Scientific Publications / University of Economics in Katowice

2014 | Economics and Business Communication Challenges : International Week | 229--239

Website Content Recommendations Based On The Signals From Social Networks

Artur Strzelecki

Abstract

The Internet, social media and network communities have been developed successfully for the last years. The implementation of the recommender systems is an obvious result of that development. The author argues that each product presented in Internet can be recommended as well as it can collect the recommendations signals. For that hypothesis's verification, author has gathered SEO and SEM data and he has analysed several websites.

Introduction

Social media have been enjoying a great deal of success in recent years, with millions of users visiting sites like Facebook and Google Plus for social networking, WordPress for blogging, Twitter for microblogging, YouTube and Flickr for video and photo sharing, Digg for social news reading, Delicious and Buffer for social bookmarking, and Pinterest for infographics sharing. These social media sites rely principally on their users to create and contribute content and to mark other users'

content with shares, likes and comments. Social media sites help with establishing the online relationships and join online communities.

Social media sites continue to spread rapidly and excessively, and their volumes of content keep growing. Users are having more and more difficulty choosing sites in which they want to become actively involved. Furthermore, they are overloaded with information from different feed readers, news alert systems and many other social media reminders. Easy access to so much information, as well as difficulty in judging the validity of so much content leads to information overload. Simply put, users have more information available than they can assimilate.

Social media sites resolve these issues by providing users with personalized recommendations via social signals. In traditional websites or e-commerce (i.e. hotels, movies, books), the goal of the recommendation system is to adapt content based on characteristics of individual users [Guy et al. 2010]. Social media sites introduce new types of public data and metadata, such as shares, likes, comments and people relationships, which can be utilized to enhance recommendations.

This paper does not aim at getting simple sums and deciding that a website which gets the biggest number of likes, tweets or plus ones is the best. The determining factor is strongly correlated with the number of published articles. It is obvious, however that quantity has nothing to do with quality. In this paper the author wants to show that each particular article in the Internet can be recommended and can collect recommendation signals. However, the recommendations are usually given to content which attracts readers more than they usual read.

The first part of this article is devoted to the most popular social networks. It provides a description of Facebook, Twitter and Google Plus

and explains what the social signals are. The second part mentions the related works on social media and social signals. In the third part the author describes SEO Planet, a service which collects content from one topic area. The next part presents the results of an analysis made in Excel, based on data received from SEO Planet. The author did a research based on the comparison of different websites monitored in SEO Planet. The author compared social signals for each website monitored in SEO Planet in the three main social media. The conclusions drawn from the research are the subject of the last part of this article.

Twitter

Not so many would have foreseen the impact of social media on the Web, but today blogs, reviews, wikis, and social networks are as much part of the Web as HTML and JavaScript. Even less would still have predicted the success of Twitter, a part of Web 2.0. Twitter is very simple itself. Twitter borrowed ideas from social networks and SMS texts. It allows users to send 140 character text messages (or tweets) to a constantly updating public timeline of user messages. Users can join in to tweets by explicitly following other users, and access or respond to a timeline of messages from these people. The simplicity of Twitter is one of its most well known features. It is a simple way for users to provide status updates to their followers. Twitter users have made a wide variety of use-cases, from political campaigning to education, and from emergency news reporting to marketing and public relations.

Worldwide Twitter is used by many as a form of RSS reader, where users follow their favourite bloggers and news organizations. Twitter has also proven to be a very popular way to share pages. Some commentators

speculate about the potential for social media services like Twitter to be a significant threat to the major search engines because it helps users discover new content in the Internet. The advertisers and marketers have also noticed the potential of Twitter as a way to engage with customers in real-life.

Twitter represents a significant opportunity for users to recommend content in the Internet [Hannon et al., 2010]. The Twitter activity of a user's social graph (their tweets, and the tweets of their followers and followees) provides a useful source of information that can be used as the basis for evaluating content from different websites.

Google Plus

Google Plus (more often written as Google+) is a social networking and identity service that is owned and operated by Google. Google describes Google Plus as a social layer that connects many of its online services, and that it is not simply a social network site, but also an authorship tool that associates web-content directly with its owner/author. Now it is the second largest social network site in the world after Facebook.

