

DEFINITION OF EDUCATIONAL INTERESTS OF SCHOOLCHILDREN ON THE BASIS OF THE ANALYSIS OF USER DATA OF "V KONTAKTE"²

Ulyana S. Zacharova, Artem V. Feshenko

National Research Tomsk State University, Tomsk, Russia
e-mail: fav@ido.tsu.ru

Abstract: *Social media are an important element in the communication policy of the modern university; they allow delivering information to the target audience without intermediaries, providing wide territorial coverage at low cost. But the existing methods of advertising targeting in social networks do not allow universities to determine the individual educational needs and interests of prospective students and to consult them on the choice of educational programme. For this reason, during recruiting campaigns universities create universal social media communities meant for advertising of all educational programs at once. Using this approach, it is difficult to segment the target audience by interests and focus their attention on study programmes correlating with their personal interests. Modern methods of user data analyzing allow universities to conduct a recruiting campaign more efficiently. This article shares the experience that Tomsk State University made with different methods of data analysis to identify in social networks prospective students with pronounced interests in a particular subject area. Methods of content analysis, statistics, surveys, and data mining were employed by a research group.*

Keywords: *data analysis, social networks, educational interests, entrants.*

When the university uses social networks for recruiting students, the biggest challenge is related to the selection of applicants with a strong interest in a particular subject area and motivation for learning. Standard social networking tools to segment the target audience use mainly social, demographic and geographic filters. To determine the needs and interests of prospective students, this data is not sufficient, but they can be supplemented with information about the user contained in his profile: subscriptions to thematic groups and pages, publications on the wall, a network of links, etc. Approaches that allow analyzing and interpreting user data to invoke an

² The study was carried out with the financial support of the Russian Foundation for Basic Research in the framework of the scientific project No. 17-16-70004.

effective information impact are already being used in politics and marketing. The basis of these approaches is the methods of linguistic analysis and psychodiagnosics [Schwartz H., Kosinski M., Markovikj D., Mangal]. But so far the marked solutions are not applied by universities for revealing educational interests and recruiting entrants. Therefore, one of the tasks of our research is to test the hypothesis whether it is possible to recruit applicants for a particular study programme via in social networks for specific faculties and areas of training through identifying their interests in the relevant subject areas.

Thus, the aim of the research is to find methods for identifying high school students among users of a social network with manifestations of interest in a particular science, segmenting the audience in the areas of training, ranking according to the degree of interest, recruiting entrants with the most expressed interests in the relevant departments.

In our opinion, the interest of a high school student in this or that field of knowledge is connected with the probability of admission to a certain faculty of the university. In a social network, the user's interests are manifested through texts, published on his profile page and participation in communities whose careful examination, in our opinion, allows us to determine interest in a particular knowledge area and classify all potential entrants into three groups of interests: humanities, natural sciences, physic -mathematical sciences, and then differentiate between groups of users in terms of the degree of manifestation (power) of interest.

To analyze the text, we resort to the content analysis methodology [Riffe, Lacy & Fico 2005]. By automating the calculation of relevant text units (in our case, separate thematically related words), we planned to identify the interest of an individual user-potential entrant-in a particular area of knowledge.

In addition to the methodology of content analysis, the statistical method was used: variance analysis according to Kraskel-Wallis (Kruskal-Wallis one-way analysis of variance).

The main tool for obtaining data from a social network is the Application programming interface (API). Using the API it is possible to obtain all the public data of the user, including the user's profile fields (name and surname, city, country, gender, education, interests, favorite books, etc.), the content of the user's personal page (walls), and list of thematic groups.

Verification of the hypothesis about the possibility to determine the interests of a social network user by analysis of the texts on her/his wall was carried out by TSU students. The admission campaign to the university will begin only in June 2017 and will end in September, so it is not yet possible to check the methods of analysis on real applicants. At the current stage of the research in the social network "Vkontakte" the profiles of the first year TSU students were selected, and the texts from the wall, published before the moment of entering the university (up to 01.08.2016) were collected. Texts less than 10 KB were excluded from the sample. A total of 232 texts, which amounted to 17% of the original sample.

With the help of the content analysis of thematic communities, dictionaries, defining the text's belonging to one of three topics are compiled: humanities, natural and physical-mathematical texts. Each of the three dictionaries consists of 400 words - markers. We compared the texts from students' profiles with the dictionaries obtained to test the hypothesis about the existence of a connection between the topic of texts on the user's wall and the choice of the faculty when entering the university (Tab. 1). For 85% humanities students, the share of linguistic markers from the humanities dictionary was more than from the natural sciences and physics and mathematics, 9% - less, and for 6% the share of humanitarian texts is proportional to either physico-mathematical or natural-science.

