

Refining Network Information on the Example of Parliamentary Elections. Part 2. Internet Portals, Media and Content Contexts

KEY WORDS

Internet, Big Data, internet portals, refining Big Data, new sources of journalistic information, parliamentary elections 2011

ABSTRACT

An important part of information resources available online as part of Big Data are Internet portals created by professionals. Frequently, though not always, they are an expanded, updated version of the printed editions of dailies and magazines. An important feature influencing the topics, range and quality of content created on these sites is their creation by professional editorial offices. Refining these resources – as new sources of journalistic information – is an addition to the first part of the presented research results, which concerned information sources provided by refining social networks, where content is usually created by non-professionals.

An important part of information resources available online, comprising peculiar data warehouses known as Big Data, are internet portals created by professionals. They often are expanded, updated versions of the printed editions of dailies and magazines. An important feature influencing the topics, range and quality of content created on these sites is their creation by professional editorial offices. An analysis – refining – of these resources was undertaken at the Institute of Journalism, University of Warsaw as part of research aiming to indicate and verify the information processing tools to allow the assessment of the current voting preferences before the parliamentary elections in Poland in 2011. A fundament to achieve the objective pursued was the evaluation of quantitative and qualitative data, as well as the dynamics of the content appearing in the social media (the results have been described in the first part of this article¹) and in the online editions of some newspapers. The results of the refining of data from online newspapers are presented in this article.

The aim of the second part of the research was, besides confirming the correctness of adopted methodology, to indicate another authoritative source of journalistic information. It was based on the example of collecting data in order to diagnose the state and dynamics of the

¹ W. Gogolek, P. Kuczma, “Rafinacja informacji sieciowych na przykładzie wyborów parlamentarnych. Część 1. Blogi, fora, analiza sentymentów” [Refining Network Information on the Example of Parliamentary Elections. Part 1. Blogs, Forums, Sentiment Analysis], *Studia Medioznawcze* 2013, No 2, p. 89–109.

change of the information view of the election committees (political parties) participating in the parliamentary elections in Poland in 2011. The following hypothesis was proposed: refining the content of internet portals enables an ongoing and reliable monitoring of the variables describing the voting preferences of Poles in the period preceding the parliamentary elections of 2011. The analysed data from internet portals come from the period 1 August–9 October 2011².

One of the first identified measures of the parties participating in the parliamentary elections was the number of entries containing their names present on internet portals. Chart 1 shows, on one hand, the relatively low informative value of the data concerning the final results of the elections, on the other hand, the abrupt rise in the number of publications in the first days of October. The predominance of the information about the ruling party seems natural and, in the face of the small number of publications about the second largest party, disproportionate in relation to the results of the elections, does not contribute significant information.

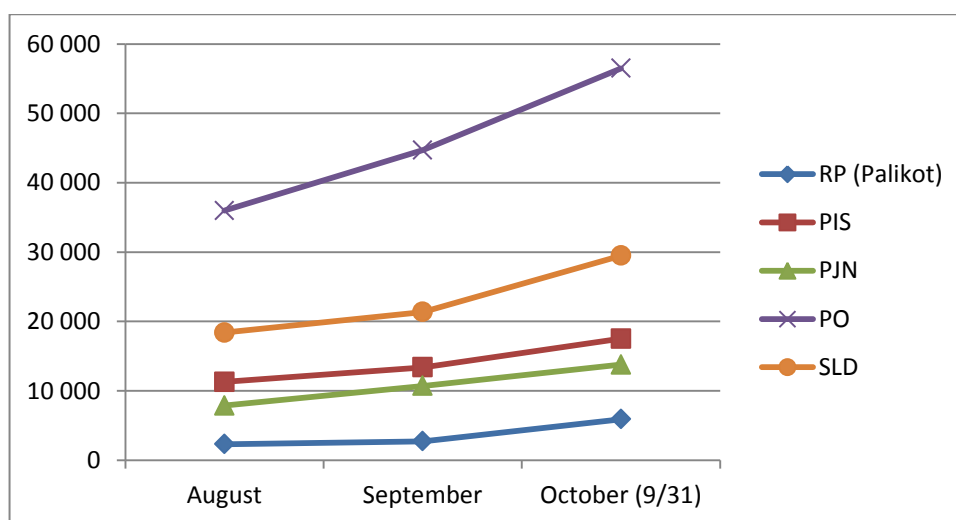


Chart 1. Changes in the representation of the names of political parties on the internet portals
Source: Tables and charts developed by the authors based on the conducted research.