Facebook

Facebook is an online social network site. Facebook allows users to create a personal profile, adds other users as friends, sends and receives messages and receives automatic notifications when friends update their profile. Users can join user groups organized by workplace, school or university or other common characteristics and categorize their friends into lists.

Social Signals

Each of social network mentioned before has its own tools that help to share, recommend and comment website content. Social buttons from these networks work inside each social network site, as well as on regular websites. Facebook has the *'like'* button which is a popular and quick way for users to share content with their friends (Gerlitz and Helmond, 2013). A single click on the 'like' button signals approval of content on the web and shares it on Facebook. Twitter has a 'tweet' button which is a small widget that allows users to easily share a website with their followers (Grabowicz et al., 2012). Google has '+1' button which can be added to a website and allows users to recommend content to their circles and drive traffic to a website. The '+1' button can also improve the time spent on a site by providing recommendations for further reading.

Usually the process of recommending content in social media works in a few steps. A user clicks the social media button ('tweet' on Twitter, 'like' or 'share' on Facebook, '+1' on Google Plus). The user is then asked to log in to social network if they are not already logged. In case of Facebook and Google, at this point, the user has already recommended content by giving it 'like' and '+1'. On Twitter, after clicking the 'tweet' button a new window appears which is already completed with the information provided in the properties box. This allows the user to share this content further in the social network. Sharing is a similar activity on Twitter, Facebook and Google Plus. After sharing, a post with the published content appears in the user's timeline in social network.

What is an activity in social media? In social media there are different activities including:

- In Facebook – sharing, commenting, liking;

- In Google Plus – sharing, commenting, giving plus one;
- In Twitter – posting a tweet with a link.

Related work

Currently a considerable research attention is being paid to social networks and integration of social networks and websites in general. Social networks provide access to new types of information and the real-time nature of this data streams provide many opportunities and challenges. Moreover, companies like Twitter, Google and Facebook have opted to make their data available. The developers from social networks created API that provides researchers with an access to a huge volume of information. That is why the recent literature includes a number of interesting analyses of social networks and data generated by users, largely with a view of developing an early understanding of why and how people are using social networks.

The work of Kwak et al. (Kwak et al., 2010) describes a very comprehensive analysis of Twitter users and Twitter usage, covering almost 42 million users, almost 1.5 billion social connections, and more than 100 million tweets. In that publication the authors have examined the mutual exchange and similarities among Twitter users. They have compared a number of different ways to evaluate users' influence and investigated how information flows through the Twitter network as a result of social relationships and retweeting behaviour.

Related work has been carried out by Freyne et al. (Freyne et al., 2009) and Geyer et al. (Geyer et al., 2008) who have explored a number of recommendation techniques for improving user engagement within social media and social networks. It has been seen that the subject matter

of this paper fits in to this body of research interest in recommending content. Furthermore, it is largely driven by the desire to help social network users to create interesting connections to recommended content.

Perhaps closest to the subject of this article is a very recent work by Chen et al. (Chen et al., 2010) on the development and evaluation of URL recommendation strategies to demonstrate the utility of using various combinations of social signals and social graph information in the process of recommending. Like the present work, Chen et al. are clarifying the significant role that social network data can play in the future to recommend content.

SEO Planet – content websites network

SEO Planet (www.planeta-seo.pl) is a website that syndicates content about search engine marketing published in the Polish language. Using this data, which has been collected since 2008 from the articles with specific URLs, the main topic of which was search engine marketing, a summary was prepared on the most frequently recommended websites on which this content is published. It should be established in what way a recommendation to a published content in this network should be given. Recommending content is very similar to forming an opinion or to having influence on somebody else's opinion. A measure that determines if a given content can be treated as opinion-forming and be recommended further are other opinions which have arisen around this content.

Contents published in the internet vary, e.g. there is a difference between an over 2,000-word article which is very interesting and is packed with information and a very short one which looks like news release. Us-

ers tend to feel more loyal to the website where they can spend more time as they have much more to read and analyse.

One of many methods that can be used to evaluate a real website value is to collect opinions that have appeared on other websites. This approach is very similar to PageRank algorithm. PageRank studies the links that interconnect different websites (Page et al., 1998).

Similar to PageRank's recommending system is used in e-commerce. E-commerce would rather collect opinions on them outside, using independent opinion services. It is exactly how the author of this paper wants to collect opinions and comments which arise on published content in the most popular social networks like: Facebook, Twitter and Google Plus.

