UDC 004

MATHEMATICAL MODEL FOR INTERNET PUBLICATIONS' POTENTIAL INFLUENCE ON TARGET AUDIENCE EVALUATION

N. Fursova, S. Kocherezhko
Poltava National Technical Yuri Kondratyuk University
prosp. Pershotravnevyi, 24, Poltava, 36011, Ukraine.

Purpose. To define all the factors that impact on publications' relevancy and potential influence regardless of their source of origin. To instantiate the mathematical model of Internet publications' potential influence on target audience evaluating. **Methodology.** We have applied mathematical simulation of growth of engaged audience field. We have defined correlation between growth of social activity indices and publications' influence considering expansion of engaged audience field's area. Additionally the correlation between relevancy of the search key and Internet publication's influence has been discovered. Thus we have applied predictive analysis in order to instantiate both exact possible value of search request relevancy. **Result.** It was successfully instantiated the term of internet publications' potential influence evaluation that is being exerted on the target auditory basing on statistical data of social activity. It was analyzedwhether the activity of Internet users has any influence on publications' influence. It was also analyzed if there is any possibility to evaluate influence of publications from within search engine look up results considering search key relevancy. As a result, the new mathematical model of internet publications' potential influence evaluation was developed and instantiated. The new formulae of potential influence calculation were successfully instantiated for both publications from Search Engine look up results and social networks publications. **Originality.** This article is the first to apply Pareto principle to internet publications' potential influence evaluation. For the first time, we have discovered the exact formulae to evaluate publication's influence on the target audience. **Practical value.** This model can beused as the primary factor of data ranking. Additionally, its use may simplify the process of reach field evaluation and enhance engaged people search models. References 10, tables 1, figures 4.

Key words: relevancy, influence, reach field, publication, social activity.

МАТЕМАТИЧНА МОДЕЛЬ ОЦІНЮВАННЯ ПОТЕНЦІЙНОГО ВПЛИВУ ПУБЛІКАЦІЙ НА ЦІЛЬОВУ АУДИТОРІЮ В МЕРЕЖІ «ІНТЕРНЕТ»

Н. А. Фурсова, С. С. Кочережко
Полтавський національний технічний університет імені Юрія Кондратюка
просп. Першотравневий, 24, м. Полтава, 36011, Україна.

Визначено фактори, що впливають на показник релевантності публікації та здійснюють потенційний вплив в мережі «Інтернет» незалежно від джерела походження. Проведено математичне моделювання приросту поля охоплення цільової аудиторії та визначено кореляцію між зростанням показників соціальної активності та впливом публікації, враховуючи широту поля охоплення конкретної публікації. Встановлено взаємозалежність між впливом інтернет-публікації та релевантністю пошукового запиту. Це дозволило провести прогноз впливу інтернет-публікації в певний момент часу. Розроблено математичні моделі визначення потенційного впливу інтернет-публікацій на цільову аудиторію, представлено розрахунки на основі статистичних показників соціальної активності. Досліджено поняття «поля охоплення» публікації. Запропоновано використати принцип Парето для визначення впливу публікацій в мережі «Інтернет».

Ключові слова: релевантність, вплив, поле охоплення, публікація, соціальна активність.

PROBLEM STATEMENT. In the modern time of information redundancy, it is becoming more and more difficult to orient in the turbulent flow of incoming data. Hence, both data systematization and data processing increase their complexity as well. It is worth to claim that data relevancy and data influence are considered as two main factors of data processing. Relevancy is an index that reveals to what extent the mined data correlates with the subject field. Influence is an index revealing the target audience reach by some publication or article. Thus, both theseindices are closely connected because irrelevant publication can hardly exert any influence on anyone. However, Internet publications' relevance evaluation is a non-trivial task due to numerous specific features of different publications provider. It isquite hard to represent the resultant indices in a user-friendly manner.However, this task becomes even more complicated by taking into account the fact that different publications sources provide vastly different data sets.

As of now, Internet publications' relevancy and influence evaluation problem is considered as a top-tier task. However, it is not a surprise that both IT-stalwarts (like Google and Microsoft) and less extensive businesses dedicate vast resources to solve this problem nowadays. This hype is caused by tremendous possibilities provided by this procedure.

