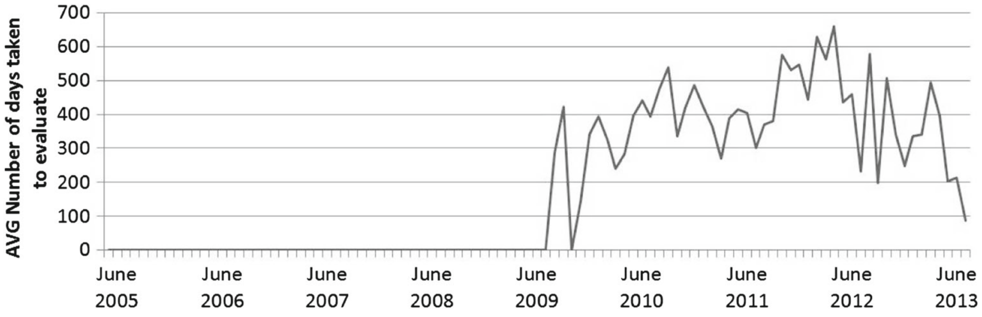


Fig. 10 The average 'age' of evaluated promises, measured in days

pop up very frequently in daily political debates. Court processes (topic: Judicial) have been separated from passing laws (topic: Law), since applying and proposing regulations are two different things.

Automatic topic detection in statements and comments was performed via regular expressions that contained sets of of positive and negative indicator keywords for each topic within the text. The used keywords were mostly stems of words that are often used in certain contexts. A list of keywords was compiled after several iterations of careful examination of the Istinomer data and the vocabulary that is used within the texts.

The vocabulary that the politicians use when addressing the public is kept quite simple, unlike in literary texts. This is beneficial for this type of regular expression topic extraction. In fact, an initial evaluation of the implemented regular expression based topic extraction component showed good performance on the observed subsample of the comment data. It achieved a precision of 95 % and a recall of 91 %, which is quite satisfactory given its simplicity. In future work we intend to explore more complex topic models as well in order to achieve even better performance and to make the approach easier to generalize to other data sources that incorporate richer vocabularies and more challenging phrasing.

Figure 6 shows the number of positive topic matches for statements and comments, respectively. Not surprisingly, there were more topic matches among the comments (6368 compared to 3265), as they are much longer than the statements themselves. More than 2 topics per statement are detected on average and more that 4 topics per comment. This is quite satisfactory. Also, the shape of the two discrete topic distributions is very similar.

In both topic distributions, Business and Local are the next leading topics. The Local topic discusses issues regarding governance of individual cities or regions. They are followed by several other frequent topics: Money, Public, Law, Infrastructure, Political Parties, Education, European Union and Police. The least frequent topic is the humanitarian one. It should also be noted that the detected topic frequency is not directly correlated with the number of keywords used for topic detection and is not an artefact of the topic extraction process.

Different levels of truthfulness can be seen when addressing different topics, as shown in Fig. 7. Serbian politicians are most truthful when addressing humanitarian issues in case of natural disasters and least truthful when addressing the elections, infrastructure, privatisation and the European Union. It should be mentioned, though, that the high truthfulness in case of humanitarian issues stems from a very low number of positive humanitarian examples in the Istinomer promise dataset.

4.5 Measuring promise similarity

In some of the following experiments, mostly in Sections 6 and 8, a similarity-based analysis of the promise data is discussed. As the data is represented by the standard vector space model (Raghavan and Wong 1986) in the 4-gram feature space, the similarity between two statements and/or comments is defined as the cosine similarity between the corresponding vectors (Feldman and Sanger 2006). V =

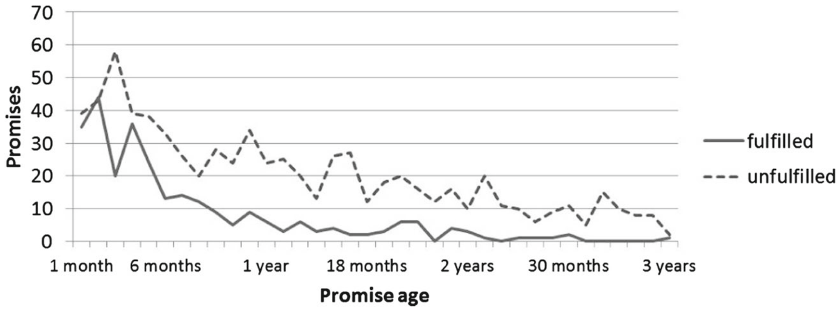


Fig. 11 The number of fulfilled and unfulfilled promises of various ages

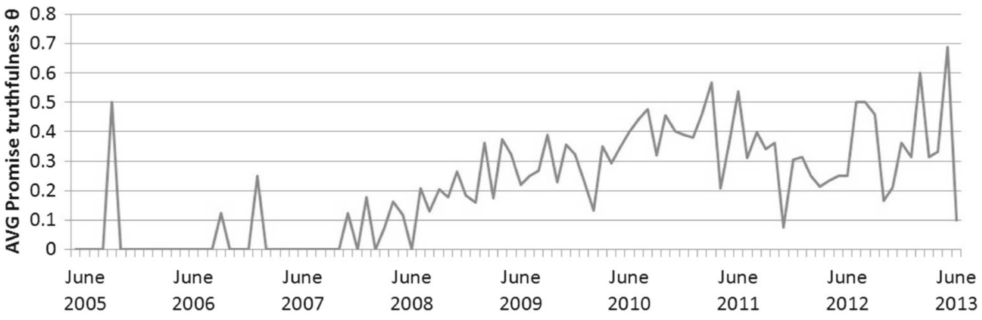


Fig. 12 The evolution of truthfulness in Serbian political discourse, based on Istinomer promise data

$[v_1, v_2..v_W]$ is the 4-gram vocabulary that defines a $|V|$-dimensional feature space. Statements and comments are represented as vectors of weights, determined by the widely used TF-IDF (term frequency, inverse document frequency) rule (Feldman and Sanger 2006). Let P be the set of all examined promises and let $\vec{x_i} = [x_i^1, x_i^2 \ldots x_i^W]$ and $\vec{x_j} = [x_j^1, x_j^2 \ldots x_j^W]$ be the vectors representing two promise texts. The cosine similarity between the promises is then calculated as in Eq. 1.

$$sim(\vec{x_i}, \vec{x_j}) = \frac{\vec{x_i} \cdot \vec{x_j}}{\|\vec{x_i}\| \cdot \|\vec{x_j}\|} = \frac{\sum_{z=1}^{W} x_i^z \cdot x_j^z}{\sqrt{(\sum_{z=1}^{W}(x_i^z)^2) \cdot (\sum_{z=1}^{W}(x_j^z)^2)}} \qquad (1)$$

5 Visualizing the emerging trends

Political discourse evolves over time and changes with governments, coalitions, treaties, as well as various internal and external factors. As the promise and evaluation dates were attached to each Istinomer article, we have used this information to plot several interesting temporal trends.

