Psychology for Predicting Internet Behavior Patterns

Nadejda Abramova[1], Olga Shurygina[1], Alexander Kondratiev[1], Ivan Yamshchikov[1`2]

[1]Yandex, Moscow, Russia

[2] University of Applied Sciences Zittau/Goerlitz, Germany

{abramovanadia, o-shurygina, fefa4ka}@yandex-team.ru , i.yamshchikov@hszg.de

Abstract—**Yandex[10] is an international search engine with more than 19 million users daily only in Russia. In order to improve their daily search-experience and make interfaces more user-friendly, Yandex carries out a great number of research activities. This research was aimed at finding a reasonably fragmented audience segmentation that should be based on psychological principles and be automatically processed. We assume that every person has his search behavior characteristics that could be explained by some stable psychological type. By search behavior we mean all the actions the users undertake to find the answers on the Internet: request formulation, quantity, timing and duration of clicks and views, returns to the pages visited before, the habitual usage of tabs and windows, navigation within the service and so on. Using cognitive psychology, we have managed to create the segmentation that has certain predictive power and, therefore, could be used in industrial applications. Based on the qualitative, as well as on the quantitative, analysis of the user-behavior, we have developed two binary scales that split the audience in four groups. The first scale "Analytic – Synthetic" could be roughly characterized as a scale describing the style of information processing, that is more natural for a user and classifies the attention markers. The second one "Logical – Ethical" deals with the informational priorities. We describe these two scales and show how they can be found from the users' behavior. We give a brief explanation of how interactions with the service could be improved based on the knowledge of user's psychological characteristics. Though all results obtained are based on a specific service and can vary from country to country due to the cultural differences, we strongly believe that these two scales, addressing the fundamental priorities of human personality, could be applied for other products and services. Moreover, understanding of general principles that guide the behavior of each user group could help generate user behavior hypotheses to other behavior researchers.**

Keywords - user segmentation; cognitive psychology; Internet behavior.

I. INTRODUCTION

Internet is an information environment that is a part of almost every person's life nowadays. Though every person is unique, in order to improve daily user-experience, a company needs to classify users according to their similarities. When behavior is classified qualitatively, somebody needs to find a quantitative metric that would allow targeting all the users that share the same characteristics. The knowledge of such metrics and the understanding of how we can improve the product or the service for these specific people could be applied to the product and increase user-satisfaction, because it would make uses' lives a bit better and simpler. This approach is called user segmentation and our research is carried out exactly in this framework for Russian search engine Yandex. Our method is based on concepts of cognitive psychology and is to give a possibility to track down people of each psychological type automatically, without any questionnaires, which is rather new and interesting. Firstly, psychological insights can give a broad and detailed description of many different aspects of human behavior; thus giving us great predictive power. Secondly, psychological theories claim to be universal and therefore, applicable to all users. A variety of theories and approaches gave us hope, that we could find an insightful idea of a segmentation that could be general, stable, measurable and weakly dependent on the timely factors such as the user's physical conditions or current mood.

To find out that, we divided our work into three parts: theory and hypothesis setting, qualitative experiments deployment and quantitative checking to verify our hypothesis. So, this paper is organized as follows: firstly, we describe findings of related works and theory behind our classification, and then, we will present the hypothesis, qualitative and quantitative experiments and findings. The conclusion and acknowledgement close the article.

II. RELATED WORKS

The whole concept of using psychology for Internet behavior studies is not really new. A number of researchers have applied psychological ideas and principles in their research of the Internet behavior. Therefore, we mention the most important areas that have a direct connection with this paper.

One of the most successful syncretic areas of psychology and IT (Information Technologies) is the area of adaptive interfaces. In 2011, Susan Weischenk published a book "100 Things Every Designer Needs to Know About People" [1], where she gives a cookbook of practical advices for the interface-designers that is entirely based on psychological theories and concepts. The whole industry of usability is largely based on the idea of applying empirical psychology to the interface-design. The concept of improving the whole user experience with the knowledge of the user's motivation, social-demographics and personal qualities is profoundly discussed by Anderson in [2].