The most relevant and influential publications selecting allows:

- small businesses and start-ups to use these publications as boost for some products or services (i.e. social promotion via publication comments);
- large businesses to collect customer activity data, intercept negative reviews etc., or in other words – perform brand social monitoring;
- all businesses to detect would-be customers via their social activity thus making it possible to offer goods and services without intermediaries;
- much more tailored data ranking etc.

Social networks publications' relevancy evaluation poses even more challenges. Lack of data set consistency and thus limited set of applicative data are the first to mention. In addition, it is impossible to apply cross-reference mechanism (as one used by Google to evaluate webpages relevancy) makes this task even more complicated. However, ignoring the fact that social networks exert the tremendous influence on target auditory is not a good choice. Vast reach audience and global usage makes every single social network the primary mass media force. Internet publications' reach and influence. That is why selecting relevant publications that can potentially affect the widest possible audience is now very important.

There have not been too many researches conducted on this topic for the last couple of years. However, this factdoes not make this theme less important or valuable.

In particular, the problem of users' social activitywas thoroughly explored by K. Sokolova. The researcher distinguishes three groups of respondents - socially active, moderately active and passive social users [1]. However V. Esina states that only 29 % of people who consider themselves socially active, actually participated in the project as part of the experiment. Only 19 % of them were leading force of the project generating ideas and concepts, the remaining 81 % of volunteers were acting as developers or panders [2].

According to Jacob Nielsen, each social community embraces not less than 90 % of passive users. Their activity is usually limited to reviewing the available materials. These users neither produce any content nor react to any publication exposing their social activity. There is only 10 % of users who may be considered as content-generators – 9 % of them publish a tiny chunk of all publication and only 1 % produce the vast majority of content. Thus 90 % of Internet publications descend from 1 % percent of users [3].

Annouck Welhuis has experimentally proved that the Pareto principle is applicableto user activity distribution considering Twitter users. According to the results he has obtained, 20 % of users generate 84 % of the content. The remaining 16 % percent of tweets were published by 9 % of users that represent moderately active Twitter users. The others do not produce any content [4].

Keith Hampton, Lauren Goulet, Lee Rainey and Kristen Purcell teamed up in 2012 in order to analyze the main patterns of users' social activity in Facebook. They have discovered that only 15 % of Facebook users update their status, 22 % comment on posts and status of other users, 20 % comment on photos of other users, 25 % react to publication provided by other users, 10 % communicate in private messages [5].

All the studies mentioned above have proven the fact that the proportion of active users does not exceed 25 % of the public. However, none of them goes beyond the empirical experiment.

WORK GOAL. This article is aimed to perform a comprehensive research of all the factors that determine the potential influence of Internet publications Mathematical model of the ad-hocInternet publications' influence evaluation development is also on the table. To achieve the goal should be to solve the following problem should be solved:

- define all the factors directly affecting on the potential influence calculation;
- develop mathematical model of the ad-hoc Internet publications' evaluation.
- develop a way to represent resultant indices.

MATERIAL AND RESULTS. There's used the two-group classification of Internet publications in this study. According to it, all the Internet publications are divided into two main groups – the publications containing social engineering elements and ones without such elements. The difference between these couple of groups is obvious - the first group includes only those publications that reveal users' social activity indicators, i.e. the number of «likes», «shares», comments etc. Such a small difference makes many adjustments for further calculations.

First, it is necessary to outline those aspects and factors that define what publications with elements of social engineering are popular, and what are not. Such type of Internet publications is mainly provided by the web platform with advanced social element, social networks, micro-blogging, some blogging platforms etc. (to be short, such platforms will be called - social networks). The number of social network users who have somehow reacted on some publication determines publication's popularity to it - left «like», «reposted» it on the wall of theirs or left a comment [1]. Such reaction indicates the impact on a particular individual. Thus, one of the leading aspects of the Internet publications' popularity and influence are the people - social network users that represent the reach audience of these internet publications. Time is yet another criterion to impact on publication's influence. Time is the cornerstone of calculating the influence factor, as is the time to determine the ratio between the group members, who have reacted (whoare exposed to these Internet publications) and those who have not yet responded to this publication.