Figure 8 shows the temporal promise landscape - the number of promises given by various government officials changes over time in a non-uniform way. The curve is far from smooth, with many maxima and minima. The last major recorded peak was in 2012, at the time of the Serbian parliamentary and presidential elections (see Fig. 9). These elections were quite uncertain until the very end and all the involved parties were trying to convince the voters by promising many improvements and future investments. The apparent decline at the very end of the curve is a bit deceiving, as there might be some long-term promises there that are yet to be evaluated by the Istinomer team in the future.

After a promise is made, some time needs to pass before its truthfulness can be evaluated, as large projects take a lot of time to complete. Therefore, there is a notable delay between the dates at which the statements are made and the evaluation dates when Istinomer comments are posted on the website. Figure 10 shows the temporal evolution of the average time between the promise statement and the promise evaluation. Even though there are some short-term promises that are immediately evaluated, the average evaluated promise age often exceeds one year. This means that there are many long-term projects that have been announced. Unfortunately, according to the assigned truthfulness score, most of them have not been accomplished.

The number of fulfilled and unfulfilled promises for different promise ages is shown in Fig. 11. There is essentially the same number of fulfilled and unfulfilled promises among those that are evaluated within a month or two from when they are made. Afterwards, there is a much higher number of unfulfilled that fulfilled promises, across the examined promise age range. This indicates that the very short term promises might be somewhat easier to fulfil.

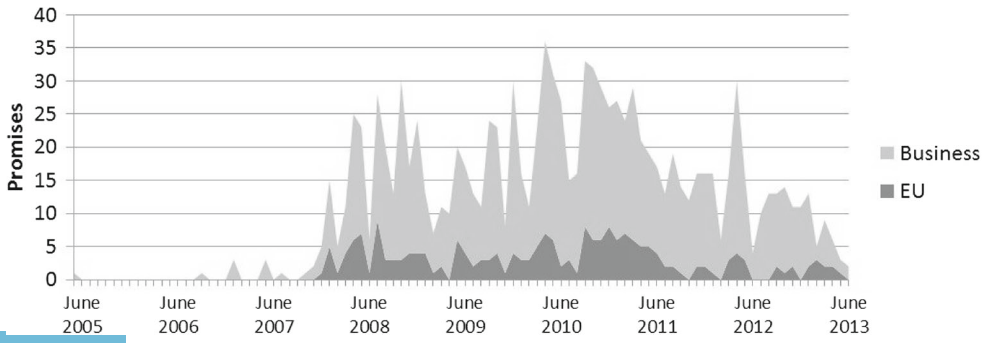


Fig. 13 The evolution of business and EU topics in Istinomer promise data

The average promise fulfillment levels have evolved over time with different ruling governments and this trend is shown in Fig. 12. Unfortunately, there is very little data for the period prior to 2008 and it has only been one year since the 2012 elections at the time this data is being analysed - so we are unable to make statistically valid comparisons between the truthfulness of different ruling coalitions at this time.

Promise topics also evolve over time. We have generated the monthly-binned topic trends for all the topics extracted from statements and comments. Figure 13 shows the co-evolution of business and European Union mentions in Serbian political promises. There is an interesting trend: each peak in EU mentions corresponds to an even larger peak in business-related promises. This might indicate that the government is considering all sorts of trade agreements and EU-based investments and projects at the same time when it is working on finishing certain stages of the process towards the EU membership. Of course, the correlation could partly also be attributed to the overall increase in promise rates at certain pre-election times, though it would not be able to account for most of the existing peak alignments.

6 Typical statements as promise hubs

In order to better summarize the data, it is useful to examine the central and most influential concepts, in this case - the most typical promises being made in various topics. We have extracted the typical promise statements by analyzing the promise n-gram feature representations and exploiting some recently described high-dimensional phenomena related to the structure of the k-nearest neighbor graphs (kNN).

Textual data is intrinsically high-dimensional, due to its rich semantics. Therefore, it is known to exhibit substantial *hubness* (Nanopoulos et al. 2009). In intrinsically high dimensional data, centers of influence emerge as exceedingly frequent k-nearest neighbors, under many common similarity measures. These highly influential points are known as *hubs*. Hubness is an important phenomenon, since k-nearest neighbor methods are widely used in many machine learning systems and applications.

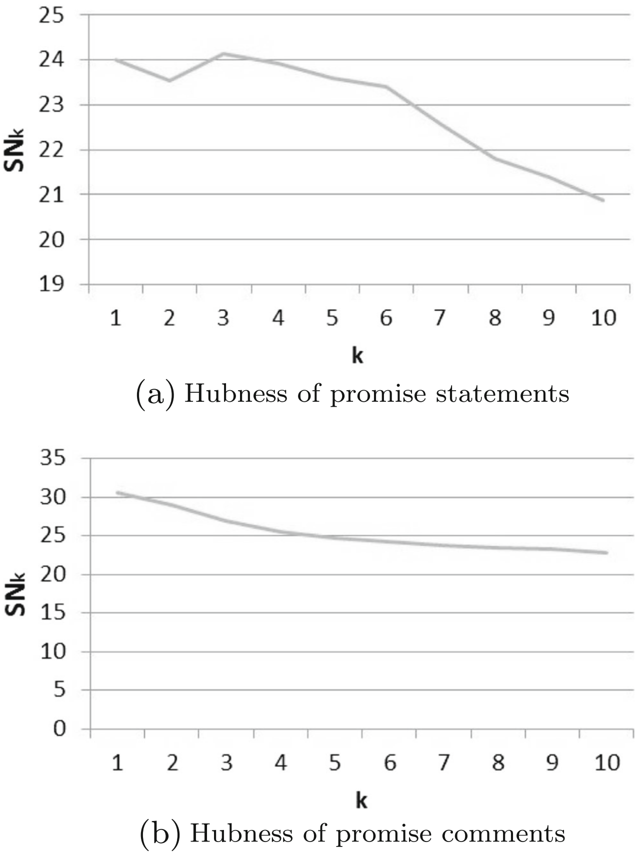
Let $N_k(x_i)$ denote the total neighbor k-occurrence frequency of x_i. The overall hubness of the datasets is defined as the skewness of the neighbor occurrence frequency distribution, as follows:

$$SN_k = \frac{\frac{1}{n}\sum_{i=1}^{n}(N_k(x_i) - k)^3}{(\frac{1}{n}\sum_{i=1}^{n}(N_k(x_i) - k)^2)^{3/2}} \quad (2)$$

A high skewness of the neighbor occurrence distribution means that a small number of points occurs very frequently in k-neighbor sets (*hubs*) and most points either never occur as neighbors or do so very rarely. Hubs are, therefore, on average, very similar to many other points, which allows them to become frequent neighbors in the feature space. The hubness phenomenon is closely related to the concentration of distances (François et al. 2007), a pathological behavior of many standard metrics in intrinsically high-dimensional data.

In high-dimensional feature spaces, data lies approximately on hyper-spheres around local cluster means. It has recently been shown that the neighbor occurrence frequency is highly correlated with cluster centrality (Tomašev et al. 2013) and that hubs emerge as the most central concepts in the observed metric space.