The author received data from the SEO Planet owner. This data contains the URL address where each content was published, the name of the website, the date of publication, the title of the specific article and the status of the website activity. Some websites, due to different reasons, are already inactive on SEO Planet. After removing the inactive websites 7,731 published articles on 91 websites remained as the final data for analysis. Almost every analysed website was run as a blog and content also had the form and visibility of a blog.

SEO Planet is a data source that stores data in one place. It is possible to download the lists of all the articles ever published on every website, e.g. by downloading a sitemap drawn according to the sitemap protocol for search engines. However, this operation needs to read the entire file and this kind of file should be created for every website (on the sites which are built on WordPress CMS the sitemaps are provided) This ap-

proach has flaws, because what is important in studying recommendations from social media is the date of publication which they lack.

The frequency of recommending content in social media grows with time, that is why the researched data had to be segmented, and the primary segment was the year of publication. Unfortunately, in sitemap files are dated according to their last modification, which makes even the oldest articles acquire a very recent date of publication, e.g. due to plugins that constantly upgrade the articles in case of links failing.

For making an analysis of the data the author used Excel add-on Seotools for Excel. SeoTools is an Excel add-in that adds a number of useful functions for working with search engine optimization and other web metrics directly in Excel. This tool is indispensible when handling vast data as it connects functions from different online services. For example, SeoTools can be used for an onpage analysis and debugging, scraping any website or web service and getting data directly to Excel using regular expressions or XPATH and to monitor social media efforts (Bosma, 2014).

The author validated every URL which was on the list of data for final analysis. Some of them were not working any more so they were removed from the list. After URLs validation, it turned out that not every website is read by SEO Planet via RSS channel. Some of them use a specified RSS publishing service like feedproxy or have their own modified RSS channels. SEO Planet collects articles by reading RSS channel. Seotools for Excel allows to read HTTP Status from specified RSS URLs, which usually returns 301 redirection to base URL of an article.

The author received all core URL addresses which were located after 2013 redirection. After completing the URL list, the author divided

data into segments. Each segment was a period of one year. There are five separate summaries for years 2013, 2012, 2011, 2010 and 2009. This segmentation was needed, because in these years the usage of social networks was different. This is the reason why this period was fragmented. After creating segments, some of websites had double URLs, e.g. with parameter or not, or had duplicated regular URL and URL from proxyfeed. In periods 2012, 2011, 2010 and 2009 doubled URLs were filtered and removed from the author's analysis list**.**

Social media signals on content in SEO Planet network

Table 1 presents the rationale for creating segments. Primary data is divided into years due to an increased number of articles published, as well as the growing usage of social signals like social button in Facebook, Twitter and Google Plus. This table shows the number of articles published during each researched year provided that they are still accessible in the Internet. The articles which are no more accessible have not been taken into account.

Table 1.

Total content recommendations across SEO Planet network in social media.

Year	Articles	Facebook	Twitter	Google+
2009	449	32	0	2
2010	1,279	2,686	346	84
2011	1,553	6,264	823	2,025
2012	2,246	9,335	3,119	7,301
2013	2,005	17,376	3,943	8,702

Source: SEO Planet and Seotools for Excel

Usually, when an author wants to share their articles' content in social media they publish them on Facebook or Google Plus. Thus, right at the start the author acquires at least one activity in each social media net-

work. What is more, for some time now (the author of this paper does not know exactly how long) SEO Planet has its own Facebook and Twitter channel where new content is published. SEO Planet uses URL shortening service bit.ly and the counting function does not see publications on Facebook, but it works well with Twitter, so each newly published content that SEO Planet collects has at least one more Twitter action.

Only average values recorded over a long period of time display whether a website is in fact noticed by users or not. Content published on this website is read and evaluated by users. Users recommend content using social signals in social media. Average values which were collected by articles in the time period 2009 – 2013 are presented in Table 2.

Table 2.

Average value of social signals from social networks for articles in SEO Planet network.