Another area is behavior analytics, where one can find two major research concepts. The first one could be called a unified approach, where one tries to find a psychological feature applicable to all human beings. Stecher and Counts in [3] studied attention, memory and thinking process and came to the conclusion that some characteristics are shared among all users. For example, they proved that trait information is remembered preferentially to the content. But as noted, such patterns are inherent to all people, without going into details why some people are more successful at remembering context or other type of information. The second approach could be called a segment-based approach. The concept behind is that there are distinguishing markers - personality parameters - that could split people into certain groups with the same behavioral patterns. Such approach seems to be really useful in the area of targeting, e.g., [4] [5]; therefore, we decided to follow this path. While it is typical to use segmentation for a specific use-case, we wanted to create a psychological segmentation that could be applied to a vast majority of cases and, consequently, would be of a great industrial interest.

Since the Internet is an informational environment, we decided to focus on the psychological theories covering issues of dealing with information. To be more specific, we were interested in information perception and processing. As a starting point, we used a classical concept of a cognitive style that was first introduced by Bieri [6] in 1955. This concept is well known and widely used in academic psychology, as well as in industrial applications. It was profoundly studied and applied to the needs of education by Riding in [7], [8] and [9]. Assuming that interaction with the Internet is an elementary exchange of information between users and environment, we could apply the concept of cognitive styles to every user action. Moreover, cognitive styles are stable, they are hardly altered by mood or physical conditions and they characterize people behavior over the long haul. All these advantages make cognitive styles extremely promising and interesting for our problem.

III. PSYCHOLOGICAL SEGMENTATION

Our research is simultaneously carried out in several directions. We started with a pure theoretical hypothesis based on the classical psychological concepts and our demands on the segmentation, such as stability and measurability that we have already briefly mentioned. Then, we verified this hypothesis in a qualitative as well as in a quantitative way and now are trying to apply obtained expertise to the products.

A. Theory and Hypothesis

Relying upon the ideas and the theoretical background mentioned in the Introduction we were looking for a segmentation that could correspond to the following conditions:

- Applicable to a variety of use cases, problem contexts and users
- Relevant for our target audience
- Measurable with some standard face-to-face means

- Time-stable
- Detectable, leaving digital imprint
- Able to find predictable behavior patterns applicable to the whole segment

Finally, we have found two cognitive scales that seem especially interesting and promising for us, see Figure 1. The first scale is Analytic - Synthetic. It describes how people deal with different information. Analytic tends to specify, go into details instead of scanning and building up the whole picture. Analytic tends to be focused on and act stepwise solving one problem at a time. Synthetic, on the contrary, is a multitasking person, who tends to generalize and does not like to pay any attention to the details. He switches between issues quickly and could be easily distracted.

Figure 1: Four psychological types based on the cognitive concept.

Logical - Ethical is the second scale that describes which kind of information a person would consider as most valid, interesting and reliable. Logical people tend to focus on facts, measurements, depersonalized and objective arguments, while Ethical people value subjective, person-focused feedback and experience. They also tend to look for personal emotional insights, rather than to look at bare facts.

These two scales with the binary outcome for each split the users in four different groups. In our qualitative and quantitative research, we tried to find and describe these four groups in a greater detail.

B. Qualitaive experiments

The aim of this part of research was to develop a portrait of each psychological type through the deep interviews of 40 people. We were going to find out their Internet preferences and habitats. Therefore, we have carried out a series of interviews with the following procedure:

- At first, a respondent was asked to fill in the forms based on the standard questionnaires used for these two cognitive styles in the classical psychology to understand a person's type.
- Then, the user was asked to carry out a series of simple tasks that involved the usage of the interface.

- • We tracked his attention, focus and motor activity to compare our expert-based impression with the results of the test and with the collected data.
- • Afterwards, we tried different user cases with the same user varying his level of competence in the area and given time for the task.