Due to their centrality, hubs can be used as cluster prototypes (Tomašev et al. 2013) and they are often used as representatives when analyzing PPI (protein-protein interaction) networks (He and Zhang 2006), where they often play the role of essential proteins. They have also been used for representing typical word meanings (Agirre et al. 2006).



(a) Hubness of promise statements

(b) Hubness of promise comments

Fig. 14 The k-neighbor occurrence frequency skewness (hubness) of promise statements and comments in 4-gram representations, under cosine similarity, for several different neighborhood sizes (k)

The Istinomer promise data exhibits very high hubness, as shown in Fig. 14. Any value of $SN_k > 1$ can be considered high, and here the neighbor occurrence skewness exceeds 20, both for the statements themselves and the associated comments. This means that the Istinomer promise data harbours hubs and that it is possible to extract the central and influential concepts by finding the major hubs in the data.

In order to extract the typical promises, we have constructed a 1-NN graph both for statement and comment texts and calculated the neighbor occurrence frequencies. The hub articles with most occurrences have been taken as typical promise representatives.

We have applied the proposed approach for detecting topic-specific typical statements, by looking for the major hubs as the most frequent nearest neighbors among the articles within separate topics. The list of typical topic-specific statements extracted as major topic-hubs is shown in Table 2. In the Money topic, the article on highway construction was assigned to it because a bank loan is mentioned in the statement text. “Several weeks, several weeks” is an article that sets the deadline for passing a strategy for Kosovo. In the statement under the title “I will be a president of Serbia, not of SNS”, Tomislav Nikolić promises to resign as the president of SNS if he gets elected president of Serbia, which he did. All extracted hub promises seem to correspond well to their topics.

7 Detecting similar promise-makers

We have compared politicians based on their promise-making profiles. It is possible to create these profiles both from the statement and the comment texts - and it is possible to use either the n-gram representation or the assigned topic model. We have performed all 4 possible types of similarity comparisons. Truthfulness was not taken into account in this particular case.

We have created the politician promise-making profiles by averaging the representation vectors that represent all the promises that they have made. The n-gram promise representations were previously weighted by TF-IDF. The cosine similarity was used to measure the similarity between different politicians. In Table 3, we report the top 5 most similar

Table 2 The most typical statements for all the automatically extracted topics individually, selected based on article hubness

Topic	Title (original)	Title (translation)	θ
Police	Novi Sad dobija načelnika policije	Novi Sad is going to get a chief of police	0
Judicial	Advokati, izdavaćete fiskalne račune!	Lawyers will issue fiscal bills.	0
Org. Director	Novi Sad dobija načelnika policije	Novi Sad is going to get a chief of police	0
Local	A načelnika nema	And the chief is not there	0
Business	Jura u Leskovcu do kraja 2011.	Jura in Leskovac by the end of 2011	0.75
Energy	Model otkupa električne energije do 10. jula	Model of electric energy purchase by July 10th	0
Privatisation	Privatizacija JATa do kraja godina	Privatisation of JAT by the end of the year	0.25
Political Parties	Resori čekaju na većnike	Sections are waiting for city councilors	0
Public	Sa Bosnom protiv kriminala	With Bosnia against crime	1
Law	Penzije 70 % plate	Pensions 70 % of the salaries	0.25
Money	Autoput počinje, samo ugovor da imamo	Starting with the highway, we only need a contract	0
Elections	Ja predsednik	Me president	0
Deadline	Socijalni zakon na čekanju	The ‘social law’ on wait	0
Sports	Jagodinski plivači na suvom	Swimmers in Jagodina are left dry	0
Infrastructure	Izbori donose autoput	Elections will bring the highway	0
Natural Disasters	Pomoć za poplavljeni jug	Aid for the flooded South	0.75
European Union	I Evropa će nam zavideti na porezima	Europe will envy us for our taxes	0
Pensions	Penzije 70 % plate	Pensions 70 % of the salaries	0.25
Resignations	Biću predsednik Srbije a ne SNS	I will be a president of Serbia, not of SNS	1
Humanitarian	Srpsko-ruski humanitarni centar	Serbian-Russian humanitarian center	1
Kosovo	Nekoliko nedelja, nekoliko nedelja	Several weeks, several weeks	1
Education	Niški studenti dobijaju dom	The students in Niš are getting a dormitory	0
Culture	Medijska strategija na pomolu	Media strategy on the horizon	1
Health	Kraće liste čekanja	Shorter waiting lists	0
Ecology	Na proleće muljamo mulj	In Spring we will be mulling the mud	0

Table 3 The most similar promise-makers

(a) Similarity based on average extracted statement topics	
Igor Pavličić	Miloš Simonović
Dragan Šutanovac	Zoran Drobnjak
Zoran Stanković	Tomica Milosavljević
Milan Krkobabić	Dragan Đilas
Miloš Simonović	Slobodan Kocić
(b) Similarity based on average extracted comment topics	
Igor Pavličić	Dragan Đilas
Igor Pavličić	Miloš Simonović
Miloš Simonović	Dragan Đilas
Mlađan Dinkić	Mirko Cvetković
Mileta Mihajlov	Dragan Đilas
(c) Similarity based on average extracted statement n-grams	
Zorana Mihajlović	Petar Škundrić
Goran Knežević	Dušan Petrović
Zoran Stanković	Tomica Milosavljević
Mlađan Dinkić	Nebojša Ćirić
Dušan Petrović	Saša Dragin
(d) Similarity based on average extracted comment n-grams	
Zorana Mihajlović	Petar Škundrić
Slobodan Homen	Snežana Malović
Mlađan Dinkić	Nebojša Ćirić
Zoran Stanković	Tomica Milosavljević
Igor Pavličić	Dragan Đilas

politicians, according to each of the 4 examined approaches. Only the politicians that made at least 5 promises were taken into account, in order to avoid trivial similarities between almost empty representation vectors.

Many of the reported similarities can easily be accounted for. Igor Pavličić, DraganĐilas, Miloš Simonović, Mileta Mihajlov and Slobodan Kocić are all current or former city mayors and Milan Krkobabić was an advisor to the mayor of Belgrade for a while. Zoran Stanković and Tomica Milosavljević have both led the Ministry of Health in the past. Goran Knežević, Saša Dragin and Dušan Petrović were both ministers of agriculture and both Mlađan Dinkić and Nebojša Ćirić are economic experts. Zorana Mihajlović and Petar Škundrić were heads of the Ministry of Energetics.

We can conclude that our models are able to automatically infer semantically correct connections between different current and ex government officials based on their previous statements. This also means that people in similar positions tend to make similar types of promises. An example in Fig. 15 is given by a comparison between Mlađan Dinkić and Mirko Cvetković, who were both leading the finance sector at their time.

These comparisons between the topic models of different politicians are possible for any pair of politicians tracked by the Istinomer team and we hope that they might help in revealing some potentially interesting correlations in the context of political promise tendencies.

8 Promise fulfilment prediction

There are many reasons why some promises might remain unfulfilled or only partially fulfilled. Sometimes, there is no true intent for fulfilling a promise and it is only made to appease the public short-term and advertise or promote the political party prior to elections. At other times, it might be due to incompetence or a lack of required organizational and management skills. Finally, it might just be a random confluence of unexpected events. However, if the promises are being made responsibly and carefully, such random problems should not be occurring too frequently.

