Collecting and Analyzing Data from E-Government Facebook Pages

Ivana Grujic[1], Sanja Bogdanovic-Dinic[2], and Leonid Stoimenov[2]

[1] Higher Technological School of applied studies,
Hajduk Veljkova 10, 15000 Sabac, Republic of Serbia
[2] Faculty of Electronic Engineering,
Aleksandra Medvedeva 14, 18000 Nis, Republic of Serbia

Abstract. Facebook has introduced a new ways for people to communicate and, at the same time, to share stories and experiences. Therefore, its usage in the area of e-Government becomes every day more and more important. In this paper we will give an overview of an application for collecting data from e-Government Facebook pages of Republic of Serbia which we have analyzed. Found results suggest in which way using Facebook we can improve interaction between citizens and e-Government which leads to maximal potential and usage of these social network.

Keywords: E-Government, Facebook, data mining

1 Introduction

We live in knowledge era and our society in global is dependent on information. As the world becomes increasingly digital, more and more data are being generated. An increasing amount of data is available on the Internet. Only collecting the data can not lead to new knowledge. However, that does not mean that we can not get useful information from collected data.

Some ways for transforming raw data into useful information are by applying data mining techniques or using statistical analysis. Data mining and statistical analaysis mean the discovery of previously unknown patterns and information which, in general, brings a lot of benefits. For example, discovered information can help us make better business decisions or to better realize interest of people, their habits or behaviors. They are applied effectively in medicine, retail, business environment, banking, insurance and what is the matter of this study - in electronic government.

E-Government is defined as the use of ICT (Information and communication technologies) and its application by the government for the provision of information and public services to the people [1]. The aim of e-Government therefore is to provide efficient government management of information to the citizens, better service delivery to them and empowerment of the people through access to information and participation in public policy decision-making [1]. The wide usage of social media sites by people, private sector and public organizations,

A. Madevska Bogdanova, D. Gjorgjevikj (Editors): ICT Innovations 2014, Web Proceedings, ISSN 1857-7288

makes government want to think carefully on how it can benefit from such usage in rebuilding its relationship with citizens, increasing their participation and user engagement level [4]. The terms such as user engagement, user interaction and user activity are used interchangeably in this paper. In the future and with the growth of technology, the significance of e-Government will be wider and more open to new innovative solutions.

Advent of social networks has introduced new methods of communication between users worldwide. Millions of people daily visit Facebook in order to inform or share their experience or impression with other. Also it has allowed users not to feel the evident distance between them.

Facebook is an ideal starting point from which the idea of e-Government should be spread as solution to state administration in order to meet the demands of citizens [3]. The activity about above-mentioned social network becomes unbreakable part of every contemporary mans life.

Ministries of e-Government of Republic of Serbia has recognized the potential of Facebook as a means of reaching massive online population. They have followed this increasing trend and utilized Facebook page mechanism for contributing to promotion and development of e-Government.

The contributions of this study are best shown through the development of application which collects and integrates public data found on Facebook pages of public administration bodies. Application provides a way of active partaking of the users thus providing promotion and incitement of using e-Government services.

On the basis on collected data, the analysis has been undertaken. The analysis is conducted in order to determine how much public administration bodies use Facebook as well as weather the changing of habits could make larger benefits. We tried to find the answers on following questions:

- ☐ In which time public administration bodies most frequently post on Facebook?
- ☐ Which time is the best for posting on Facebook?
- ☐ Is there any connection between the length of posted messages and the user engagement?
- ☐ How many user engagements have been made by 10% of most active users?

Given answers clearly suggest that there is potential and if public administration bodies change the way of posting, they would achieve more benefits from social networks. Eventually, important suggestions are given so that published messages could be seen by more people.

This paper is divided into 5 sections including this introductory remark. In section 2 there are examples of similar initiatives which are conducted in the world. Section 3 gives an overview of architecture of the system. In section 4 there is an explanation data collecting and analyzing. The disccusion is given in the Section 5. Section 6 provides conclusion.

