Is it better to build multi-language Stackoverflow site?

Australian National University
u5870997@anu.edu.com

May 11, 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 devided into two sets, one is the set of posts and comments who use links to stackoverflowcom and the other is those to ru.stackoverflow.com. Some statistics are presented in figures below.
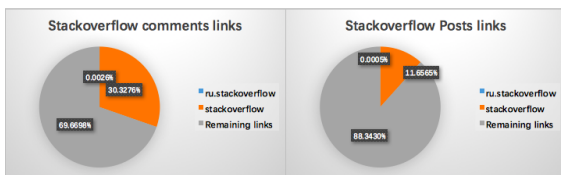
For the posts and comments in Stackoverflow:

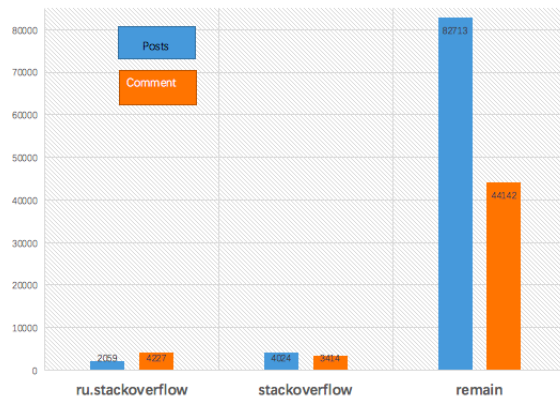**Figure 1:** *Links in Stackoverflow*

Similarly, for Russian Stackoverflow:

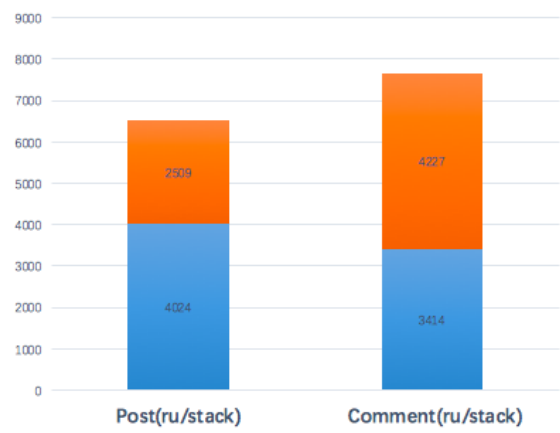**Figure 2:** *Amount of links activily in Russian Stackoverflow*

Figure 3: *Compare betweenthe amount of links to stackoverflow site and Russian subsite*

Apparetly, 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 theis 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 Stackoverlfow 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 refenrence to the main site according to the compare of destination between the amount of links.

Figure 4: *Creation date compare*

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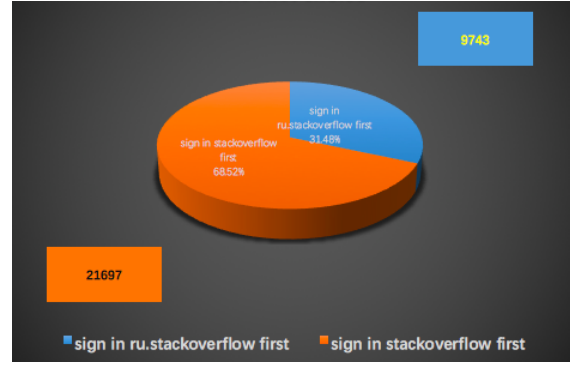
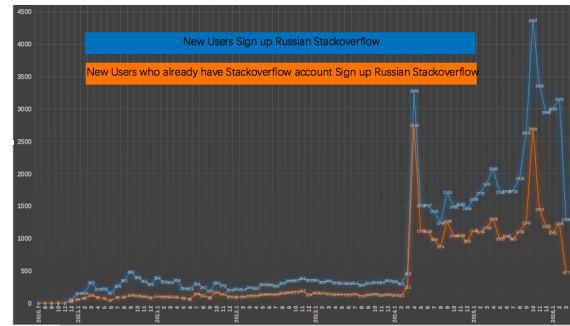Here is the statistics for the users sign up Russian Stackoverflow in figure 5.