Our hypothesis was that political promise fulfilment is not entirely random and that there exist some patterns and regularities that could be discovered via standard data mining techniques.

8.1 Discovering hidden patterns with association rules

Association rule mining (Adamo 2001) is frequently used in data analysis (Cagliero and Fiori 2013; Hong et al. 1999) as it allows for unsupervised pattern discovery. Many algorithms exist that enable association rule extraction and they are implemented across a variety of data mining libraries. In our experiments, we have used the MagnumOpus software (http://www.giwebb.com/).

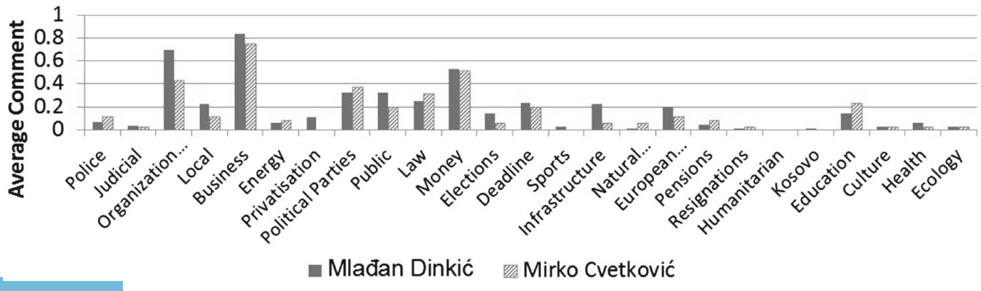


Fig. 15 A comparison between the promise comment topic models of Mlađan Dinkić and Mirko Cvetković, displaying high average topic similarity

The approach was applied to a list of topic vectors extracted from the Istinomer article comments, extended by truthfulness and politicians' names. Truthfulness was set as the target attribute for prediction. The numeric ranges for truthfulness were set as $\theta \in [0, 0.5]$ and $\theta \in (0.5, 1]$. This way, we have aggregated the promises that were mostly fulfilled and those that were mostly not fulfilled.

There are several ways to rank the association rules based on their informativeness (Adamo 2001). We have decided to rank them based on rule strength (confidence) (Agrawal et al. 1993), followed by leverage (Piatetsky-Shapiro 1991), for tie-breaking. Rule confidence measures the probability of the right-hand side being true, given the left hand side. Leverage also takes coverage into account.

Table 4 a list of top association rules extracted from Istinomer promise data, as ranked by the MagnumOpus association rule discovery tool. All of the below listed rules have a confidence of 1 and are only distinguished by the leverage, which measures statistical dependence between the involved variables on the left hand side and the right hand side.

All of the strongest rules shown in Table 4 predict a low chance of promise fulfillment. Rule AR4 says that if a promise involves the elections and the European Union, it is most likely not going to be fulfilled. Even more prominent is AR2, where we see that promises that involve building new infrastructure right before, during or after the elections and are initiated by certain political parties are also most likely not going to be fulfilled. The promises made by Mlađan Dinkić involving elections and the promises made by Petar Škundric involving monetary issues of loans and investments are not likely to be fulfilled.

While the more general uncovered truthfulness patterns may not come as a surprise to those familiar with the properties of general political discourse, the person-specific rules can be very helpful in understanding and interpreting the emerging political trends. In particular, it is possible to observe these rules in aggregate, based on political party affiliation and political orientation of the politicians being observed.

However, these rules are not so appropriate for on-line real-time promise fulfilment prediction. Strong rules have a rather low coverage, so it would not be possible to apply them very often. While these rules remain a useful abstraction of the semantics in the discourse, they do not take into account all of the promise content and some of the discarded details can prove to be very useful for prediction as they relate to the exact phrasing of promises and more specific micro-topics.

Table 4 The strongest association rules for predicting the promise truthfulness on Istinomer data

Index	Association rule	Conf.	Lever.
AR1	Politician=Igor Pavličić $\wedge$ Topic=Public $\Longrightarrow \theta \leq 0.5$	1	0.0033
AR2	Topic=Political Parties $\wedge$ Topic=Elections $\wedge$ Topic=Infrastructure $\Longrightarrow \theta \leq 0.5$	1	0.0031
AR3	Topic=Energy $\wedge$ Topic=Public $\wedge$ Topic=Law $\Longrightarrow \theta \leq 0.5$	1	0.0033
AR4	Topic=Elections $\wedge$ Topic=European Union $\Longrightarrow \theta \leq 0.5$	1	0.0031
AR5	Politician=Oliver Dulić $\wedge$ Topic=Infrastructure $\Longrightarrow \theta \leq 0.5$	1	0.0029
AR6	Topic=Law $\wedge$ Topic=Deadline $\wedge$ Topic=Infrastructure $\Longrightarrow \theta \leq 0.5$	1	0.0027
AR7	Topic=Political Parties $\wedge$ Topic=Law $\wedge$ Topic=Money $\wedge$ Topic=Deadline $\Longrightarrow \theta \leq 0.5$	1	0.0025
AR8	Topic=Political Parties $\wedge$ Topic=Elections $\wedge$ Topic=Deadline $\Longrightarrow \theta \leq 0.5$	1	0.0025
AR9	Topic=Police $\wedge$ Topic=Political Parties $\wedge$ Topic=Sports $\Longrightarrow \theta \leq 0.5$	1	0.0025
AR10	Politician=Božidar Delić $\wedge$ Topic=Public $\Longrightarrow \theta \leq 0.5$	1	0.0025
AR11	Topic=Public $\wedge$ Topic=Law $\wedge$ Topic=Culture $\Longrightarrow \theta \leq 0.5$	1	0.0025
AR12	Politician=Slobodan Homen $\wedge$ Topic=Political Parties $\wedge$ Topic=Public $\wedge$ Topic=Law $\Longrightarrow \theta \leq 0.5$	1	0.0023
AR13	Topic=Political Parties $\wedge$ Topic=European Union $\wedge$ Topic=Elections $\Longrightarrow \theta \leq 0.5$	1	0.0023
AR14	Politician=Milan Marković $\wedge$ Topic=Public $\Longrightarrow \theta \leq 0.5$	1	0.0023
AR15	Topic=Law $\wedge$ Topic=Elections $\Longrightarrow \theta \leq 0.5$	1	0.0023
AR16	Politician=Milutin Mrkonjić $\wedge$ Topic=Education $\Longrightarrow \theta \leq 0.5$	1	0.0023
AR17	Politician=Petar Škundrić $\wedge$ Topic=Money $\Longrightarrow \theta \leq 0.5$	1	0.0020
AR18	Politician=Mlađan Dinkić $\wedge$ Topic=Political Parties $\wedge$ Topic=Elections $\Longrightarrow \theta \leq 0.5$	1	0.0020
AR19	Topic=Police $\wedge$ Topic=Political Parties $\wedge$ Topic=Elections $\Longrightarrow \theta \leq 0.5$	1	0.0020
AR20	Topic=Energy $\wedge$ Topic=Political Parties $\wedge$ Topic=Public $\wedge$ Topic=Law $\Longrightarrow \theta \leq 0.5$	1	0.0020