A. Madevska Bogdanova, D. Gjorgjevikj (Editors): ICT Innovations 2014, Web Proceedings, ISSN 1857-7288

2 Related Work

Today, Facebook is the second most popular website and home for a large amounts of data: 250 billion of photos, 1.13 trillion of likes, more than 300 PB user data and more than 10 million applications [2]. Regarding big popularity of Facebook social network as well as big amount of data which appears on that network, it is not surprising that a number of studies have been executed in the world in order to analyze Facebook data.

Facebook can be very influential social media in life of a modern man. The paper "Identifying Influential and Susceptible Members of Social Networks" [11] shows perfectly its influence and susceptibility. The experiment conducted in this paper took 44 days during which 7730 product adopters sent 41.686 automated notifications to randomly chosen targets among their 1.3 million friends. It, also, proved that younger users are more susceptible to influence than older users, men are more influential than women, women influence man more than they influence other women, and married individuals are the least susceptible to influence in the decision to adopt the product offered.

One statistical review, powered by Buddy Media [6] is designed to help retail brands maximize and leverage their Facebook presences. They analyze user engagement of Facebook Wall Posts from nearly 100 of the worlds largest brands during a six month period from January 1 through June 30, 2011. They found out that posts made during non-busy hours receive 20% more user engagement and posts with 80 characters or less receive 66% higher engagement. What can be found very interesting is the fact that Wednesday appears as the best day for posting, while the Friday is the worst day.

Researchers from Swiss Federal Institute of Technology in Zurich in their study [7] analyze the effects caused by the posts shared on a Facebook page in terms of user interactions. The analysis was performed over data collected from fourteen Facebook brand pages over the period of four months, from February to June, 2011. The results have shown that different post cause different effect on the user interaction. Statuses have caused the greatest number of comments while videos have caused the most likes. Interaction duration on the photos was longest. Research has also shown that posting day was not significantly influenced on user interaction.

Similar research was carried out in the Republic of Croatia at the University of Zagreb [8]. In the first quarter of 2013, 63 Croatian Facebook pages were observed. Focus was on the determination of the reactions rate when it comes to posts according to hours or days in a week. Found results have shown that posts published between 12:00 PM and 1:00 PM cause the biggest interaction. The best day for publication of new content is Thursday.

Slightly different analyses have been performed in area of e-Government. Authors of paper "E-Government and Social Media Sites: The Role and Impact" [4] have examined the performances of Jordan e-Government Facebook pages regarding Page Insight. Page Insight gives a number of aggregate metrics, including the number of people who like their Page, how many people see a post and how many people clicked the post [5]. Considering Page Insights, there is

indication of good level of engagement by citizens where they interact positively with the posts published by the page through making a like, comments, share, answering a question or just see the page. Also, the insight shows that page fun base is mostly comprised of men between 25 and 34.

In a close collaboration entities such as the European and German Parliament, local authorities and political parties from different EU countries has developed WeGov [9] Web application that support policy-makers in the analysis of social networks. This project aimed to improve the engagement between government policy-makers and citizens via popular social networking sites. WeGov toolbox include the following: searches and monitors discussions on social networking sites, reorganizes and summarizes comments according to groups of words, highlights important posts by predicting which posts and users will generate more attention and classifies users according to their behavior and interactions within the social networking sites. The project lasted from January 2012 to September 2012.

Considering publicly available papers there is noticeably big interest in analyses and mining Facebook data. Also, according to stated researches it is clear that a similar analysis in the area of e-Government has not been conducted.

3 System Architecture

The application for collecting Facebook data has been implemented as three tiers Web application. It is comprised of server, client and database layers, as indicated at Fig. 1. Client layer is a thin client that runs on a Web browser and simply displays information to end-users. Database layer archives collected data and provides them back on user request.