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.
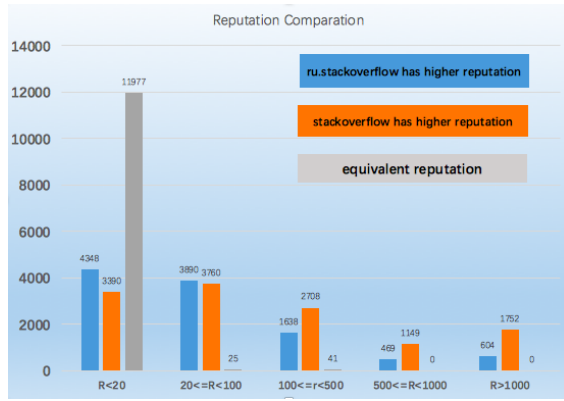

Figure 6: *Reputation compare*

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.

V. Post

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, we can simply get a group of core user by calculating their posts. For the Russian users, simply calculating their post count and the results are presented in figure 8. This result is very consistent with the twenty-eight law, that is, about twenty percent of the user made a contribution of about eighty percent
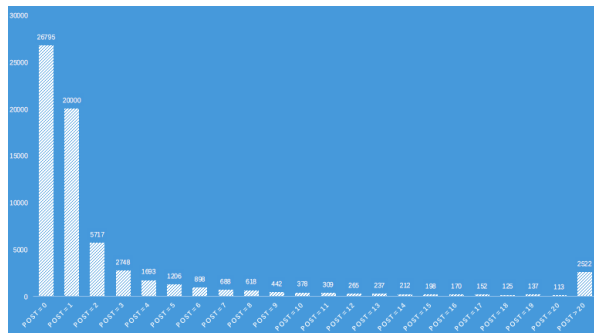

Figure 7: *russian users posts count result*

So we can see that 4818 users who contributed more than 10(can equal) posts own 237,557 posts, which is almost 76.8% of the total amount of Russian posts. So it is better to make an assumption that these kind of users are the core users for the Russian sub site. On the other hind,s we need to use their creation date again to defenite their direction of migration between these two sites, which is showed in figure 9.
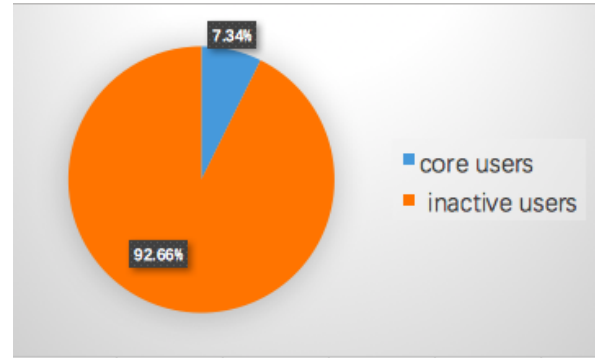

Figure 8: *Core users that owns both Stackoverflow account and Russian Stackoverflow account at the same time*
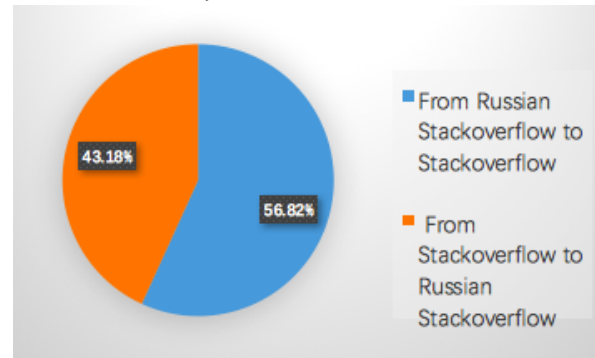

Figure 9: *The compare of migration drection*

As the results above, 1845 out of 11253 overlap users who migrated to Stackoverflow, which is 16.40% are proved still be core users in Russian stackoverflow site, while 1402 out of overlap 24498 users who migrated from Stackoverflow, which is 5.72% are proved to be core users. There are and 310,539 posts in Russian Stackoverflow, and 237,557 posts in Russian Stackoverflow are contributed by the core user group that owns 4818 people. In other words, the 7.34% core users contribute 76.50% of the total posts in Russian Stackoverflow.