Table 5 Promise fulfilment prediction on Istinomer promise data

(a) Statement texts

Perf.	Not true!	$\theta(Pol.)$	$AVG_m(\theta)$	kNN	HIKNN	NHBNN	NB	NB(kNN)	NB(HIKNN)
F_1^M	0.27	0.32	0.38	0.38	**0.39**	0.33	0.32	0.31	0.38
Acc.	**66.6 %**	59.6 %	57.4 %	62.1 %	61.1 %	56.3 %	50.7 %	62.2 %	61.5 %
Err_θ	0.31	0.33	**0.30**	0.33	0.33	0.37	0.36	0.31	0.33

(b) Comment texts

Perf.	Not true!	$\theta(Pol.)$	$AVG_m(\theta)$	kNN	HIKNN	NHBNN	NB	NB(kNN)	NB(HIKNN)
F_1^M	0.27	0.32	0.38	0.38	**0.39**	0.34	0.32	0.31	0.36
Acc.	**66.6 %**	59.6 %	57.4 %	58.3 %	58.5 %	56.3 %	50.9 %	64.9 %	56.9 %
Err_θ	0.31	0.33	**0.30**	0.36	0.38	0.35	0.35	0.30	0.36

Average estimator performance is shown, after examining all the promises in chronological order. Best result in each line is given in bold

8.2 Evaluating different promise fulfilment prediction strategies

We have implemented and examined several different approaches for promise fulfilment prediction based on n-gram textual representations and extracted topic models. These approaches were also compared to some simple baselines. The promises were evaluated in the chronological order and the predictive models were built incrementally as new promises were made and evaluated.

Zero-rule ("Not true!") was taken as the first baseline, due to a high percentage of unfulfilled promises. For the second baseline, we have considered the average truthfulness of the politician, up until that time ($\theta(Pol.)$). The third baseline was the moving average of truthfulness within a short time-window m ($AVG_m(\theta)$), where we have used $m = 5$.

We have evaluated the performance of kNN, HIKNN (Tomašev and Mladenić 2012), NHBNN (Tomašev et al. 2011) and Naive Bayes (NB) (Witten and Frank 2005). Furthermore, we have investigated a hybrid approach of using either kNN or HIKNN to determine the priors in the Naive Bayes rule. We will refer to these variants as NB(kNN) and NB(HIKNN). Only the positive topic matches were used as features in both NB(kNN) and NB(HIKNN). Using the local neighborhood information in Naive Bayes is quite common and we wanted to see if the n-gram and topic representations could be used together for better prediction performance. In the k-nearest neighbor approaches, the neighborhood size of $k = 1$ was used, as it gave the best results for this particular data in our initial experiments.

Three different classes were defined, based on the truthfulness scores. Promises with truthfulness $\theta < 0.375$ were considered 'probably false'. Promises with truthfulness $0.375 \leq \theta < 0.625$ were considered 'undetermined'. Finally, promises with truthfulness $0.625 \leq \theta$ were considered 'probably truthful'.

Even though the labels have been assigned to certain truthfulness ranges, the prediction models themselves were trained on the fuzzy truthfulness scores θ and have been set to return a fulfillment prediction value between 0 and 1. Binning was performed afterwards and had no impact on model training.

Due to a high class imbalance of the Istinomer promise data, which is highly skewed towards the negative examples, the macro-averaged F_1 score (F_1^M) was taken as a primary evaluation metric (Witten and Frank 2005). It is frequently used for class imbalanced classification evaluation. The global accuracy was also calculated, as well as the error in truthfulness probability itself.

Table 5 gives a summary of average estimator performances. Clearly, trying to predict promise truthfulness based on short statement texts alone is not an easy task. The highest average accuracy is actually achieved by zero-rule, which assigns a truthfulness of $\theta = 0$ to every statement it encounters. The moving average of truthfulness achieves the lowest error in probability, though it still remains high.

The content-based approaches achieve a significantly higher F_1^M score than zero-rule or $\theta(Pol.)$. This shows that

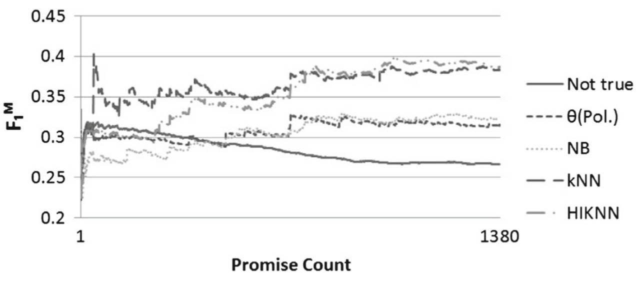


Fig. 16 The evolution of F_1^M score in truthfulness prediction, over time

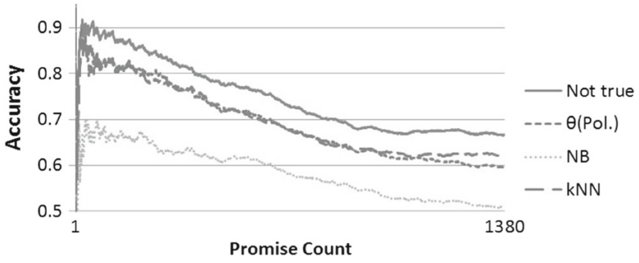


Fig. 17 The evolution of truthfulness prediction accuracy, over time

it is possible to utilize at least some of the textual information correlated with promise fulfillment. Also, disregarding the zero-rule, the content-based methods achieve a significantly higher accuracy than the other two baselines. This is important, as the zero-rule does not in fact attempt to make meaningful truthfulness predictions. Saying that everything is false might be fairly accurate on such imbalanced data, but it is not a potentially useful real-life estimation strategy (Table 5).

Estimating promise truthfulness from the detailed comments seems not to be better than basing the decisions on the short statements. It also seems that the augmented automatically extracted topic vectors do not encode enough truthfulness-related information. Naive Bayes performs very badly on this data. However, the topic vectors themselves are still useful in other contexts, as we have already seen that they can be used to capture semantically correct similarities between different politicians (Section 7).

The hybrid approaches were not much better than the rest, though they did clearly outperform the basic Naive Bayes, suggesting that correctly setting the priors is quite important.

HIKNN, a hubness aware *k*NN method, achieved the highest F_1^M score, though it was not significantly better than *k*NN. As *k*NN is easier to interpret and implement, we will focus on *k*NN as the primary estimator in the remaining discussion.