Table 1. Correspondence of the text posted by students on the VK wall to the subject area they aspire to study.

Study area	Correspondence of text subjects to the selected study area		
	Correspond	Does not correspond	Controversial
Natural Sciences	64%	27%	9%
Physics and Mathematics	32%	58%	10%
Humanities	85%	9%	6%

Thus, the method of analyzing texts from the wall of users of "Vkontakte" to determine the interest in a particular subject profile has a number of limitations. First, the amount of text on the wall for objective analysis should exceed 10 kb, which significantly limits the number of analyzed objects from 100% to 17%. That is, this method is not applicable to most profiles of high school students in "Vkontakte". Secondly, the method of text analysis using specialized dictionaries is fairly accurate (85%) only to identify users with humanitarian interests, to identify users with natural-science and physical-mathematical interests, the accuracy of the method is not high enough.

The method of analyzing the content of the wall when identifying the educational interests of applicants is planned to be supplemented by an analysis of the thematic communities in which they are members. Joining a community and subscribing to a page in social networks can characterize the interests of the entrant. If you select from the spectrum of identified interests of the entrant topics, related to education and cognition, then the accuracy of the classification of applicants for subject areas can be increased.

Within the framework of the study, the thematic content of the communities for 18,000 applicants for only one city, Tomsk, was analyzed. From the profiles of

entrants, the communities in which they participate are unloaded and summarized. From the total number of communities were selected 959, only those that occur in profiles of at least 10 users. The definition of the subject matter of the community was carried out manually. As a result, a classifier of communities was compiled and the share of each heading in the total number of communities was determined.

The check of the classifier on 992 TSU students showed, that 66% of them are subscribed to groups and pages, the topics of which can be related to one or another subject area. Comparison of the demanded subject area with the subjects of SM communities to which they are subscribed is presented in the table (Tab. 2).

Table 2. Correspondence of subjects of communities in “Vkontakte” among students in the field of preparation.

Study area	Number of student SM profiles analyzed	Correspondence of SM community subject to the selected study area		
		Correspond on 100%	Correspond more than 30%	Correspond less than 30%
Humanities	324	88%	6%	6%
Physics and mathematics	199	17%	1%	82%
Natural Sciences	139	4%	0%	96%

The accuracy of identifying humanities students with the help of communities' classifier was 94%. The low accuracy of the definition of interests in physics and mathematics and natural science contents can be explained by a limited sample of communities for compilation of a classifier: out of 959 analyzed communities, 231 correspond to topics relevant for humanities, 22 to physics and mathematics, and only one to natural sciences.

At the current stage of research, methods of analyzing texts to be found in user profile, as well as the SM groups and pages they are subscribed to, make it possible to identify only the humanities with high accuracy. The application of these methods to the profiles of potential applicants in 2017 will allow to determine the relative frequency of the mention of linguistic markers in texts on the wall and the absolute values by the number of subject subscriptions for content, related to interest in the humanities. We assume, that these data will allow all users to rank according to

the strength of interest and narrow down the target audience while working to attract entrants to social networks. We will be able to assess the effect of this approach after the end of the admission campaign in August 2017. Expected results: expanding the geography of applicants, increasing the competition for humanities programmes, increasing the share of first-year students, who learned about the university through social networks, reducing the number of deductions from the university in the first year of study, increasing academic performance in the first year of study.

References

1. *Schwartz H.A.* et al. Personality, gender, and age in the language of social media: The open-vocabulary approach // *PloS one.* – 2013. – T. 8. – №. 9. – C. e73791.
2. *Kosinski M.* et al. Manifestations of user personality in website choice and behavior on online social networks // *Machine learning.* – 2014. – T. 95. – №. 3. – C. 357-380.
3. *Markovikj D.* et al. Mining facebook data for predictive personality modeling // *Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013), Boston, MA, USA.* – 2013.
4. *Rutter R., Roper S., Lettice F.* Social media interaction, the university brand and recruitment performance // *Journal of Business Research.* – 2016. – T. 69. – №. 8. – P. 3096-3104.
5. *Mangal N., Niyogi R., MilaniA.* Analysis of Users' Interest Based on Tweets // *Computational Science and Its Applications.* – Springer, 2016. – V. 9790. – pp. 12-23.
6. *Fagerstrøm A., Ghinea G.* Co-creation of value in higher education: using social network marketing in the recruitment of students // *Journal of Higher Education Policy and Management.* – 2013. – T. 35. – №. 1. – P. 45-53.
7. *Riffe D., Lacy S., Fico F.* Analyzing media messages: Using quantitative content analysis in research. — Mahwah, NJ : Erlbaum, 2005.