

Is it better to build multi-language Stackoverflow site?

Australian National University
u5870997@anu.edu.com

April 20, 2017

Abstract

I. INTRODUCTION

II. CONTENTS

- Link
- User
- Post
- Tag

III. LINK

This part of work is based on the links of stackoverflow and Russian stackoverflow. According to the previous results by some other similar research, links are good proof to present the user activity and community activity. Combine the amount of links and the activities of link between these two communities can provide some evidence for the necessity of a multi-language sub site. First of all, we need to find out how many posts and comments used links, and how many people links there evidence in their answers and questions. Then the analysis would be presented in some charts to show the results.

Currently, there are 34,857,917 posts and 55,852,373 comments in Stackoverflow site while 280,424 posts and 499,186 comments in Russian Stackoverflow subsite. Basically, they are divided into two sets, one is the set of posts and comments who use links to stackoverflow.com and the other is those to ru.stackoverflow.com. Some statistics are presented in figures below.

For the posts and comments in Stackoverflow:

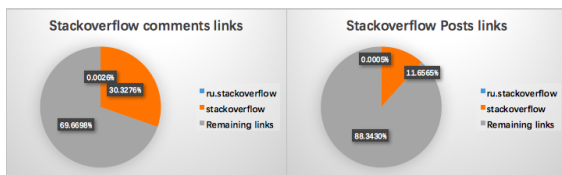


Figure 1: Links in Stackoverflow

Similarly, for Russian Stackoverflow:

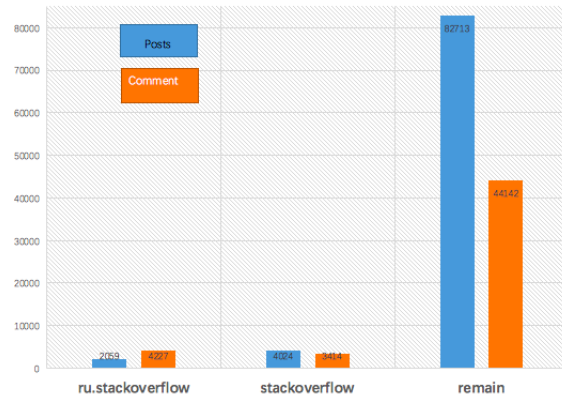


Figure 2: Amount of links activity in Russian Stackoverflow

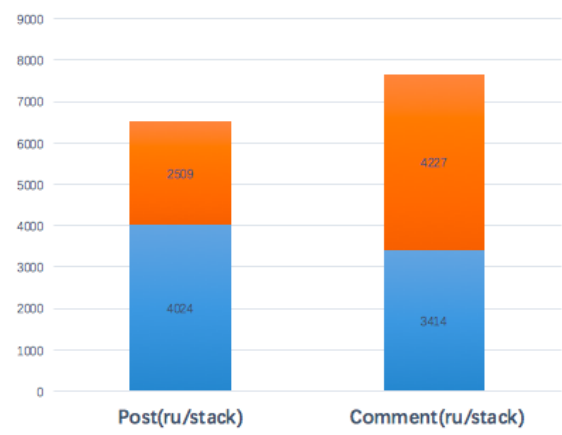


Figure 3: Compare between the amount of links to stackoverflow site and Russian subsite

Apparently, in stackoverflow set, 12,220,205 posts used links in their body text. 11.66% of them link to stackoverflow itself, while only 67 links are lead to Russian subsite. In Russian stackoverflow, 88,796 posts used links in their body text. 4.53% of them link to stackoverflow main site,

while 2.32% of them link to Russian subsite itself. The result of the link activities shows that Russian Stackoverflow subsite is relatively independent community, which support itself by a lot of link activities instead of making the majority of their references to the main Stackoverflow site. However, in some ways the Russian Stackoverflow subsite has a non-negligible demand of knowledge reference to the main site according to the compare of destination between the amount of links.

IV. USERS

It has been widely believed that users data is highly important for the community analysis. User activity can be revealed by data, which have already been structurized into different types. Hence, the analysis of user would be expanded with different aspects.