Now it is time to get down to data presentation principle in social networks. The most common content providers are communities, public pages and user profiles. A user subscribing to any of these providers acquires access to publications generated by them. Here comes the third criterion for social network publications' influence evaluation – source channel's (or in other words, content provider's) popularity. This connection is quite logical because the more popular a channel becomesthe larger its target audience becomes and therefore increasing publication's reach coverage. In addition, the publications' potential influence is affected by new publications generating speed - in other words, the amount of new publications per fixed period. Too frequent publication generating rapidly lowerssome specific publications down the source channel's feed. Thus the less publications are generated per day, the more influential some specific publication becomes [6].

The first step is to determine the common set of data provided by each social network. The comparative analysis of leading social networks' API showed that all

the major market players provide all the necessary data for further calculations. Hence, the following indices will be used in the upcoming calculations:

– the total number of source community/source public page followers;

– the total count of «likes» for some publication;

– the number of times this publication has been shared;

– feed position;

– new publications generating speed.

These indicescan be divided into two groups, represented in Tab. 1, those that depend on users, and those that depend on the time

Table 1 – Formal classification of potential impact evaluation factors

Factors dependent on users – X-group	Factors dependent on time – C-group
Amount of page followers, (a)	Feed position, (p)
Amount of user reactions, (b)	New publications generating speed, (q)
Number of publication shares, (c)	

In addition, its worth do define the meaning of Internet publications potential influence concept. This ratio indicates the proportion between the ratio of the current reach field's square and the ideal reach field's square considering the possibility of reach field's potential area under ideal conditions.

Here comes the linear functional relationship – publications' influence depends on the number of channel followers Fig. 1.

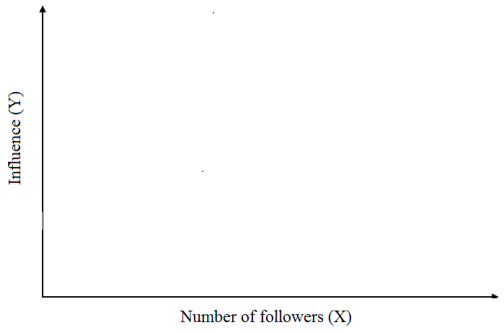

Figure 1 –Visual representation of the relationship between the number of followers and influence considering coordinate system

Both groups of factors (i.e. user-dependent and time-dependent factors) are affecting the resultant influence rate value. Hence, we have a formula to express the function that sets the upper limit of the reach field coverage.

$$Y(X) = (b + 4 \cdot c) \cdot X + \left(\frac{1}{q} \cdot \frac{a}{p}\right) \qquad (1)$$

The first term represents the shape of function considering the number of «likes» (b) and shares (c) by those who users have already seen the publication. The second term represents a potential gain of those users who might see this publication considering publication's feed position (p) and feed upgrade speed (q). According to the Pareto principle, the total number of followers can be divided into 20 % active and 80 % inactive users. Thus the weight ratios: 0,2 to 0,8 can be applied to the first and the second terms respectively [3].

Active members of the group can be divided into 20 % of those who have already seen, and 80 % - those who have not yet. Inactive followers on the same principle can be divided into 80 % of those who may read an article and 20 % of those who will definitely ignore it.

Here we have:

– 4 % - active followers, who have already seen the publication;
– 16 % - active followers, who's yet to see the publication;
– 64 % - non-active followers who will probably see the publication;
– 16 % - non-active followers who will definitely ignore the publication.

Thus, at the initial time we have got:

$$Y(X) = 0{,}04 \cdot (b + 4c) \cdot X + 0{,}96 \cdot \left(\frac{a}{q \cdot p}\right) \qquad (2)$$

Increasing the total value of $0.04 \cdot (\mathrm{b} + 4 \cdot \mathrm{c})$ leads to increasing of the reach field's square.

It remains to determine the reach field's area by integration beyond the total count of channel's followers.

$$K = \int_0^a \left[0{,}04 \cdot (b + 4c) \cdot X + 0{,}96 \cdot \left(\frac{a}{q \cdot p}\right)\right] dX \qquad (3)$$

Thus, it is obvious, that the more followers the channel has got, the larger the reach field is.

Having got the save number of followers, the publication with more "likes" and shares will get the preference Fig. 2.