In order to perform a quantitative assessment of the number of times the names of political parties occurred in the online texts, appropriate contexts were identified. The keys (words/expressions) associated with the government and its functions, and with the

² The October data were normalised to highlight the dynamics of content changes from month to month. In order to do so, the number of entries published in the first 9 days was divided by 9 to obtain the daily mean number of entries, and subsequently multiplied by 31, the number of days in October.

competencies of individual ministries³, adopted as substantive contexts: 1) education, 2) finance, 3) economy, 4) infrastructure, 5) culture, 6) science and higher education, 7) defence, 8) labour and social policy, 9) agriculture, 10) regional development, 11) state treasury, 12) sport and tourism, 13) justice, 14) domestic affairs and administration, 15) foreign affairs, 16) environment, and 17) health. The words describing the competencies of each ministry were based on the competencies specified in their statutes.⁴

The second group of contexts – media-related – consists of the contexts associated with the ongoing events being covered by the media. The two largest opinion dailies in Poland⁵, of differing political profiles, were subjected to refining: *Gazeta Wyborcza* and *Rzeczpospolita*. Electronic editions of both daily newspapers, available via the Festiva search engine, were used in the analysis⁶. The analysis enabled to distinguish the following groups of media-related contexts: 1) EU (European Union) – including, among others, such words as union, EU, European, presidency, Europe; 2) (Smolensk) air crash – catastrophe, Smoleńsk, Russia, IAC, tragedy, Tupolev, assassination; 3) power – government, administration, power, Sejm, leader, president; 4) media – media, newspaper, TVP, TVN, television; 5) money – money, finance, budget, NBP; 6) reforms – reforms; 7) Church – Church; 8) law – prosecutor's office, law, statute, court, tribunal, etc.

Media-related context obtained this way were used for the analysis of the substantive character of the electoral campaign of 2011. The analysis attempted, among other goals, to answer the question whether the content published by internet portals represented more substantive contexts, or media-related contexts, taken quantitatively. Contexts are understood here as strings containing in close proximity the names of parties and words adopted as substantive and media-related contexts. These strings are further referred to as entries, like in the refining of social media.

The methodology of the research, including the study of sentiments, was described in the first part of the article.

³ Based on the structure of the Council of Ministers, see: *Postanowienie Prezydenta Rzeczypospolitej Polskiej z dnia 16 listopada 2007 r. nr 1131-50-07 o powołaniu w skład Rady Ministrów* [Resolution of the President of the Republic of Poland on 16 November 2007, No 1131-50-07, on the appointment to the Council of Ministers], *Monitor Polski (M.P.)* [Polish Monitor] 2007, No 87, item 947.

⁴ For the list of statutes, see: http://www.id.uw.edu.pl/zasoby/profile/59/Aneks_nr_2-Wykaz_statutow_ministerstw.pdf [accessed: 23 Apr 2013].

⁵ See: <http://www.wirtualnemedial.pl/artykul/gazeta-wyborcza-i-fakt-to-najchetniej-czytane-dzienniki#> [accessed: May 2010].

⁶ https://han.buw.uw.edu.pl/han/ISIEM/site.securities.com/search/pub_search.html?pc=PL&sv=EMIS [accessed: May 2010].

The study of the content of selected internet portals showed a relatively small prevalence of substantive contexts over media-related ones.

Adopting refining as the source of journalistic information suggested in the description of the research, the distribution of the media-related and substantive contexts in the news about all the parties participating in the elections may be considered interesting.