First of all, the intersection of these two sets is the foundation of this part of research, which is the users that owns th main site account and Russian subsite account at the same time. A point worth mentioning is that not only people from Russia use Russian for introduction and answers, but also some other countries like Ukraine and Belarus. That is the user base called "Russian Speakers".

Stackoverflow main site has 6,833,276 users, while Ru.stackoverflow has 65,623 users art the same time. Each stackexchange user account has a unique accountId, which is always the same if the user sign up for a subsite account like Russian or Spanish Stackoverflow. Using this information, the number of the intersection can be found as 35,751, which is nearly 54.48% of the total number of users in Russian Stackoverflow.

i. CreationDate

Focusing on this intersection user base, the user movement is calculated by their account creation date of each site. As the result shows in figure 4, 11,253 people have their Russian Stackoverflow account first and then sign up for a Stackoverflow account. While the other 24,498 users are the opposite of them. On the one hand, user base is segregated by the new subsite and the number of users leaving the main site to their native language subsite is growing with the time goes by. On the other hand, when 3 out of 10 users flow to Stackoverflow from Russian Stackoverflow, the otehr 7 of these 10 people flow to Russian stackoverflow from Stackoverflow, which means Russian site suprisingly offers a lot of users as feedback to Stackoverflow.

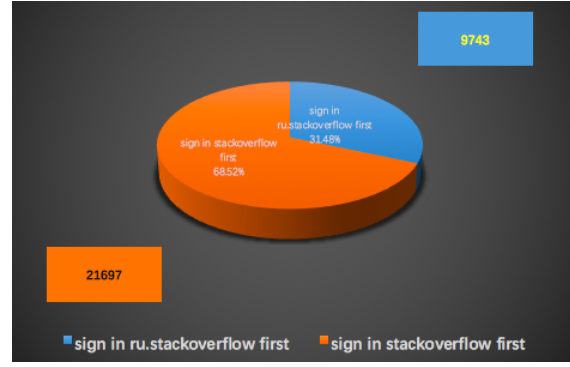


Figure 4: Creation date compare

Here is the statistics for the users sign up Russian Stackoverflow in figure 5.



Figure 5: User movement for signing up Russian Stackoverflow in each month

ii. Location and AboutMe

Basically, all the users who can speak Russian can be found in a simple way. The Russian Alphabet can filter all the people using Russian in their DIsplayname or self introduction of the main Stackoverflow account. In this way, 19,300 users can be seen as a set. Surprisingly, only 3,946 of them have a Russian stackoverflow account, which means 79.55% of these Russian speakers choose to use Stackoverflow site only. There are not a pretty huge number of users use their native language, which is Russian, in Stackoverflow site. And compare this number with the total number of Russian Stackoverflow user base, it is still not a big number. The conclusion is that the majority of Russian Speakers prefer to use English as the only general language on the main site, and the fragmenting level is slight.

iii. Reputation

Reputation is a very valuable information that can be used to reveal the active level of users. The assumption is the user whose reputation is under 20 will be seen as an inactive

user. As the result shows in figure 6, a huge number of inactive users, which includes a number of barely used accounts whose reputation is 1.

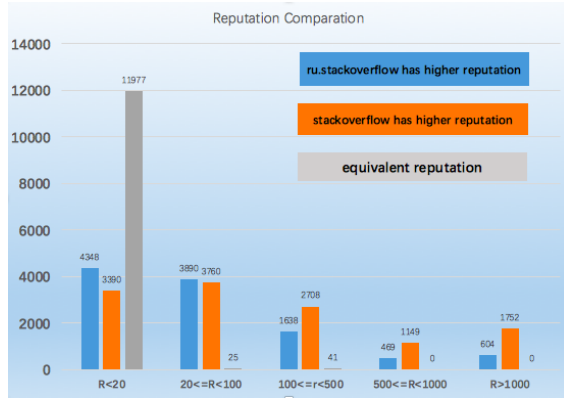


Figure 6: Reputation compare

And the remaining data shows there are more users' stackoverflow account owns a higher reputation than their Russian ones. Only 8,360 users in the 31,440 overlapping set have a over 100 reputation in either of their accounts, and in those 24,498 users who migrated from Stackoverflow to Russian Stackoverflow 6,489 users have a over 100 reputation, which means they are active users.