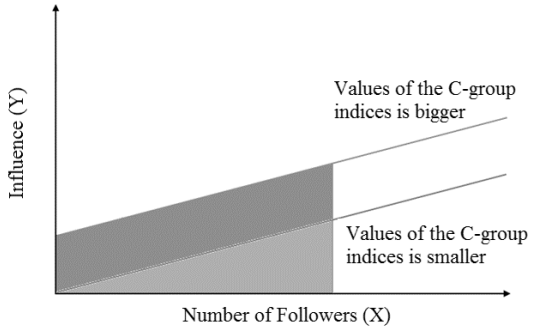

Figure 2 – Reach field's area increasing caused by growth of C-group indices' total value

Having got the save number user reactions, the publication with more potential audience will get the preference Fig. 3.

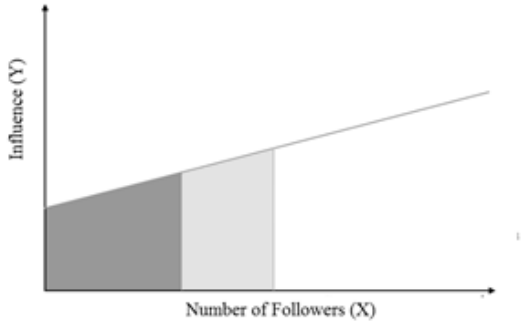

Figure 3 – Reach field's area increasing caused by growth of followers count

Formula 2 reveals current value of the reach field's area.In order to convert it to user-friendly percentage form, we must find the proportion between the ratio of the current reach field's square and the ideal reach field's square considering the possibility of reach field's potential area under ideal conditions. Having applied the Pareto principle, we have 84 % of those users who would probably see the publication under ideal conditions. That is:

$$\bar{K} = \int_0^a \left[0{,}84 \cdot (b + 4c) \cdot X + 0{,}16 \cdot \left(\frac{a}{q \cdot p}\right)\right] dX \quad (4)$$

Thus, we can evaluate the resultant value of publication's potential influence coefficient (PIC).

$$PIC = 1 - \frac{K}{\bar{K}} \quad (5)$$

The publications without social engineering elements are mainly represented by articles and posts that form up search engines' look up results. These publications do not contain any social activity counters, so the potential influence calculation principle given above is applicable to them. In this case, the publication's SERP (SERP – Search Engine Results Position) and the relevance of used search key should be considered as the main factors to be used while evaluating potential influence [7].

As of now, almost every search engine makes it possible to analyze retrospective of the search request's relevance, i.e.provides with graphs depicting how the search request's count has been changing in course of some fixed period Fig. 4.

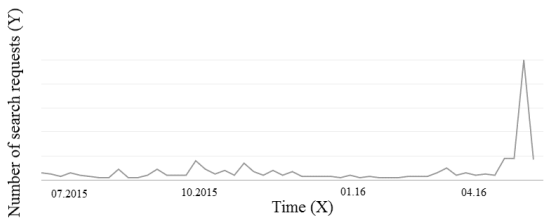

Figure 4 – Search requests count dynamics provided by Google Trends

Having acquired this data set and applied correlation analysis, it is possible to build a predicting function that is to illustrate trending increase or decrease in course of time. That is the distance between the current and predicted, which illustrates search request's relevance.

According to recent studies, SERP directly affects Internet publication's influence [8]. It is obvious that the greatest chance to be visited and read has the publication from the top-ten search results, while the top-5 intercepts ~75 % of total views. 90 % of users limit their search with the first page of SERP Fig. 5 [9].

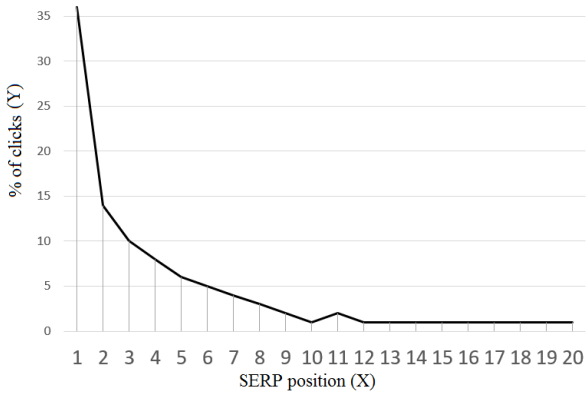

Figure5 – Average count of page views depending on Google SERP

Hence, we can calculate the estimated value of Internet publication's potential influence by taking average values of the search request's relevance and the probability of some page to be viewed depending on SERP. This approach allows us to balance both the publications from within top-10 of SERP by relatively irrelevant demand and vastly relevant publications that did not get to the top-10 of SERP [10].