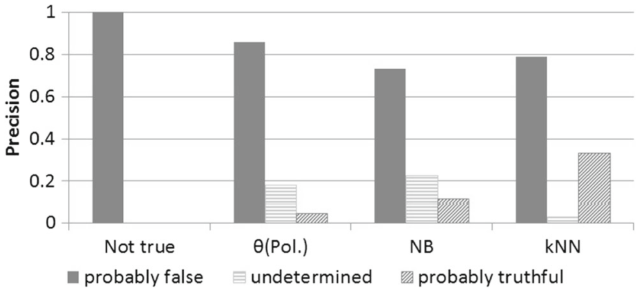


Fig. 18 Precision of different truthfulness estimation methods on different types of promises. Truthful and partially truthful promises seem to be much more difficult to recognize than those that are mostly or entirely false and end up not being fulfilled

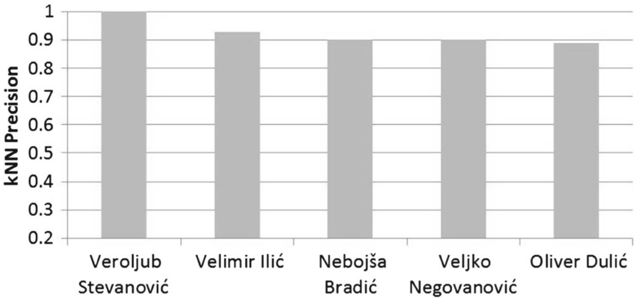


Fig. 19 The 5 most predictable major promise makers, based on *k*NN truthfulness estimation

Figures 16 and 17 show that the average performance summaries do in fact reflect a difference between the compared approaches, as their average performance metrics remain clearly separated over time, as promises are introduced and evaluated.

In Fig. 17, it can be seen that the accuracy of all approaches is decreasing along with the increasing truthfulness of promises (this easily follows from the drop in zero-rule performance). This indicates that it might be more difficult to detect the positive examples, promises that are probably going to be fulfilled. Therefore, most errors are pessimistic, in a sense that the truthful promises are being estimated as non-truthful. This has been confirmed by examining the confusion matrices. Figure 18 shows different truthfulness class precisions for different estimators on Istinomer promise data. The truthfulness skewness of the dataset certainly plays a role in the pessimistic bias of the models, but other factors might be involved as well.

A detailed analysis of the results has revealed that the prediction precision varies for different politicians and different topics, as some tend to be easier and some more difficult to predict accurately.

On average, better prediction precision is achieved for the politicians that fulfil fewer promises, as prediction quality is better for the less likely promises. A list of 5 most predictable promise makers and 5 least predictable promise

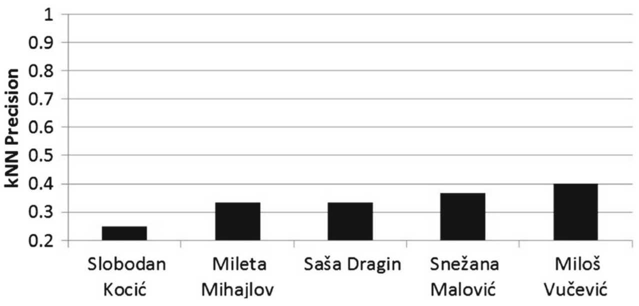


Fig. 20 The 5 least predictable major promise makers, based on *k*NN truthfulness estimation

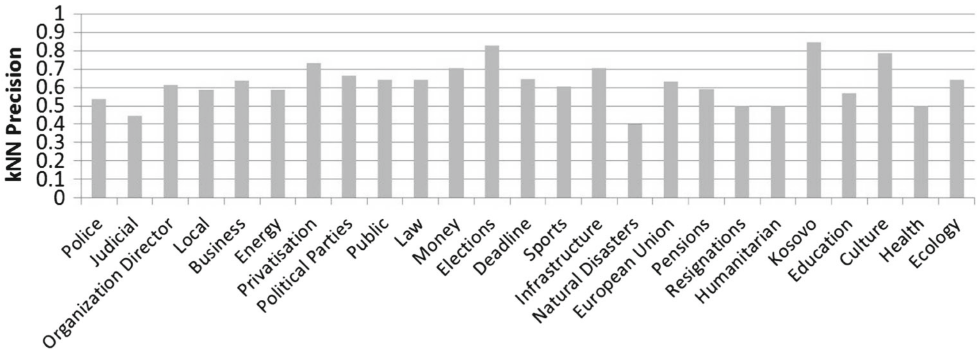


Fig. 21 The precision of *k*NN truthfulness estimation for different automatically extracted topics on Istinomer promise data

makers (with at least 5 promises in the database) is shown in Figs. 19 and 20, respectively.

The least predictable major promise maker, based on *k*NN, is Slobodan Kocić. His Istinomer promise profile reads as follows (according to the considered promises by July 2013): 3 unfulfilled promises, 2 projects that have not been started, still working on 2 projects and 1 promise that has been fully fulfilled. The most predictable major promise maker, based on *k*NN, is Veroljub Stevanović. Out of his promises, 3 have not been fulfilled and 3 have been started and aborted. In other words, all 6 of them belonged to the merged 'probably false' truthfulness category that was used in precision/accuracy calculations.

There results suggest that it is still possible to achieve a high truthfulness prediction accuracy for certain promise makers. It might be possible to develop separate content-based truthfulness models for different politicians instead, in order to improve the prediction performance.

Similarly, it seems that not all topics are equally well suited for the similarity-based *k*NN approach. This can be seen in Fig. 21. The *k*-NN truthfulness estimator was best at estimating the truthfulness of promises pertaining to Kosovo, Culture, Elections and Privatisation.

9 Conclusions

We have implemented a pipeline for analyzing Serbian political discourse by applying standard data mining and visualization techniques to an information system of annotated political promises publicly available on the Serbian Truth-o-meter (Istinomer) website.

The textual data was represented via n-grams and was semantically enriched by a set of 25 topics that were semi-automatically extracted from the text.

Topical promise profiles were generated for the politicians present in the data and were used alongside n-gram representations for cross-politician comparisons. The comparisons have revealed that politicians in similar roles tend to make similar promises. This shows that the chosen data representation is able to capture the hidden semantics in the data and is appropriate for conducting further analysis.

In order to better summarize the topics, we have proposed to use statement hubs as promise prototypes. This possibility is backed up by recent discoveries in high-dimensional data clustering and analysis. This might help with gaining quick insights into the topical statement structure.

Statement topic vectors were used for discovering truthfulness patterns in Istinomer data via association rules. The analysis has revealed some interesting correlations. Most of the strong rules predict negative outcomes, promises not being fulfilled.

Several truthfulness estimators were implemented and estimated incrementally on chronologically ordered statements. The content-based methods outperformed most simple baselines in terms of the F_1^M score, but had a lower accuracy compared to a baseline that views all promises as 'probably false', which is a consequence of severe class imbalance in the data. However, the results suggest that some topics and politicians are easier for prediction than others, so it should be possible to make future improvements, by using feature selection and metric learning.

The proposed approach is language-independent and can easily be used for analyzing political statements from other countries, in cases where the appropriate annotated databases of factual political statements exist and are publicly available.

10 Limitations of the proposed approach and future work

While the proposed method enables us to gain some useful insights, there is plenty of room for improvement.

The truthfulness labels assigned by the annotators are highly subjective and it is unclear whether the fuzzy interpretation presented in the paper can fully capture the underlying semantics. This is a limitation in data collection.