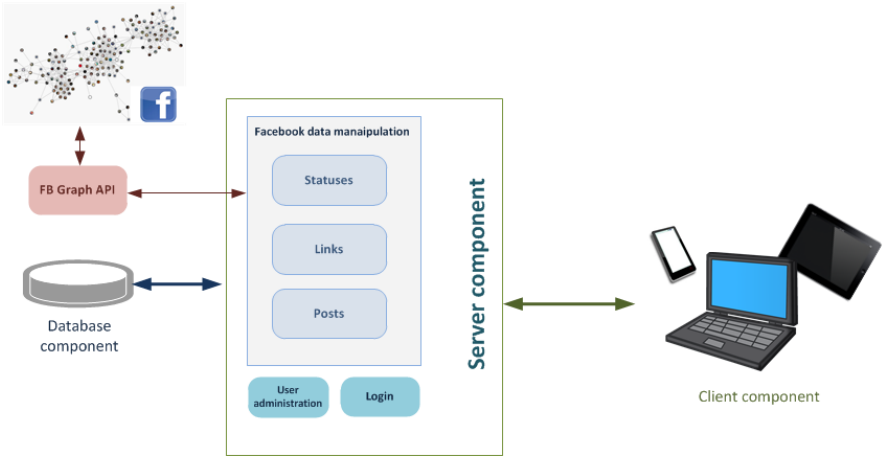

Fig. 1. System architecture.

Server component implements the entire application logic and processes. It is modular and comprised of the following modules: Statuses, Links, Posts, User administration, Login. Login module provides mechanisms for logging into application with an existing Facebook account. User administration module provides functions for managing user profile of a logged user. Statuses, Links and Posts are in charge of manipulating with Facebook data. We have decided to classify these data into three categories in order to be able to perform efficient separate analyses later. Statuses module enables collecting statuses published by the observed Facebook Page and related information of statistical importance: number of likes, time of publishing, user comments. Links are recognized as a special case of statuses: status with a url pointing to resource of importance, as it is very common and could influence completely different reaction of Facebook followers. Finally, Posts represent content posted by observed Facebook page on other pages and topics. This final category could provide insightful information regarding topics that are of interest for observed organization.

Technologies used for the development of this application were open source related: PHP on the server side and HTML 5 and jQuery on the client side. For collecting Facebook data, Facebooks Graph API was used. The Graph API is the primary way to get data in and out of Facebooks social graph. It's a low-level HTTP-based API that you can use to query data, post new stories, upload photos and a variety of other tasks [5].

Based on collected data, the analysis has been conducted using RapidMiner [10] and Weka [12] tools. RapidMiner and Weka are software platforms that provide an integrated environment for machine learning, data mining, text mining and predictive analytics.

4 Data Collecting and Analysing

We should keep in mind that the subject of this study is e-Government and that user through Facebook can get useful information. In order to achieve the goal of this research we have identified thirteen Facebook pages of public administration bodies of Republic of Serbia and then collected relevant data. The list of these pages is given in Table 1.

We collected posts objects found on that pages such as: statuses, notes, links, photos and video clips. Post objects contain the following relevant information: ID of institution, post type, message, creation time, number of likes, number of comments and number of shares.

Gathered data are referred to period of time from the moment in which every single page has been created to the moment of collecting. The collecting has been performed from the end of November to the beginning of December in the year 2013.

Facebook does not restrict when it comes to number of requests or amount of data. Therefore, the collected amount of data has been represented in theoretical maximum that we could get in a particular moment. In total, we have collected 15653 postings (of which 701 are statuses and notes, 4439 are links, 10162 are

Table 1. List of Facebook pages.

Name of ministry
Ministry of Defence
Ministry of Youth and Sport
Ministry of Finance - Tax Administration
Ministry of Agriculture, Forestry and Water Management
Ministry of Justice and Public Administration
Ministry of Energy, Development and Environmental Protection
Ministry of Regional Development and Local Self-Government
Ministry of Finance
Ministry of Culture and Information
Ministry of Finance - Customs Administration
Ministry of Foreign and Internal Trade and Telecommunications
Ministry of Economy
Ministry of Interior (inofficial page)

photos and 351 video clips). On the other hand, when it comes to user activity, we have collected 105373 likes, 4439 comments and 10162 shares.