CONCLUSIONS. The research successfully identified those factors that directly determine the Internet publication's potential influence regardless of its type and source of origin, including how they influence the resultant value in particular. The mathematical model of internet publication's potential influence has successfully been developed and instantiated. This model can be applied to any type of publication from any source (regardless is it a web site or some sort of social network post) if it is provided with the full set of data that are necessary to perform calculations.

The new formulae of potential influence calculation allow representing in user-friendly manner the potential influence rate of any type of internet publications. Thus, it is simple to define what internet publication have more chances to be viewed and therefore have much better potency to reach the target audience.

Data relevancy and influence calculation market is quite extensive, however, as of now, there is not so many fully functional monitoring mechanisms of this kind present on the market. Nearly all existing tools, such as Alexa Page Rank, are proprietary and not publicly available. However, the combination of the principles described above allows creating an intelligent system that can successfully replace some of the existing tools.

REFERENCES

1. Sokolova, E.S. (2012), *Social Activity of the Youth*, *Nauka*, Moscow University of Humanitarian Sciences, Moscow, Russia.

2. Yesina, V.I. (2011), *Developing social activity of the Youth throw project activities*: *Guide for organizational and pedagogical work at the university*, Grodno National University, Grodno, Belarus.

3. Nielsen, J. (2006), "Participation Inequality: Encouraging More Users to Contribute", available at: http://www.useit.com/alertbox/participation_inequality.html (accessed December 15, 2015).

4. Welhuis, A. (2012), "Twitter and Pareto Principle", available at: http://www.annouckwelhuis.nl/twitter-and-the-pareto-principle-2/ (accessed December 10, 2015).

5. Hampton, K., Goulet, L., Rainie, L., Purcell, K. (2012), available at: http://www.pew internet.org/files/oldmedia//Files/Reports/2011/PIP%20%20Social%20networking%20sites%20and%20our%20lives.pdf (accessed December 10, 2015).

6. Boyd, D.M, Ellison, N.B. (2008), "Social Network Sites: Definition, History and Scholarship", *Journal of Computer-Mediated Communication*, no. 13.1, pp. 210–230.

7. Knox, H., Savage, M., Harvey, P. (2006), "Social Networks and the Study of Relations: Networks as Method, Metaphor and Form", *Economy and Society*, no. 35.1, pp. 113–140.

8. Petrescu, P. (2014), "Google Organic Click-Through Rates in 2014", available at: https://moz.com/blog/google-organic-click-through-rates-in-2014 (accessed January 12, 2016).

9. Tober, M. (2014), "The 2014 Rank Correlation Analysis and SEO Ranking Factors for Google", available at: http://blog.searchmetrics.com/us/2014/09/07/the-2014-rank-correlation-analysis-and-seo-ranking-factors-for-google-u-s/ (accessed November 23, 2015).

10. Feeley, M., Rein, D. (2014), "Method and system for achieving an ordinal position in a list of search results returned by a bid-for-position search engine", US 20040068460 A1.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОЦЕНИВАНИЯ ПОТЕНЦИАЛЬНОГО ВЛИЯНИЯ ПУБЛИКАЦИЙ НА ЦЕЛЕВУЮ АУДИТОРИЮ В СЕТИ «ИНТЕРНЕТ»

Н. А. Фурсова, С. С. Кочережко

Полтавский национальный технический университет имени Юрия Кондратюка

просп. Первомайский, 24, г. Полтава, 36011, Украина.

Определены факторы, которые влияют на показатель релевантности публикации и осуществляют потенциальное влияние в сети «Интернет» независимо от источника происхождения. Проведено математическое моделирование прироста поля охвата целевой аудитории и определена корреляция между ростом показателей социальной активности и влиянием публикации, учитывая широту поля охвата конкретной публикации. Установлена взаимозависимость между влиянием интернет-публикации и релевантностью поискового запроса. Это позволило осуществить прогноз влияния интернет-публикации в определенный момент времени. Разработаны математические модели определения потенциального воздействия интернет-публикаций на целевую аудиторию, представлены расчеты на основе статистических показателей социальной активности. Исследовано понятие «поля охвата» публикации. Предложено использовать принцип Парето для определения влияния публикаций в сети «Интернет».

Ключевые слова: релевантность, влияние, поле охвата, публикация, социальная активность.

Стаття надійшла 12.06.2016.