For the 1402 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 definitly reveal their emphasized aspect. The idea is that the object of reference of the users it themselve. To com-

pare the post frequency before and after they create a new subsite account, which called user migration.

$$f(x) = \frac{PostCountAfterMigration/TimeLengthAfterMigration}{PostCountBeforeMigration/TimeLengthBeforeMigration}$$

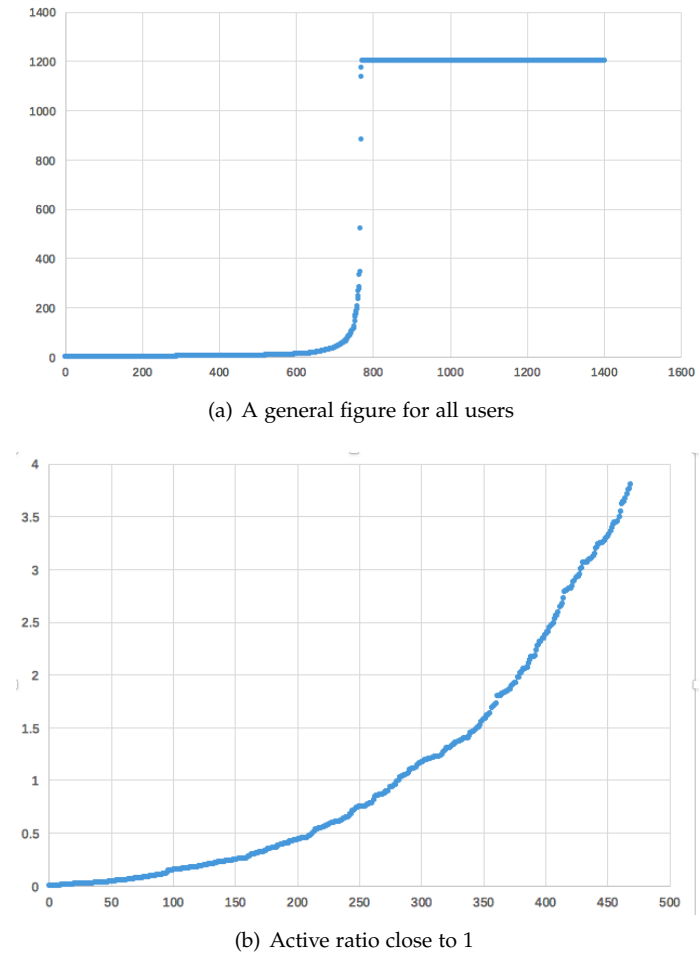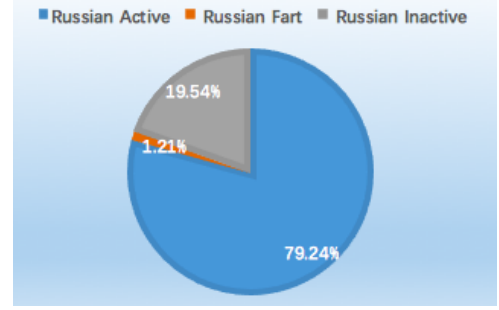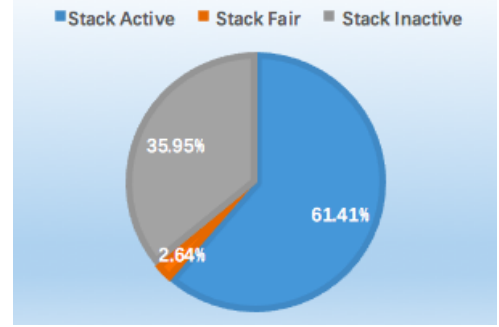Briefly, for all the Russian users, the result of their active proportion is in figure 10

(a) A general figure for all users

(b) Active ratio close to 1

Figure 10: *active ratio chart*

(a) Russian Post Activity.