In terms of data cleaning not much was required, except several posts which have more than 1000 likes we observed only 500 of them. For analysis of data we used Started edition of RapidMiner tool which can read data from Excel or CSV (Comma Separated Value) file formats. We have exported data in Excel file format. Weka tool uses only ARFF (Attribute-Relation File Format), so we have converted data in ARFF file format.

4.1 Time of Posting

In order to answer to this question, we observed the dependence of numbers of posts based on the time of their publication. To make this process as simple as possible, a small approximation was done in the phase of data preparation. For example, all the posts created in the period from 10:00 AM to 10:59 AM have posting time set on 10:00 AM. Analysis has proved that ministries publish the biggest number of posts in the period from 14:00 PM to 17:00 PM. In the Fig. 2 given results are graphically shown via scatterplot diagram.

Using both tools we got the same results. Also, we used K-means algorithm with number of clusters k=24. Graphical representation of clusters is given in the Fig. 3.

4.2 Best Time for Posting

The purpose of this question is to determine in which time (period of time) users are the most available on social network. According to that, which time would

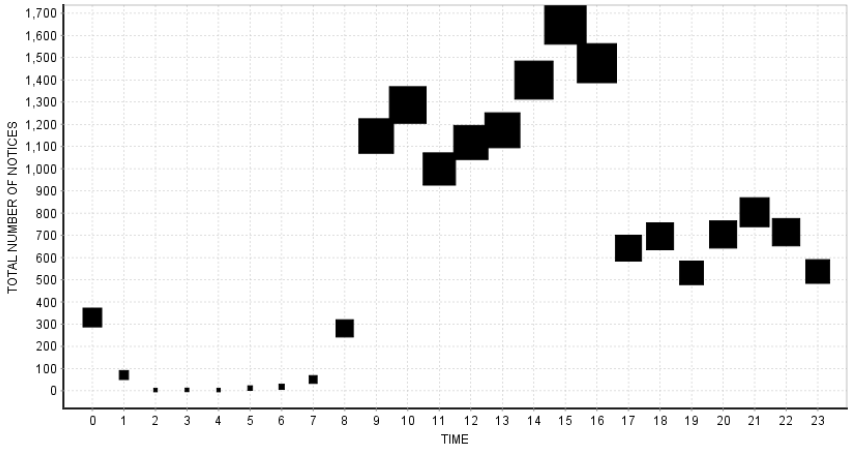

Fig. 2. Numbers of posts by the time of their publication.

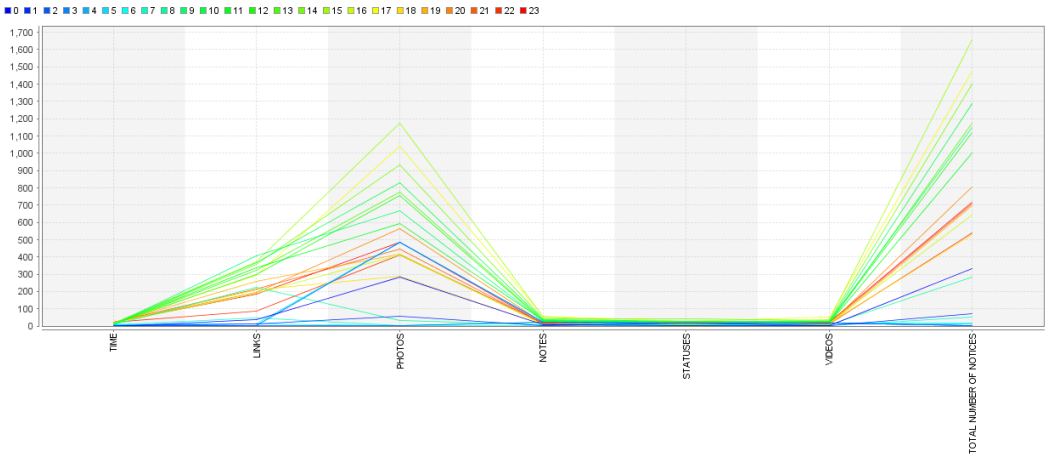

Fig. 3. Graphical representation of clusters using RapidMiner.

be the most convenient for posting because during that time the posts would be accessible to the most number of users. As we realized it is not possible to determine the time in which like was created, the determination of time when user most present was based on number of likes, comments and shares. As a matter of fact, we observed periods of time in which posts are created and number of user activities for that posts. Results suggest that users are most present on social network in the period from 9:00 AM to 17:00 PM, which is illustrated in Fig. 4.