Based on these interviews and collected data we have managed to see representatives of all four groups and build a qualitative description of the typical representative of each group. Here is the summary of our results, describing four psychological types behavioral patterns:

- • Logical Analytic type
  - ○ searches step-by-step
  - ○ plans his actions before doing them
  - ○ concentrates on objective facts
  - ○ specifies every search parameter
  - ○ tends to fact-check and compare parameters
  - ○ eager to solve the problem in the most effective way
- • Ethical Analytic type
  - ○ searches step-by-step
  - ○ does not perform multi-tasks
  - ○ pays attention to any related type of user generated content eagerly
  - ○ tends to find the most useful solution of the problem
- • Logical Synthetic type
  - ○ searches multidimensionally
  - ○ performs spontaneous changes in the task
  - ○ tends to prefer facts and numbers over a personal experiences, but does not bother with fact-checking
- • Ethical Synthetic type
  - ○ searches multidimensionally
  - ○ sometimes does three or more tasks in parallel
  - ○ switches easily from one task to another one
  - ○ can be easily distracted
  - ○ focuses on personalized experience
  - ○ looks for the most impressive solution of the problem

We have also found some interesting features of the behavior that could be attributed just to the Analytic - Synthetic scale. For instance, Analytics tend to do one thing at a time. Even if you ask them a question, while they are surfing the Internet, they tend to continue browsing only after the question is answered. This observation gave us a clue that content of the site and advertisement should be strictly connected to the query. Synthetics can have more objects in their attention at once but they pay for it with some negligence and carelessness. They even make more typos. That means that content of the web page can contain both context information and additional information as well. So, the design for Synthetics should be easy enough for them to follow several topics or objects they are interested in.

Moreover, as we tracked eye-paths, we have discovered that Analytic people, especially Logical Analytical type people are mostly consistent in their actions. They view every page slowly, from the top to the bottom and do it very attentively. Synthetic type people, on the contrary, are very chaotic. They scan the page in a random way. To get detailed attention patterns of each psychological type is an aim for our future research. However, these findings gave us a clue that interfaces design as well as content should differ for each type in order to increase convenience for the user.

Such substantial differences in the behavior of each type encouraged us to think of further qualitative experiments. Furthermore, all gathered qualitative research results motivated us to move on to quantitative experiments and gave lots of insights for the calculations and data-analyses that we are still doing at the moment.

C. Qualitative research

The first and foremost, we decided to focus on the scale Analytic - Synthetic as on more promising and more insightful one in our quantitative research. We have constructed a number of metrics to verify the fact that these two groups exist, such as reaction rate, number of entry points to start search, query length, regression in navigation and change of topics. But, in this paper, we would like to focus on only two of them, which are the most illustrative.

1) Regressions in navigation: Based on the evidence we have got from the qualitative experiments that Analytics should have more linear navigation patterns in comparison with Synthetics, which we believe have more Internet navigation recursions and loops, we have constructed the following metric.
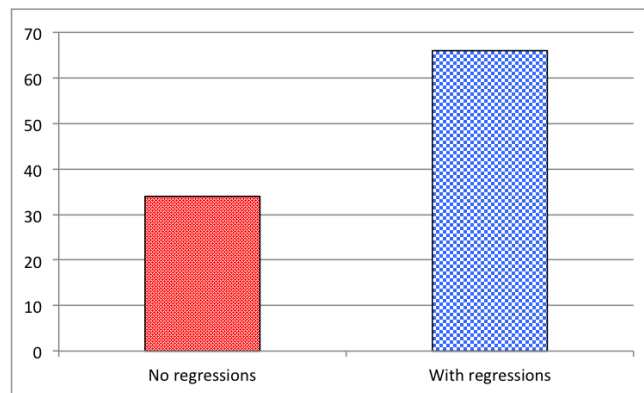

Figure 2: Percentage of the users without and with regressions in the monthly audience of Yandex Russia.

Let us assume that the user has opened a certain page A then moved from it to the page B. Now, if this user does not show any activity on the page B and comes back to the page A within a short interval let us call such pattern a regression. The time interval was estimated as an average time between the click opening the page and the first click on it calculated for this user. We have used one month dataset of approximately 35 million unique users. As it turned out, almost a third of the users did not have any regressions within a month; see Figure 2.