Year	Articles	Facebook	Twitter	Google+
2009	449	0.1	0.0	0.0
2010	1,279	2.1	0.3	0.1
2011	1,553	4.0	0.5	1.3
2012	2,266	4.2	1.4	3.2
2013	1,192	8.7	2.0	4.3

Source: SEO Planet and Seotools for Excel

A number of recommendations could be analysed for each year. For further analysis though there was one more condition introduced. Recommendations were calculated only for the websites which in the analysed period had at least 10 articles published on them. Sometimes there are websites which receive many social signals in social media but, for example, merely 2 new articles were published on them over a period of 12 months**.** Such websites were disregarded in this part of the author's

analysis. Based on this criterion, for each year five websites that had the largest average number of activities in social media were chosen.

In 2009 the list of 5 best websites was rather symbolic, due to a very low number of activities in social media. On Facebook there were only 32 activities recorded. Five websites which were in the top positions in 2009 remained in the lead in the next few years.

Table 3.

Websites' social media activity in SEO Planet network in 2009.

Place	Website	Articles	Facebook	Twitter	Google+	Social Media Average
1	Sprawny Marketing	20	1.6	0.0	0.1	1.7
2	Magiczne SEO i SEM	42	1.5	0.0	0.1	1.7
3	Przemysław Modrzewski	35	0.1	0.4	0.3	0.8
4	Paweł Zinkiewicz	13	0.3	0.0	0.3	0.6
5	Mariusz Gąsiewski	41	0.2	0.3	0.0	0.6

Source: SEO Planet and Seotools for Excel

In 2009 the five analysed websites collected a small number of social signals. This number correlates with activity in that time and frequency of content publication. Now Paweł Zinkiewicz does not publish new content and Magiczne SEO i SEM publish sporadically.

Table 4.

Websites' social media activity in SEO Planet network in 2010.

Place	Website	Articles	Facebook	Twitter	Google+	Social Media Average
1	Silesia SEM	10	13.0	0.0	0.0	13.0
2	Magiczne SEO i SEM	19	7.4	0.6	0.0	8.0
3	Sprawny Marketing	317	5.7	0.5	0.1	6.3
4	Cezary Glijer	36	3.0	1.1	0.1	4.2
5	Sebastian Jakubiec	40	2.8	0.0	0.1	2.9

Source: SEO Planet and Seotools for Excel

In 2010, once again, Sprawny Marketing and Magiczne SEO and SEM made the top five websites and, in the first position was Silesia SEM's debut. The results were based mainly on Facebook activities.

Table 5.

Websites' social media activity in SEO Planet network in 2011.

Place	Website	Articles	Facebook	Twitter	Google+	Social Media Average
1	Sprawny Marketing	171	12.5	1.0	2.5	16.0
2	Cezary Glijer	32	9.1	1.7	4.6	15.4
3	Performance Media	111	8.8	0.4	3.3	12.4
4	Magiczne SEO i SEM	10	6.4	1.4	1.7	9.5
5	Mariusz Gąsiewski	67	4.6	0.4	1.7	6.6

Source: SEO Planet and Seotools for Excel

In 2011, once again, among 5 best websites were Sprawny Marketing and Magiczne SEO and SEM. Second year in a row on the list there were Cezary Glijer and Mariusz Gąsiewski who returned to the ranking as well. This was the first year of the Performance Media website and it had already got to the third position. In 2011 Google Plus social network was created which immediately translated into a noticeably increased number of plus ones recorded.

Table 6.

Websites' social media activity in SEO Planet network in 2012.

Place	Website	Articles	Facebook	Twitter	Google+	Social Media Average
1	Silesia SEM	44	23.6	3.3	10.8	37.6
2	Sprawny Marketing	118	18.2	2.1	4.3	24.6
3	Performance Media	69	12.1	1.8	8.9	22.8
4	Blog Wojtka	129	8.6	3.4	8.2	20.1

5	Mariusz Gąsiewski	46	5.2	1.5	9.1	15.8

Source: SEO Planet and Seotools for Excel

In 2012 Magiczne SEO i SEM disappeared from the top five websites' list. After three years of presence they simply stopped publishing. Another consecutive year Sprawny Marketing, Performance Media and Mariusz Gąsiewski remained in the lead. That same year a new website called 'Blog Wojtka' appeared and it immediately got to the fourth position. In the first place there was Silesia SEM, with a noticeable advantage over other websites.

Table 7.

Websites' social media activity in SEO Planet network in 2013.