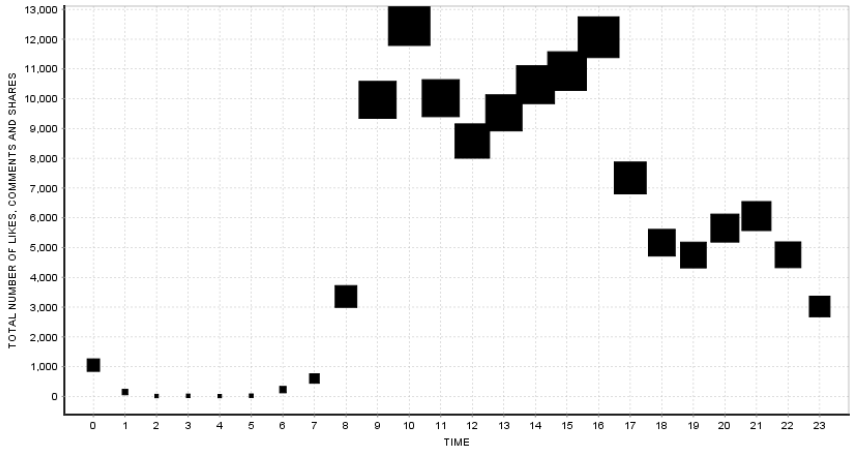

Fig. 4. Number of user activity according to the time of post.

4.3 Dependence of User Activity of Message Length

This question represents an effort to find dependence between length of message (meaning number of characters each message status or text which are followed by link) and number of users who read the particular message. Facebook does not allow insight in the matter considering how many times a new message has been read by user. The only relevant subject which can be used is user engagement. Therefore, we start from assumption that the message which has more likes or comments is at the same time the one that has been seen more. Also, we assumed that people pay more attention on shorter messages which was proven to be true. Fig. 5 shows the relation between users engagement and number of characters of the message.

4.4 Activity of the most active users

In conducted study the activity of 29131 users is depicted. Study proved that 10% of the most active users are responsible for 62% of overall number of likes,

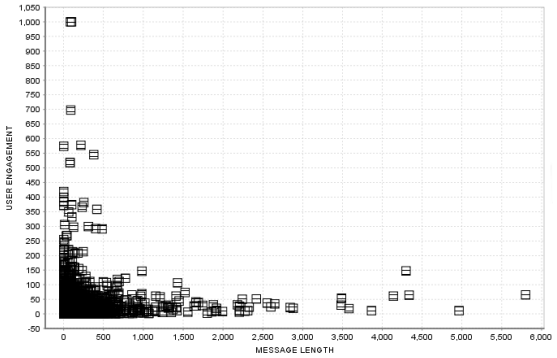

Fig. 5. Dependance of user activity of message length.

comments and shares. According to that, we can conclude that there is a smaller group of users which on daily basis follow the posts from e-Government Facebook pages. Fig. 6 shows the activity of specific groups of users.

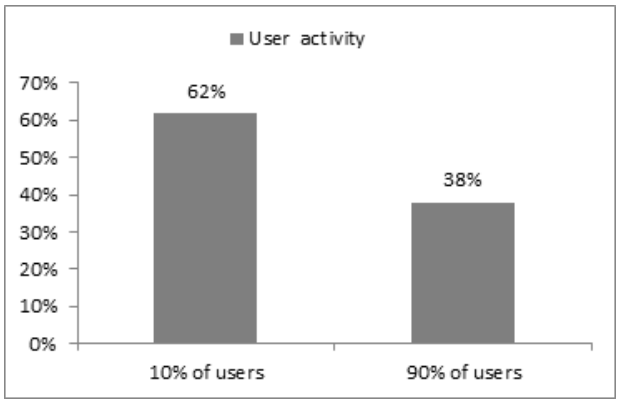

Fig. 6. Activity of specific groups of users.