2) Change of the topic: When users search something on the Internet, they naturally input search queries. One can classify if two queries going one by one within one session have the same general topic or are not even slightly connected. As we have seen from the qualitative experiments, Analytics should have no unconnected queries and Synthetics, on the contrary, should change query topics from time to time during one session. It turned out, see Figure 3, that there is another strong dichotomy with approximately 30 percent of the audience not having any single change of the topic within a month.
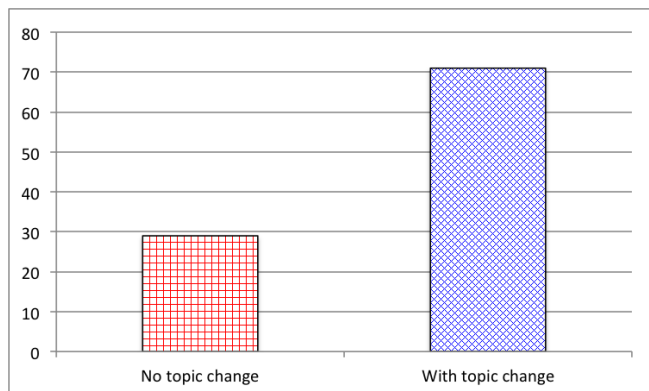

Figure 3: Percentage of the users without and with topic changes in the monthly audience of Yandex Russia.

3) Further quantitative research: We compared these two metrics and got 70% correlation. This fact is extremely interesting because metrics are built on different counters and describe different activities. Synthetics tend to have regressions and changes of the topic and Analytics tend to do everything in a more organized step-by-step way.

We are currently running a number of experiments in order to prove the direct connection between the cognitive styles and the other metrics found. This proof is extremely important for the possible results generalization.

IV. CONCLUSION AND FUTURE WORK

We believe that we have found a very fruitful psychological hypothesis that can be a basement of very general user segmentation. Segments existence was proved by a series of qualitative and quantitative experiments that we briefly mentioned, but could not describe in a greater detail. Our cross-function approach is new in terms of the scalability and generality. It characterizes human personality fundamental qualities and can be applied for a great range of issues from interface improvements, content customization to user targeting. Moreover, we have another series of experiments that should produce additional targeting metrics and prove behavioral changes dependence on cognitive styles and not on some other hidden factor. We also plan to apply the results obtained to a specific industrial product in the nearest future.

REFERENCES

[1] S. Weinschenk "100 Things Every Designer Needs to Know About People", Pearson Education, 2011.

[2] P. Anderson, "Seductive Interaction Design: Creating Playful, Fun, and Effective User Experiences", New Riders Pub, 2011.

[3] K. Stecher and S. Counts, "Spontaneous inference of personality traits and effects on memory for online profiles", in ICWSM 2008 Conference Proceedings, ICWSM, 2008

[4] S. T. D. Ryen, W. White, and J. Teevan, "Characterizing the influence of domain expertise on web search behavior," in WSDM 2009 Conference Proceedings. WSDM, 2009.

[5] D. S. M. Kosinskia and T. Graepel "Private traits and attributes are predictable for digital records of human behavior", 2012.

[6] J. Bieri, "Cognitive complexity-simplicity and predictive behavior," Journal of Abnormal and Social Psychology, vol. 51, 1955, pp. 263–268

[7] R. Riding and I. Cheema, "Cognitive styles - an overview and integration," Educational Psychology, vol. 11, no. 3/4, 1991, pp. 193–215

[8] R. Riding and E. Sadler-Smith, "Type of instructional material, cognitive style and learning performance," Educational Studies, vol. 18, no. 3, 1992, pp. 323–340

[9] R. Riding and E. Sadler-Smith, "Cognitive Styles and Learning Strategies: Understanding Style Differences in Learning and Behavior", London, David Fulton Publishers, vol. 5, no. 4, March 2002, p. 217

[10] company.yandex.com [accessed June 2014]