

# Rafinacja informacji sieciowych na przykładzie wyborów parlamentarnych.

## Część 2. Portale internetowe, konteksty medialne i merytoryczne

**Włodzimierz Gogołek, Paweł Kuczma**

Istotną częścią zasobów informacyjnych dostępnych w sieci, składających się na swoiste hurtownie danych zwane Big Data, są tworzone przez profesjonalistów, zawodowo, portale internetowe. Często są one rozszerzoną, aktualizowaną kopią wydań drukowanych dzienników i czasopism. Istotnym czynnikiem wpływającym na tematykę, zakres i jakość powstających w tych miejscach treści jest fakt ich przygotowywania przez profesjonalne redakcje. Analizę – rafinację – tych zasobów podjęto w Instytucie Dziennikarstwa Uniwersytetu Warszawskiego w ramach jednej z prac badawczych mających na celu wskazanie i weryfikację narzędzi obróbki informacji umożliwiających ocenę bieżących preferencji wyborczych przed wyborami parlamentarnymi w Polsce w 2011 r. Podstawą do osiągnięcia założonego celu była ocena danych ilościowych i jakościowych oraz dynamiki zmian treści ukazujących się w mediach społecznościowych (wyniki przedstawiono w pierwszej części artykułu)<sup>1</sup> oraz w siecio-

wych wydaniach niektórych gazet. Wyniki rafinacji danych pochodzących z gazet online zostaną przedstawione w niniejszym artykule.

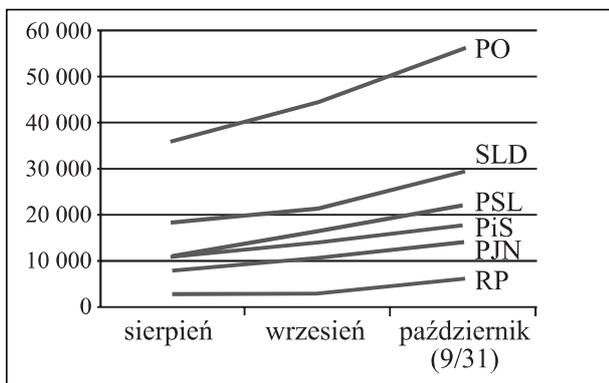
Celem drugiej części badań było, poza potwierdzeniem poprawności przyjętej metodologii badań, wskazanie kolejnego miarodajnego źródła informacji dziennikarskich. Posłużył temu przykład pozyskiwania danych służących diagnozowaniu stanu i dynamiki zmian obrazu informacyjnego aktywności komitetów wyborczych (partii politycznych) biorących udział w wyborach parlamentarnych w Polsce 2011 r. Przyjęto następującą hipotezę: rafinacja treści portali internetowych umożliwiała bieżący, wiarygodny monitoring zmiennych opisujących preferencje wyborcze Polaków w okresie poprzedzającym wybory parlamentarne w 2011 r. Analizowane dane z portali internetowych pochodzą z okresu 1 sierpnia – 9 października 2011 r.<sup>2</sup>

Jedną z pierwszych wyróżnionych miar oceny partii biorących udział w wyborach parlamentarnych była liczba wpisów z ich nazwa-

<sup>1</sup> W. Gogołek, P. Kuczma, *Rafinacja informacji sieciowych na przykładzie wyborów parlamentarnych. Część 1. Blogi, fora, analiza sentymentów*, „Studia Medioznawcze” 2013, nr 2, s. 89–109.

<sup>2</sup> Dane październikowe zostały znormalizowane, aby uwidocznić dynamikę zmian treści z miesiąca na miesiąc. W tym celu liczbę wpisów opublikowanych w pierwszych dziewięciu dniach podzielono przez 9, aby uzyskać średnią dzienną liczbę wpisów, i pomnożono przez 31 – liczbę dni w październiku.

mi występujących na portalach internetowych<sup>3</sup>. Rysunek 1 ilustruje, z jednej strony, względnie małą wartość informacyjną danych dotyczącą ostatecznych wyników wyborów, z drugiej gwałtowny wzrost liczby publikacji dotyczących partii w pierwszych dniach października. Dominacja informacji o partii rządzącej wydaje się naturalna i wobec nieproporcjonalnej w stosunku do wyników wyborów małej liczby publikacji o drugiej największej partii, nie wnosi znaczących informacji.