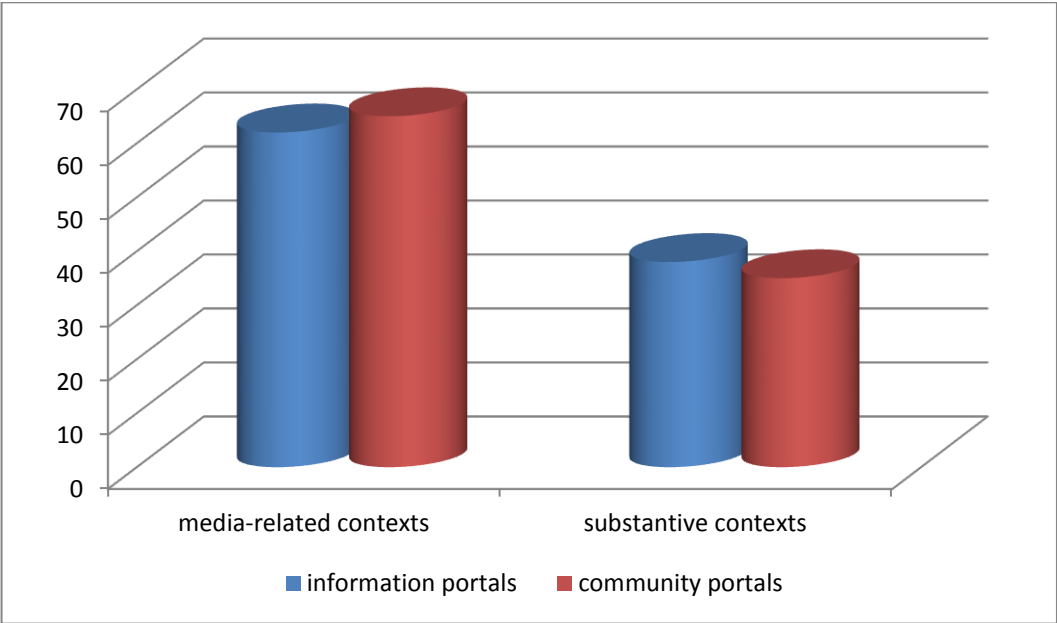


Chart 2. Distribution of contexts: media-related vs. substantive contexts in the information portals and social media (percentages)

The distribution of the substantive and media-related contexts in the information portals is very similar to the corresponding distribution in the social media. This is indicative of a great similarity of the entries in the professional media reports and community sources. Both the social media and information portals seem to pay almost as much attention to the substantive (with a negligible advantage of information portals) and media-related questions. This conclusion draws attention to the debatable issue of the differences, or rather the lack of them, in the informational content of these sources of information.

Table 1. Detailed distribution of contexts in the information portals

| Contexts in the information portals | Percentage distribution |
|-------------------------------------|-------------------------|
| Law | 14 |
| Media | 10 |
| Labour and social policy | 9 |
| Infrastructure | 8 |

| | |
|-------------------------------------|---|
| Justice | 6 |
| Foreign affairs | 6 |
| Finance | 6 |
| Education | 4 |
| Economy | 4 |
| Health | 4 |
| Air crash | 3 |
| Money | 3 |
| Sport and tourism | 3 |
| European Union | 3 |
| Culture | 2 |
| Science and higher education | 2 |
| Agriculture | 2 |
| Internal affairs and administration | 2 |
| Environment | 2 |
| Power | 2 |
| Church | 1 |
| Reforms | 1 |
| State treasury | 1 |
| Defence | 0 |
| Regional development | 0 |

What is quite telling are the differences in correlation between the number of votes gained by the parties and 1) the positive entries in social media, blogs, and forums; and 2) positive entries in the analysed portals. It attests to a much more apparent sentiment (tonality) in community networks towards the winning parties than in the more subdued content of the portals.