(b) StackOverflow Post Activity.
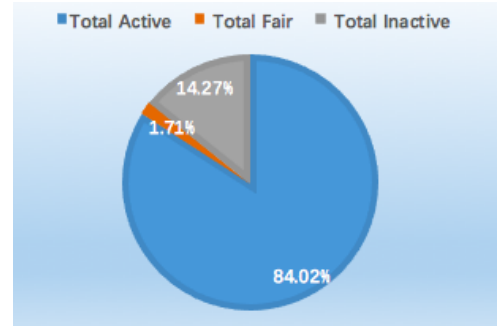
(c) Total Post Activity.

Figure 11: *User Post Activity Result.*

i. PostCount

So there are three level of activity, the first one is active, which means the count of users' post is abov 1.1 mutiple of their previous one. The second one is fair, which is between 0.9 and 1.1. The third one is inactive, which is under 0.9. The result is in figure 10.

And the in these 1402 users who own both Stackoverflow account and Russian Stackoverflow account and migrated from Stackoverflow site to Russian sub site. 861 of them are both on both sites after their migration, and 3 of them are fair, while 200 of them are inactive. The result is in figure 11
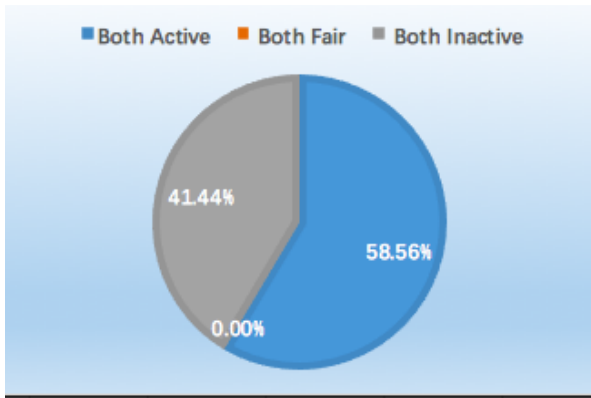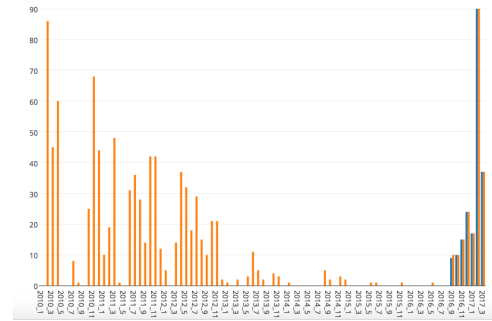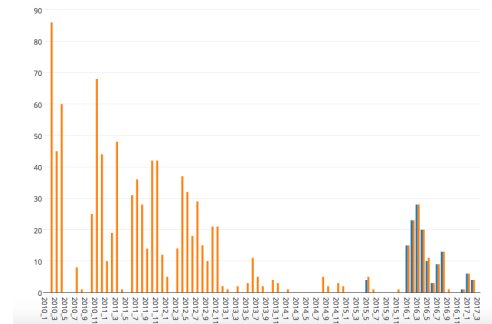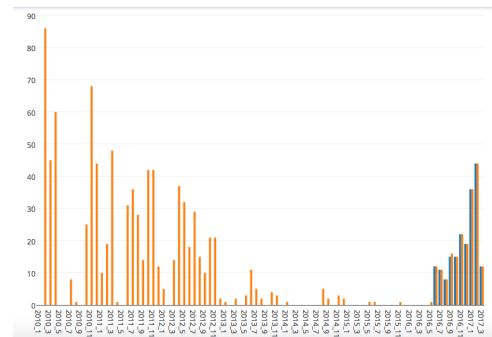
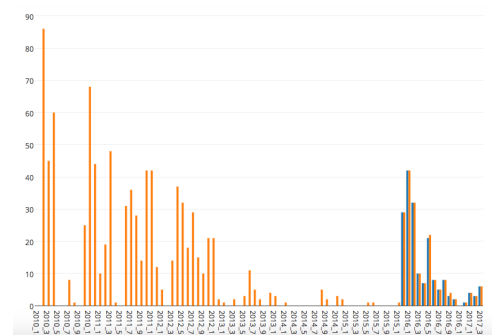**Figure 12:** *Overall result*

(a) Russain Stackoverflow Active.