Place	Website	Articles	Facebook	Twitter	Google+	Social Media Average
1	Silesia SEM	86	24.3	5.0	15.0	44.2
2	Sprawny Marketing	62	27.3	3.0	5.0	35.4
3	Blog Wojtka	104	95	9.6	14.0	33.0
4	Blog Zgreda	29	19.4	2.4	10.1	31.9
5	Performance Media	37	19.4	2.2	10.1	31.7

Source: SEO Planet and Seotools for Excel

At the end of 2013 the results were very similar to previous years. The best five included 'Blog Zgreda', which started to employ social signals. From the beginning Blog Zgreda had a noticeable number of articles, but it was short of distribution in social networks. The rest of the positions on the list in 2013 were occupied by Silesia SEM, Sprawny Marketing, Blog Wojtka and Performance Media.

Summing up the annual records of the years 2009 – 2013 analysed in this paper, Sprawny Marketing appeared among the best 5 websites every time. It proves that Sprawny Marketing was the most recommended

website during the period in question. Each year they had good content which was always readily recommended. They primarily use their own Facebook page with thousands of observers to publish content in social media.

Three times the top positions in the five best websites ranking belonged to the Performance Media, Silesia SEM, Mariusz Gąsiewski and Magiczne SEO i SEM. Blog Wojtka held the high position twice. Each of this websites has their own Facebook page and viewers who follow content of their sites. Each of them has their own author. These authors write on very specific topics and publish good quality content, which altogether has a positive reflection in social media.

Another approach to segment data is by comparing content recommendations in each social media network separately**.** In Google Plus there are still no advertisements to be found. Content can only be 'earned', not 'paid for'. Facebook, on the other hand, allows to sponsor the published content, hence increase its visibility. Nonetheless, it is still the viewers who decide whether they wish to recommend content or not. Twitter is not too popular in Poland. Its values are very low, but at the same time they are more independent as it is harder to make social activities on Twitter.

Conclusions

Recommendations in social media networks considerably increase traffic on a website. They induce users to be more loyal and to make more visits to a given website. The numbers of social media activities may not be high, as the above analysis showed, they usually are below 10**.** Typically, only few social media users give recommendations in social media.

References

[Bosma,2014] Bosma N. Work with SEO in Excel - The ultimate Excel plugin, http://nielsbosma.se/projects/seotools/.

[Chen, 2010] Chen J., Nairn R., Nelson L., Bernstein M. and Chi E. H. Short and tweet: experiments on recommending content from information streams. In CHI '10: Proceedings of the 28th international conference on Human factors in computing systems, pages 1185-1194, 2010.

[Freyne,2009] Freyne F., Jacovi M., Guy I. and Geyer W. Increasing engagement through early recommender intervention. In RecSys '09: Proceedings of the third ACM conference on Recommender systems, pages 85-92, 2009.

[Gerlit, 2013] Gerlitz C., Helmond A., The Like economy: Social buttons and the data-intensive web. New Media & Society 2013: 1461444812472322.

[Grabo, 2012] Grabowicz P.A., Ramasco J. J., Moro E., Pujol J. M., and Eguiluz V. M. Social features of online networks: The strength of intermediary ties in online social media. PloS one 7.1, 2012.

[Greyer,2008] Geyer W., Dugan C., Millen D. R., Muller M. and Freyne J. Recommending topics for self-descriptions in online user profiles. In Proceedings of the 2008 ACM conference on Recommender systems, pages 59-66, New York, NY, USA, 2008. ACM.

[Guy, 2010] Guy I., Zwerdling N., Ronen I., Carmel D. and Uziel E. Social Media Recommendation based on People and Tags. Proceedings of the 33rd international ACM SIGIR confer-

ence on Research and development in information retrieval, pages 194-201, New York, NY, USA 2010. ACM.

[Hanno,2010] Hannon J., Bennett M., Smyth B. Recommending Twitter Users to Follow Using Content and Collaborative Filtering Approaches, Proceedings of the fourth ACM conference on Recommender systems, pages 199-206, New York, NY, USA. 2010. ACM.

[Kwak, 2010] Kwak H., Lee C., Park H. and Moon S. What is twitter, a social network or a news media? In WWW '10: Proceedings of the 19th international conference on World Wide Web, pages 591-600, 2010.

[Page, 1998] Page L., Brin S., Motwani R., Winograd T., The PageRank Citation Ranking: Bringing Order to the Web, Technical Report, Stanford University, 1998.