iv. Core Users

Generally, for each single community, core user groups are the foudation because of their contribution like a large number of posts and comments. Similarly, in figure 7, for the 65623 overlapping users, using the Stackoverflow principles can [need to be figured out by literacy searching] get a group of core user, which has 8360 people whose reputation of either of the account is greater than 100. On the other hind we need to use their creation date again to defenite their direction of migration between these two sites, which is showed in figure 8.

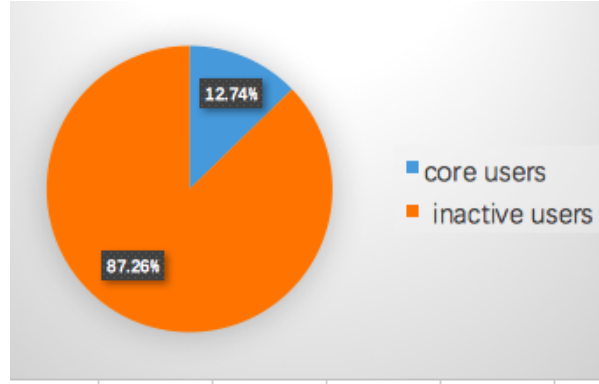


Figure 7: Core users that owns both Stackoverflow account and Russian Stackoverflow account at the same time

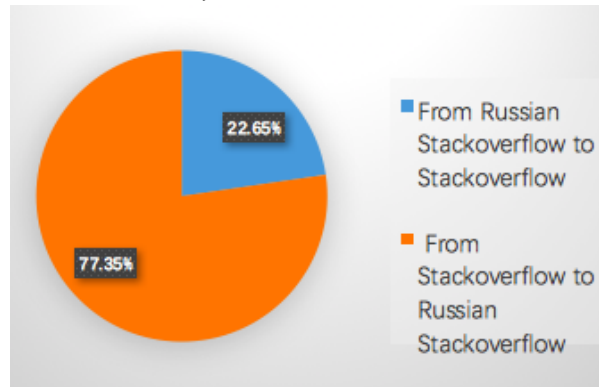


Figure 8: The compare of migration drection

As the results above, Only 1871 out of 11253 users who migrated to Stackoverflow, which is 16.6267% are proved still be core users in Russian stackoverflow site, while 6489 out of 24498 users who migrated from Stackoverflow, which is 26.4879% are proved to be core users.

V. POST

There are 34,857,917 posts in Stackoverflow and 310,539 posts in Russian Stackoverflow. While only 573,927 posts in Stackoverflow and 177,270 posts in Russian Stackoverflow are contributed by the core user group that owns 8360 people. In other words, the 12.74% core users contribute 57.08% of the total posts in Russian Stackoverflow. These core users' contribution of posts are showed in the chart below

Also these core users answer posts and question posts are counted in the chart below

For the 6,489 core users who already have Stackoverflow account before thier creation of Russian Stackoverflow account, the compare between the post activities before their migration with the post activities after their migration can definitely reveal their emphasized aspect. The idea is that

user origin	posts/posts of core users	posts/posts of all users in Russian Stackoverflow
From Stackoverflow to Russian Stackoverflow	44.21%	25.24%
From Russian Stackoverflow to Stackoverflow	55.79%	31.85%

user origin	Answer Number	Question Number
From Stackoverflow to Russian Stackoverflow	464,185	109742
From Russian Stackoverflow to Stackoverflow	124,958	52,314

the object of reference of the users it themselves. Calculate the average post/question/answer numbers during their time when they only use stackoverflow account and after their creation of Russian Stackoverflow account, then compare these two numbers to decide whether a single user is still active or already swithed their avtivity site. the results are represented in the figure 11.