Rys. 1. Zmiany liczebności występowania nazw partii w portalach internetowych

Źródło: Tabele i wykresy w opracowaniu własnym na podstawie przeprowadzonych badań.

W celu dokonania ilościowej oceny krotności występowania nazw partii w tekstach zamieszczanych w sieci wyróżniono odpowiednie konteksty. Były nimi klucze (zwroty/słowa) wiązane z rządem i jego funkcjami oraz kompetencjami poszczególnych ministerstw<sup>4</sup>, przyjęte jako konteksty merytoryczne: 1) edukacja, 2) finanse, 3) gospodarka, 4) infrastruktura, 5) kultura, 6) nauka i szkolnictwo wyższe,

7) obrona, 8) praca i polityka społeczna, 9) rolnictwo, 10) rozwój regionalny, 11) skarb państwa, 12) sport i turystyka, 13) sprawiedliwość, 14) sprawy wewnętrzne i administracja, 15) sprawy zagraniczne, 16) środowisko oraz 17) zdrowie. Słowa opisujące kompetencje każdego z ministerstw zostały oparte na kompetencjach zapisanych w ich statutach<sup>5</sup>.

Drugą grupę kontekstów – konteksty medialne – stanowią te, które związane są z bieżącymi wydarzeniami relacjonowanymi w mediach. Rafinacji poddano dwa największe polskie dzienniki opiniotwórcze<sup>6</sup> o odmiennym profilu politycznym: „Gazetę Wyborczą” oraz „Rzeczpospolitą”. Do tej analizy wykorzystano elektroniczną wersję obu dzienników dostępnych za pomocą wyszukiwarki Factiva<sup>7</sup>. W wyniku analizy wyłoniono następujące grupy kontekstów medialnych: 1) UE (Unia Europejska) – w tym m.in. takie słowa, jak unia, UE, europejski, prezydencja, Europa, 2) katastrofa (smoleńska) – katastrofa, smoleńska, Rosja, MAK, tragedia, tupolew, zamach, 3) władza – rząd, zarząd, władza, sejm, lider, prezydent, 4) media – media, gazeta, TVP, TVN, telewizja, 5) pieniądze – pieniądze, finanse, budżet, NBP, 6) reformy – reformy, 7) kościół – kościół, 8) prawo – prokuratura, prawo, ustawa, sąd, trybunał itp.

Uzyskane w ten sposób konteksty medialne zostały użyte do analizy merytorycznego charakteru kampanii wyborczej w 2011 r. Jej celem, między innymi, była próba odpowiedzi na pytanie, czy w treściach publikowanych przez portale internetowe intensywniej, w sensie ilościowym, reprezentowane są konteksty mery-

<sup>3</sup> Badaniem objęto partie: Platforma Obywatelska RP (PO), Polskie Stronnictwo Ludowe (PSL), Prawo i Sprawiedliwość (PiS), Sojusz Lewicy Demokratycznej (SLD), Polska Jest Najważniejsza (PjN) i Ruch Palikota (RP).

<sup>4</sup> Na podstawie struktury Rady Ministrów za: Postanowienie Prezydenta Rzeczypospolitej Polskiej z dnia 16 listopada 2007 r. nr 1131-50-07 o powołaniu w skład Rady Ministrów, M.P. 2007, nr 87, poz. 947.

<sup>5</sup> Spis statutów zob. [www.id.uw.edu.pl/zasoby/profile/59/Aneks\\_nr\\_2-Wykaz\\_statutow\\_ministerstw.pdf](http://www.id.uw.edu.pl/zasoby/profile/59/Aneks_nr_2-Wykaz_statutow_ministerstw.pdf) [dostęp: 23.04.2013].

<sup>6</sup> [www.wirtualnemedial.pl/artykul/gazeta-wyborcza-i-fakt-to-najchietniej-czytane-dzienniki#](http://www.wirtualnemedial.pl/artykul/gazeta-wyborcza-i-fakt-to-najchietniej-czytane-dzienniki#) [dostęp: maj 2010].