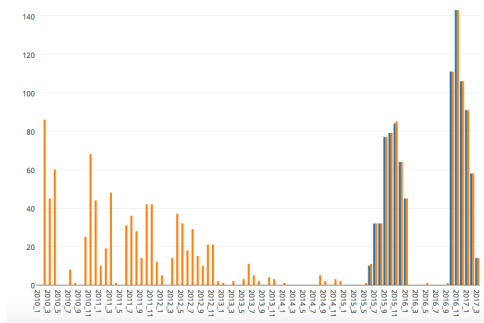
(b) Russain Stackoverflow Inactive.
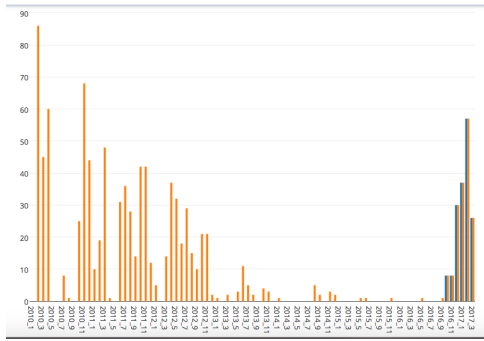

(c) Stackoverflow Activite.
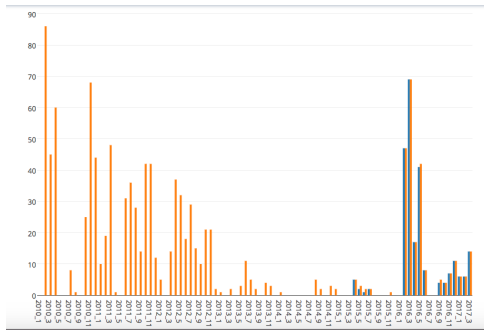

(d) Stackoverflow Inactive.

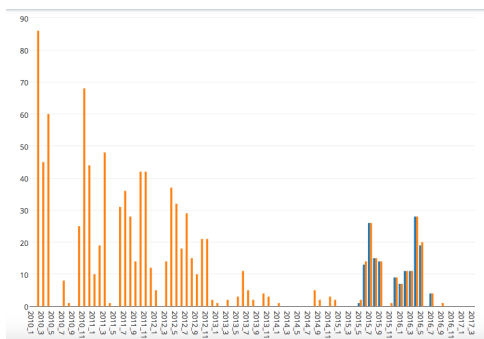**Figure 13:** *8 examples(Part 1)*

(a) Total Active.

(b) Total Inactive.

(c) Both Active.

(d) Both Inactive.

Figure 14: *8 examples(Part 2)*

ii. Comment

For these 3247 overlap core Russian Stackoverflow users , they have 181,643 posts,which includes 59312 questions and 122331 answers. And for those 1402 overlap core Russian Stackoverflow users who from stackoverflow site, they have 77466 posts, which includes 17353 questions and 60113 answers. Compare the average number is , 1:2.062 vs 1:3.464.