Ideally, the truthfulness grades would be assigned collaboratively by integrating opinions from multiple experts, thereby reducing the bias in the data and assigning each promise a singular quantitative score that is the overall truthfulness estimate.

The data that was used focuses on the most prominent politicians and their statements. It does not provide full coverage of the political discourse. It may be beneficial to attempt to extract the promises and statements automatically from large news corpora instead, in order to achieve better coverage. Such extracted statements might then be pruned and evaluated by human annotators.

Finally, the textual analysis could employ some recently developed language-independent techniques for representation learning (Mikolov et al. 2013), as well as entity extraction. Topic extraction can also be more automated, which would make it easier to build the final enriched statement representation. More mature techniques might allow for the development of interactive software that would enable people to gain better understanding of the political process.

References

Adamic, L., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. In *In LinkKDD'05: Proceedings of the 3rd international workshop on Link discovery* (pp. 36–43).

Adamo, J. (2001). *Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms*. Berlin: Springer.

Agirre, E., Martínez, D., de Lacalle, O.L., & Soroa, A. (2006). Two graph-based algorithms for state-of-the-art WSD. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 585–593).

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *SIGMOD Rec*, *22*(2), 207–216.

AlSumait, L., Barbara, D., & Domeniconi, C. (2008). On-line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking. In *Eighth IEEE International Conference on Data Mining (ICDM)* (pp. 3–12).

Baccianella, A.E.S., Sebastiani, F., & Sentiwordnet 3.0 (2010). An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta.*

Balasubramanyan, R., Routledge, B.R., & Smith, N.A. (2010). From tweets to polls : Linking text sentiment to public opinion time series.

Cagliero, L., & Fiori, A. (2013). Discovering generalized association rules from twitter. *Intelligent Data Analysis*, *17*(4), 627–648.

Campbell, J.E. (2008). Evaluating u.s. presidential election forecasts and forecasting equations. *Int. J. Forecast.*, *24*(2), 259–271.

Carruba, C., Gabel, M., Murrah, L., Clough, R., Montgomery, E., & Schambach, R. (2006). Off the Record: Unrecorded Legislative Votes, Selection Bias and Roll-Call Vote Analysis. *Br. J. Polit. Sci.*, *36*(4), 691–704.

Cate, F.H., Dempsey, J.X., & Rubinstein, I.S. (2012). Systematic government access to private-sector data. *International Data Privacy Law*, *2*(4), 195–199. doi:10.1093/idpl/ips027.

Cavnar, W.B., & Trenkle, J.M. (1994). N-gram-based text categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval* (pp. 161–175).

Charalabidis, Y., & Koussouris, S. (Eds.) (2012). *Empowering Open and Collaborative Governance - Technologies and Methods for Online Citizen Engagement in Public Policy Making*: Springer.

Charalabidis, Y., Triantafillou, A., Karkaletsis, V., & Loukis, E. (2012). *Public policy formulation through non moderated crowdsourcing in social media*, (pp. 156–169): Springer.

Cliffe, L., Ramsay, M., & Bartlett, D. (2000). *The politics of lying: Implications for democracy*: St Martin's Press.

Clinton, J., Jackman, S., & Douglas, R. (2004). The Statistical Analysis of Roll Call Data. *Am. Polit. Sci. Rev.*, *2*, 355–370.

Custers, H., Calders, T., & Zarsky, T. (2013). *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases. Studies in applied philosophy, epistemology and rational ethics*: Springer.

Dai, H.J., Chang, Y.C., Tzong-Han Tsai, R., & Hsu, W.L. (2010). New challenges for biological text-mining in the next decade. *J. Comput. Sci. Technol.*, *25*(1), 169–179.

Damashek, M. (1995). Gauging similarity with n-grams: Language-independent categorization of text. *Science*, *267*(5199), 843–849.

Danna, A. (2002). Gandy OscarH., J.: All that glitters is not gold: Digging beneath the surface of data mining. *J. Bus. Ethics*, *40*(4), 373–386.

Dörre, J., Gerstl, P., & Seiffert, R. (1999). Text mining: finding nuggets in mountains of textual data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '99* (pp. 398–401). New York: ACM. doi:10.1145/312129.312299.

Fairclough, I., & Fairclough, N. (2013). *Political Discourse Analysis: A Method for Advanced Students*: Taylor & Francis.

Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*: Cambridge University Press.

François, D., Wertz, V., & Verleysen, M. (2007). The concentration of fractional distances. *IEEE Transactions on Knowledge and Data Engineering*, *19*(7), 873–886.

Gamon, M., Basu, S., Belenko, D., Fisher, D., Hurst, M., & Konig, A.C. (2008). BLEWS: Using Blogs to Provide Context for News Articles. In *ICWSM, 2008*.

Greenberg, J. (2010). There's nothing anyone can do about it: Participation, apathy, and "successful" democratic transition in postsocialist serbia. *Slav. Rev.*, *69*(1), 41–64.

Grosskreutz, H., Boley, M., & Krause-Traudes, M. (2010). Subgroup discovery for election analysis: A case study in descriptive data mining. In *Discovery Science* (pp. 57–71). Berlin Heidelberg: Springer.

Hamamoto, M., Kitagawa, H., Pan, J.Y., & Faloutsos, C. (2005). A comparative study of feature vector-based topic detection schemes a comparative study of feature vector-based topic detection schemes. In *Web Information Retrieval and Integration, 2005. WIRI '05. Proceedings. International Workshop on Challenges in* (pp. 122–127).

He, X., & Zhang, J. (2006). Why Do Hubs Tend to Be Essential in Protein Networks *PLoS Genet.*, *2*(6).

Helbing, D., & Balietti, S. (2011). From social data mining to forecasting socio-economic crises. *The European Physical Journal Special Topics*, *195*(1), 3–68.

Hong, T.P., Kuo, C.S., & Chi, S.C. (1999). Mining association rules from quantitative data. *Intelligent Data Analysis*, *3*(5), 363–376.

Howard, P.N. (2005). Deep democracy, thin citizenship: The impact of digital media in political campaign strategy. *The ANNALS of the American Academy of Political and Social Science*, *597*(1), 153–170. doi:10.1177/0002716204270139.

Jackman, S. (2001). Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking. *Polit. Anal.*, *9*(3), 227–241.

Jackson, P., & Moulinier, I. (2007). *Natural Language Processing for Online Applications: Text retrieval, extraction and categorization. Second revised edition. Natural Language Processing*: John Benjamins Publishing Company.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Inf. Syst. Manag.*, *29*(4), 258–268.

Kes̆elj, V., Peng, F., Cercone, N., & Thomas, C. (2003). N-gram-based author profiles for authorship attribution. In *Proceedings of the conference pacific association for computational linguistics, PACLING*, (Vol. 3 pp. 255–264).

Klein, D., Smarr, J., Nguyen, H., & Manning, C.D. (2003). Named entity recognition with character-level models. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL, CONLL '03., Association for Computational Linguistics* (pp. 18–183). USA: Stroudsburg. doi:10.3115/1119176.1119204.