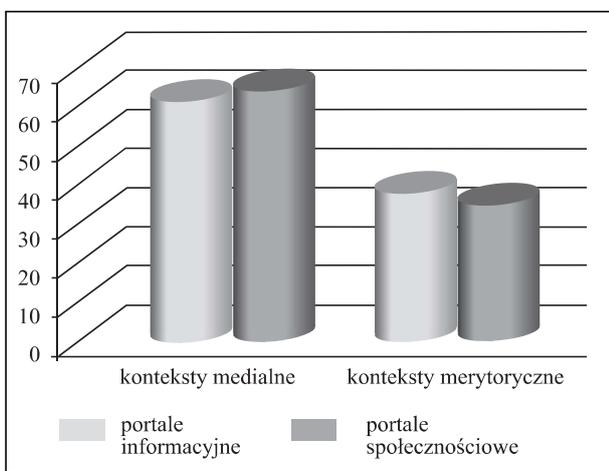
<sup>7</sup> [https://han.buw.uw.edu.pl/han/ISIEM/site.securities.com/search/pub\\_search.html?pc=PL&sv=EMIS](https://han.buw.uw.edu.pl/han/ISIEM/site.securities.com/search/pub_search.html?pc=PL&sv=EMIS) [dostęp: maj 2010].

toryczne czy konteksty medialne. Konteksty rozumiane są tutaj jako łańcuchy blisko występujących obok siebie nazw partii ze słowami uznanymi za konteksty merytoryczne i medialne. Owe łańcuchy nazywane są dalej, podobnie jak w rafinacji mediów społecznościowych, wpisami.

Metodologia badań, w tym sposób badania sentymentów, została opisana w części pierwszej artykułu.

Badania treści wyróżnionych portali internetowych dowiodły stosunkowo niewielkiej przewagi liczebnej kontekstów merytorycznych nad medialnymi.

Traktując rafinację jako sugerowane w opisie badań źródło informacji dziennikarskich, za interesujący można uznać rozkład kontekstów medialnych i merytorycznych w newsach dotyczących wszystkich partii biorących udział w wyborach.



Rys. 2. Rozkład kontekstów: medialne vs. merytoryczne w portalach informacyjnych i w mediach społecznościowych (w proc.)

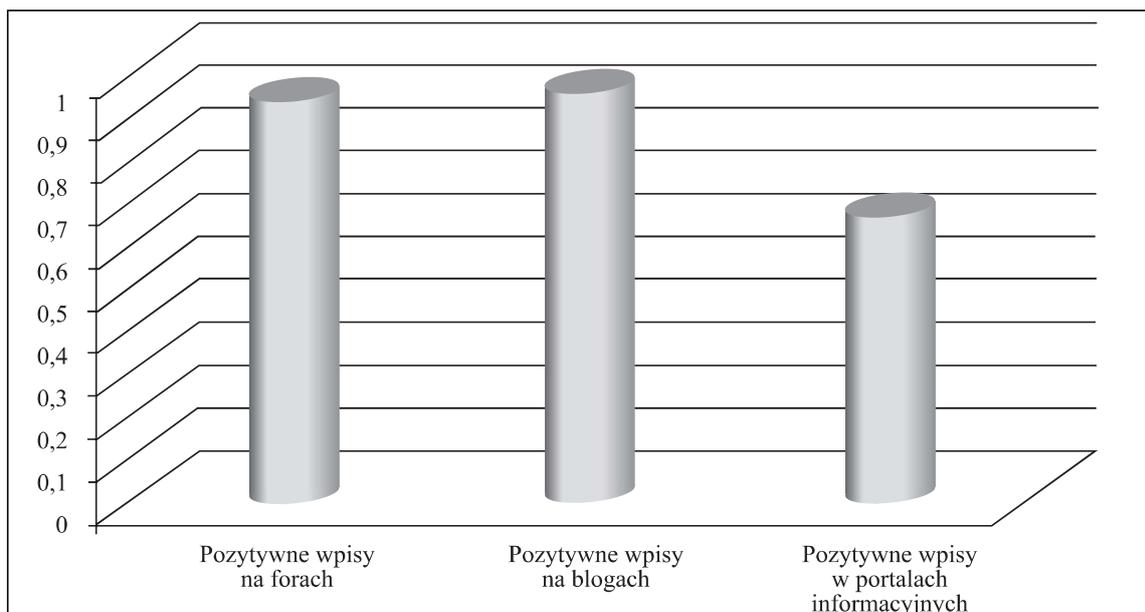
Rozkład kontekstów merytorycznych i medialnych w portalach informacyjnych jest bardzo podobny do analogicznego rozkładu w mediach społecznościowych. Wskazuje to na ogromne podobieństwo wpisów w profesjonalnych opracowaniach medialnych i źródłach społecznościowych. Niemal taką samą uwagę zarówno media społecznościowe, jak i portale informacyjne po-