The comment amount distribution chart is in figure 14. We can see that the majority of the russian users haven't post any comment. And fot that 3247 russian core users who owns two accounts, their comments number is also calculated.
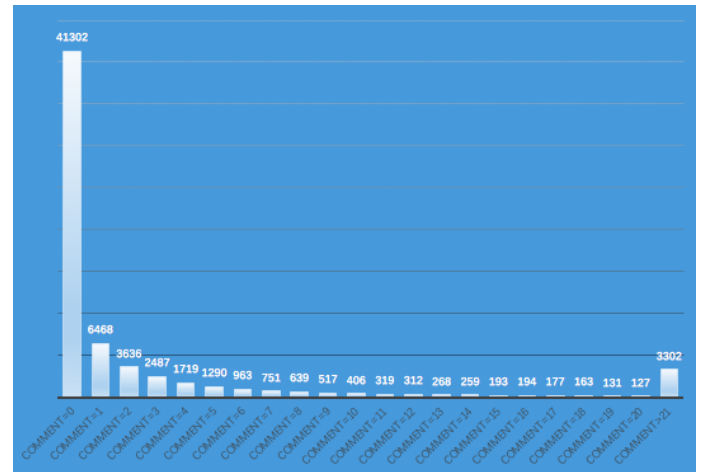
(a) Comment amount in Russian Stackoverflow
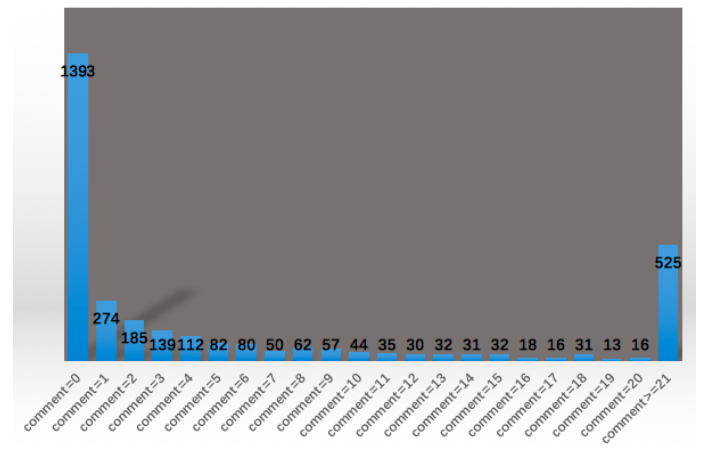
(b)

Figure 15: *Comment amount of 3247 overlap russian core users in Stackoverflow*

//65623 Russain users have 559217 comments(average number is 8.522). 4818 coreusers have 481091 comments(average number is 99.853). And the 60805 users

who are not core users have 78126 comments(average number is 1.285). 3247 core users in the overlap set have 395088 comments(average number is 121.678). 1402 core users in the overlap from Stackoverflow main site have 185845 comments(average number is132.558) some results in figure 15
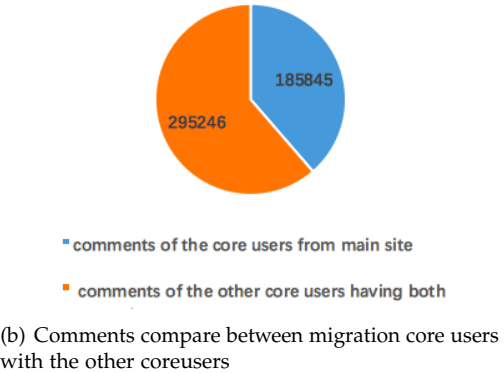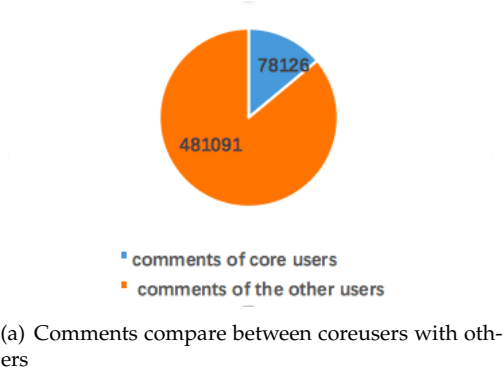

(a) Comments compare between coreusers with others

(b) Comments compare between migration core users with the other coreusers

Figure 16: *comments compare*

So the result is the core users from Main site are the most active users in posting comments. And they contributed 33.23% comments. The average comments number of this group of people is 132.558.