Liu, B. (2007). Opinion mining. In *Web Data Mining, Data-Centric Systems and Applications* (pp. 411–447). Berlin Heidelberg: Springer.

Loukis, E., & Charalabidis, Y. (2012). Participative public policy making through multiple social media platforms utilization. *Int. J. Electron. Gov. Res.*, *8*(3), 78–97. doi:10.4018/jegr.2012070105.

Malouf, R., & Mullen, T. (2008). Taking sides: user classification for informal online political discourse. *Internet Research*, *18*(2), 177–190.

Maragoudakis, M., Loukis, E., & Charalabidis, Y. (2011). A review of opinion mining methods for analyzing citizensâĂŹ contributions in public policy debate. In *Electronic Participation* (pp. 298–313). Berlin Heidelberg: Springer.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In *ICLR Workshop*.

Milošević, N. (2012). *Stemmer for Serbian language*: ArXiv e-prints.

Miner, G., Elder, J., Hill, T., Delen, D., & Fast, A. (2012). *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications. Academic Press*: Academic Press.

Mostafa, M.M., & El-Masry, A.A. (2013). Citizens as consumers: Profiling e-government servicesâĂŹ users in egypt via data mining techniques. *Int. J. Inf. Manag.*, *33*(4), 627–641. doi:10.1016/j.ijinfomgt.2013.03.007.

Murray, G.R., Riley, C., & Scime, A. (2009). Pre-election polling: Identifying likely voters using iterative expert data mining. *Public Opinion Quarterly*, *73*(1), 159–171. doi:10.1093/poq/nfp004.

Murray, G.R., & Scime, A. (2010). Microtargeting and electorate segmentation: Data mining the american national election studies. *Journal of Political Marketing*, *9*(3), 143–166. doi:10.1080/15377857.2010.497732.

Nanopoulos, A., Radovanović, M., & Ivanović, M. (2009). How does high dimensionality affect collaborative filtering? In *Proceedings of the third ACM conference on Recommender systems, RecSys '09* (pp. 293–296). USA: ACM.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, *2*(1-2), 1–135. doi:10.1561/1500000011.

Piatetsky-Shapiro, G. (1991). Discovery, analysis and presentation of strong rules. In *Knowledge Discovery in Databases* (pp. 229–248): AAAI Press.

PÃtry, F., Collette. (2009). In L.M. Imbeau (Ed.) *Measuring how political parties keep their promises: A positive perspective from political science* (Vol. 15, pp. 65–80). New York: Springer.

Raghavan, V.V., & Wong, S.K.M. (1986). A critical analysis of vector space model for information retrieval. *J. Am. Soc. Inf. Sci.*, *37*(5), 79–287. doi:10.1002/(SICI)1097-4571(198609)37:5<279::AID-ASI1>3.0.CO;2-Q.

Rana, N., Dwivedi, Y., & Williams, M. (2013). A meta-analysis of existing research on citizen adoption of e-government. *Inf. Syst. Front.*, 1–17.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Detecting and tracking political abuse in social media. In *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Sanches, P., Svee, E.O., Bylund, M., Hirsch, B., & Boman, M. (2013). *Knowing your population: Privacy-sensitive mining of massive data* Vol. 1: Network and Communication Technologies.

Scharl, A., & Weichselbraun, A. (2008). An automated approach to investigating the online media coverage of U.S. presidential elections. *Journal of Information Technology and Politics*, *5*(1), 121–132. doi:10.1080/19331680802149582.

Seo, Y.W., & Sycara, K. (2004). *Text clustering for topic detection. Tech. Rep. CMU-RI-TR-04-03*. Pittsburgh: Robotics Institute.

Stamatatos, E. (2009). Intrinsic plagiarism detection using character n-gram profiles. In *3rd PAN Workshop. Uncovering Plagiarism, Authorship and Social Software Misuse* (pp. 38–46).

Stieglitz, S., & Dang-Xuan, L. (2012). Social media and political communication: a social media analytics framework. *Soc. Netw. Anal. Min.*, 1–15.

Tomašev, N., & Mladenić, D. (2012). Nearest neighbor voting in high dimensional data: Learning from past occurrences. *Computer Science and Information Systems*, *9*, 691–712.

Tomašev, N., Radovanović, M., Mladenić, D., & Ivanović, M. (2013). The role of hubness in clustering high-dimensional data. *IEEE Trans. Knowl. Data Eng.*, *99*(PrePrints), 1.

Tomašev, N., Radovanović, M., Mladenić, D., & Ivanovicć, M. (2011). A probabilistic approach to nearest neighbor classification: Naive hubness bayesian k-nearest neighbor. In *Proceeding of the CIKM conference*.

Uramoto, N., Matsuzawa, H., Nagano, T., Murakami, A., Takeuchi, H., & Takeda, K. (2004). A text-mining system for knowledge discovery from biomedical documents. *IBM Syst. J.*, *43*(3), 516–533.

Vachudova, M.A. (2009). Corruption and compliance in the EU's post-communist members and candidates. *JCMS: Journal of Common Market Studies*, *47*, 43–62.

Vaidya, J. (2012). Privacy in the context of digital government. In *Proceedings of the 13th Annual International Conference on Digital Government Research, dg.o '12* (pp. 302–303). New York: ACM. doi:10.1145/2307729.2307796.

Vitas, D., Krstev, C., Obradović, I., Popović, L., & Pavlović-Lazetić, G. (2003). *An overview of resources and basic tools for processing of Serbian written texts*.

Vlado, K., & Šipka, D. (2008). A suffix subsumption-based approach to building stemmers and lemmatizers for highly inflectional languages with sparse resources. INFOTHECA. *Can. J. Inf. Libr. Sci.*, *9*(1), 23–33.

Wartena, C., & Brussee, R. (2008). Topic detection by clustering keywords. In *19th International Workshop on Database and Expert Systems Application, 2008. DEXA '08* (pp. 54–58).

Weber, I., Garimella, V.R.K., & Borra, E. (2012). Political search trends. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, SIGIR '12* (pp. 1012–1012). New York: ACM. doi:10.1145/2348283.2348437.

Weerakkody, V., Irani, Z., Lee, H., Osman, I., & Hindi, N. (2013). E-government implementation: A birdâĂŹs eye view of issues relating to costs, opportunities, benefits and risks. *Inf. Syst. Front.*, 1–27.

Witten, I.H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. USA: Morgan Kaufmann Publishers Inc.

Zhong, N., Li, Y., & Wu, S.T. (2012). Effective pattern discovery for text mining. Knowledge and Data Engineering. *IEEE Transactions on*, *24*(1), 30–44.

Nenad Tomašev is a software engineer. He received his PhD in machine learning from the Artificial Intelligence Laboratory at Jožef Stefan Institute in Ljubljana, Slovenia in 2013. The focus of his research was on machine learning in many dimensions and the phenomenon of hubness in particular. In his research, Dr. Tomašev designed novel algorithms for classification, clustering, metric learning, representation learning, ranking and instance selection. Dr. Tomašev previously interned at Google and Telvent DMS. He also participated in organizing summer schools in Višnjan and Petnica Science Center.