5 Disccusion

Communication between Government and citizens via social networks abolishes many barriers and provides a more relaxed way of communication which is preferred by citizens. Governments should use social networks but also to use the best of them because in that way they can increase transparency, responsibility and efficiency, anti-corruption, trust and citizens' inclusion. Possibilities for further development e-Government in Republic of Serbia are vast.

In this paper, we have tried to find out what is the effect of posting messages on Facebook on user interaction level. Of great importance is the fact that answers have shown that there is a space for changes which can contribute to better availability of information to the users.

From the charts which present answers to first two questions, we can clearly see that the biggest number of posts is in the period from 2:00 PM to 5:00 PM, as well as the user are the most active in the period from 9:00 AM to 5:00 PM. That means that posting messages in early morning hours would be read by more people.

Taking in consideration user activities according to types of posts we have evaluated that statuses and links have the greatest average number of likes and comments. In addition, recommendation would be that using that types of posts encourage more citizens to take part in it.

Additionally, this research has shown that people read shorter content. That suggests that posting short and concise messages increases reading and number of people which are daily informed through these Facebook pages.

If we compare found results with results of another analysis conducted in the world, we can see that people in Serbia have slightly different habits. People in Serbia use Facebook the most in working hours. What is more and similiar to the people in other countries the are mostly interested in shorter messages.

One of the goals of e-Government is approaching to electronic government to people. Options that Facebook allow us for development of e-Government are numerous. Four of five stages of e-Government development is two-way interaction. On the other way, Facebook provides us exactly that. General solution that was discussed in this paper should be able to contribute fulfilling of emphasized goals and encourage larges involvement citizens of Republic of Serbia.

6 Conclusion and future work

The purpose of electronic government has been recognized by almost all the countries in the world. Especially encouraging is the fact that the majority use social media in order to respond to expectations of the user.

Directions the further development of this application would be taking over different types of data (for example events) and informing the users about events via email. Revealing the posting day which is best for posting and revealing the optimal number of messages which should be published on Facebook pages during some period would be the directions for the future.

All above mentioned should contribute in putting e-Government on the higher lever, especially at the time of world crises and limited resources that we have to work with.

References

1. E-Government Development, http://unpan3.un.org/egovkb/egovernment_overview/ereadiness.htm

2. EBy the Numbers: 98 Amazing Facebook User Statistics, http://expandedramblings.com/index.php/by-the-numbers-17-amazing-facebook-stats/#.Uy8ejKhdUmE
3. Preporuci eUpravu, https://facebook.com/notes/portal-euprava
4. Khasawneh, T.R., Abu-Shanab, A. E.: E-Government and Social Media Sites: The Role and Impact. World Journal of Computer Application and Technology 1(1): 10-17 (2013)
5. The Graph API, https://developers.facebook.com/docs/graph-api
6. A statistical review for the retail industry: Strategies for effective Facebook Wall Posts. Integration. Data report, Buddy Media Inc. (2011)
7. Pletikosa-Cvijikj, I., Dubach-Spiegler, E., Michahelles, F.: The Effect of Post Type, Category and Posting Day on User Interaction Level on Facebook. 3rd International Conference on Social Computing SocialCom. Boston (2011)
8. Istraivanje: Hrvati, postavljajte sadrzaj na Facebook izmedju 13 i 14 sati, utorkom, http://netokracija.com/facebook-istrazivanje-hrvatska-55026
9. WeGov, Where eGovernment meets the eSociety, http://wegov-project.eu
10. RapidMiner Studio, http://rapidminer.com/products/rapidminer-studio
11. Sinan, A., Walker, D.: Identifying Influential and Susceptible Members of Social Networks. Science 337, 337 (2012);
12. Weka 3: Data Mining Software in Java, http://cs.waikato.ac.nz/ml/weka/index.html