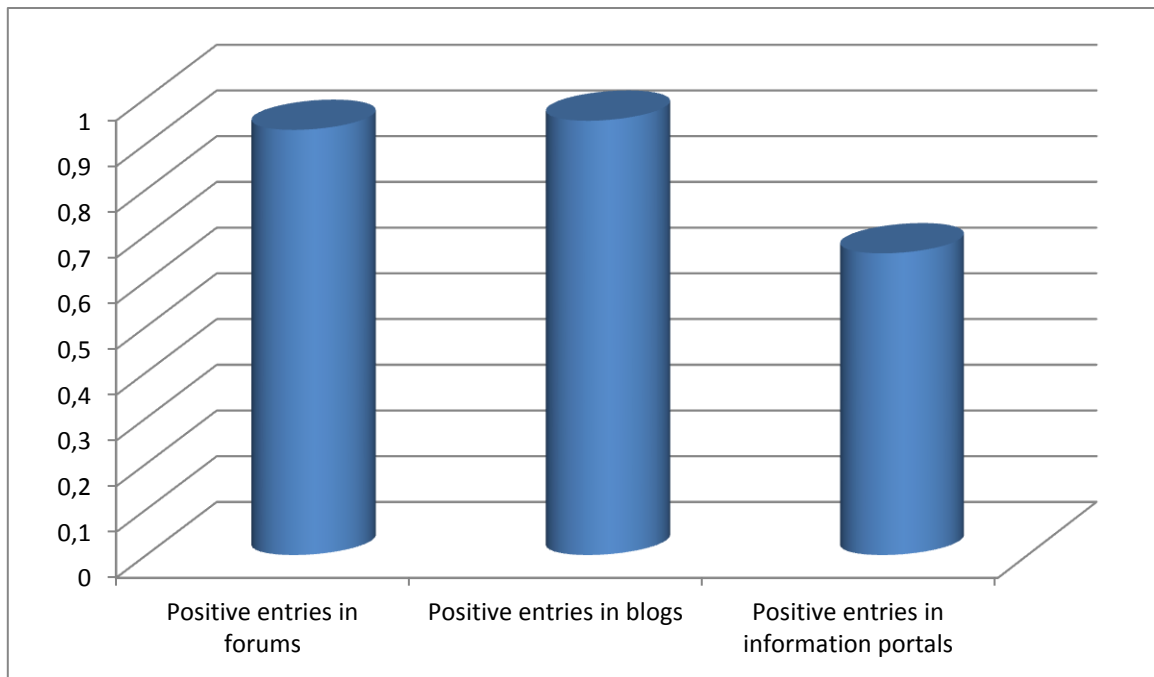


Chart 3. Proportions of the values of Pearson's correlation factors between the number of votes received by the parties and the number of positive entries (sentiments)

Conclusions

The presented research results obtained via the analysis of data collected from the entries in internet portals show an almost 40-percent share of media-related issues (contexts). It attests to the nature of the interests of the publishers of the portals, not necessarily dominated by the substantive problems, as could be expected. On one hand, this conclusion seems to be a valuable piece of information, a clue to the proportion of the content published in the portals and provided by professional editorial offices. On the other, it points out the directions in the choice of content which should be kept or changed by the most interested parties – the journalists.

A no less important conclusion is the indication of the surprisingly large informational value of the data collected from social media. It proves that the results of refining community networks (described in the first part of the article) is closer to the actual results of the elections than the result of the analysis of content published in the information portals during the electoral campaign.

The method of tapping the potential of the Big Data through refining as a journalistic tool, suggested in the description of the study, seems entirely justified as a new long-term source of journalistic information. It is of particular significance in the period of changes taking place in the traditional media, doomed to convergence with the new media and their related tools. It becomes necessary to notice the usefulness of data-driven decision making.

This phenomenon may be applied in the public sphere, e.g. during the preparation of electoral campaigns, as well as in the decision-making of the politicians in the situations between elections, and in the commercial sphere decision-making, where the data obtained through refining may prove to be an important source of data on the consumer preferences and behaviour.