święcają problematyce merytorycznej (z pomijalną przewagą portali informacyjnych) i medialnej. Wniosek ten zwraca uwagę na dyskusyjną tezę różnic, a właściwie ich braku, w zawartości informacyjnej owych źródeł informacji.

Tabela 1. Szczegółowy rozkład kontekstów w portalach informacyjnych

Konteksty w portalach informacyjnych	Udział w proc.
Prawo	14
Media	10
Praca i polityka społeczna	9
Infrastruktura	8
Sprawiedliwość	6
Sprawy zagraniczne	6
Finanse	6
Edukacja	4
Gospodarka	4
Zdrowie	4
Katastrofa	3
Pieniądze	3
Sport i turystyka	3
Unia Europejska	3
Kultura	2
Nauka i szkolnictwo wyższe	2
Rolnictwo	2
Sprawy wewnętrzne i administracja	2
Środowisko	2
Władza	2
Kościół	1
Reformy	1
Skarb Państwa	1
Obrona	0
Rozwój regionalny	0

Wymowne są różnice wielkości korelacji między liczbą uzyskanych przez partie głosów a 1) pozytywnymi wpisami w mediach społecznościowych na blogach i na forach oraz 2) pozytywnymi wpisami w analizowanych portalach. Dowodzi to znacznie wyraźniej widocznego w sieciach społecznościowych sentymentu (tonu, barwy) do zwycięskich partii niż w bardziej stonowanych treściach portali.



Rys. 3. Proporcje wielkości współczynników korelacji Pearsona między liczbą uzyskanych przez partie głosów a liczbą wpisów pozytywnych (sentymentów)

## Wnioski

Przedstawione wyniki badań wynikające z analizy danych uzyskanych z wpisów na portalach internetowych wskazują na niemal 40-procentowy udział problematyki medialnej (konteksty medialne). Dowodzi to charakteru zainteresowań wydawców portali, które niekoniecznie, jak można by oczekiwać, są zdominowane przez problematykę merytoryczną. Z jednej strony wniosek ten wydaje się cenną informacją, wskazówką co do proporcji zawartości treści zamieszczanych na portalach, a dostarczanych przez profesjonalne redakcje. Z drugiej wskazuje kierunki, które należy utrzymać lub zmienić w doborze treści przez najbardziej zainteresowanych – dziennikarzy.

Nie mniej ważnym wnioskiem jest przesłanka wskazująca na niespodziewanie dużą wartość informacyjną danych pozyskiwanych z mediów społecznościowych. Dowodzą one, że wynik rafinacji sieci społecznościowych (opisanych w pierwszej części artykułu), jest bardziej zbliżony do rzeczywistych wyników wyborów niż wynik analizy treści zamieszcza-

nych w portalach informacyjnych podczas kampanii wyborczej.

Proponowana w opisie badań metoda korzystania z potencjału Big Data, za pośrednictwem rafinacji jako narzędzia dziennikarskiego, wydaje się w pełni uzasadniona – jako nowe, perspektywiczne źródło informacji dziennikarskiej. Ma to szczególne znaczenie w okresie dokonujących się zmian w mediach tradycyjnych skazanych na konwergencję z nowymi mediami i narzędziami, które z nimi są związane. Konieczne staje się dostrzeżenie przydatności wykorzystywania w procesie podejmowania decyzji pogłębionej analizy danych (*data-driven decision making*). Zjawisko to może znaleźć zastosowanie w sferze publicznej, np. w trakcie przygotowania kampanii wyborczych, ale również w podejmowaniu decyzji przez polityków w sytuacjach między wyborami, a także w sferze komercyjnej, w której podejmowanie decyzji na podstawie danych uzyskanych z rafinacji może być istotnym źródłem danych na temat preferencji i zachowań konsumentów.