

Magnet or Sticky? A Stack Overflow Tag-by-Tag Typology

Abstract—Stack Overflow (SO) is one of the most popular question and answer website among engineers. There are many tags in Stack Overflow, and with the tag, you can find the questions of your interest smoothly. Many questions and answers are received every day for Stack Overflow. We explored how their interest shifted from how they use tags. We classified tags into four types: (1) attractive, (2) stagnant, (3) fluctuating, and (4) terminal based on magnet values and sticky values. Analysis of stack overflow revealed that: (1) There were some historical events when there were characteristics in the transition of magnet value and sticky value. (2) Since the total number of users of stack overflow increases year by year, the graphs of magnet value and sticky value fall to the right, but if they are normalized, the sticky value of the tag is relatively outputted and a new viewpoint is revealed. (3) In quadrant transitions, tags often remain in the same quadrant [1].

Index Terms—magnet, sticky, tag, user migration, OSS census

I. INTRODUCTION

Pew Research Center [2], an organization that studies problems in the United States and the world, investigated society and population using US tax survey data. States that have a high percentage of people migrating from the outside are defined as *magnet states* and states where the high proportion of the population who continues to live in the same state since birth are defined as *sticky states*. For example, 86% of Nevada's population migrated from other states so Nevada state is quite "magnet". Through such a survey, you can find the tendency how American citizens move.

For many engineers, it is important to know the changes in the interests of other engineers. To make a project better, excellent engineers need to be interested in the project over the long term. According to the definition of magnet and sticky to be defined later, we classified tags used for stack overflow into magnet tags and sticky tags. In addition we define attagnifying strong magnetic and strong sticky *attractive*, strong magnet weak sticky *stagnant*, weak magnet strong sticky *fluctuating*, weak magnet defines a weak sticky *terminal*. By classifying like this, we analyzed what we know from the popularity of tags. We examined tags' magnets and sticky values by classifying them as tags *programming language*, *framework*, *environment*, *OS* and researching theme by theme. We also compared the news and history of the IT industry, if there are characteristic changes in magnet values and sticky values, we examined why it was like that. We addressed the following two issues:

(RQ1) What are typical values of magnet and sticky in Stack Overflow?

In many cases, the sticky value tended to be higher than the magnet value. In addition, the decrease rate was higher for the magnet value than for the sticky value. Among them, the sticky value of swift and go rose.

(RQ2) How do magnet and sticky values change over the time?

In our research we can identify which tags are obsolete. When the tags move quadrant, we find that something happen.

II. DEFINITION OF MAGNET AND STICKY

Measuring Contributor Retention and Attraction in Questions of Stack Overflow

This section describes how we measure the appeal and adhesion of users on different topics on Stack Overflow in this study, we use the Magnet and Sticky metrics defined by the Pew Research Center [26] for illustrating the migratory trends of citizens in the United States.

The Pew Research Center report¹ defines magnet states as those states where a large proportion of adults who live there have moved from another state. Thus, the magnet metric for a state is the proportion of adult residents of a state who were not born in the state. Furthermore, the report also defines sticky states as those states where a large proportion of adults who were born there continue to live there. Thus, the sticky metric for a state is the proportion of adult residents who were born in the state

*These definitions are sound for a study of populations, where a single adult can only occupy one state at a time. However, the definition cannot be applied directly to the topics discussed by the users of Stack Overflow where **a user can ask or answer questions in several topics at the same time. *Therefore, we expand original definition to apply to topics in Stack Overflow as follows:

Topics discussed in Stack Overflow

Questions in Stack Overflow are composed of the content of the question, answers to the questions and comments [3], that are call Posts in the database of SOTorrent. Each question has one or more tags that separate the question into different topics. Simultaneously, posts in a question have their own creator (for the content of question, one is the questioner, and for the answer, one is the respondent) who is participant of the topics of the question. We also define the activity of asking or answering questions in some topic as discussion of the topic. For example, a classical questions in Stack overflow

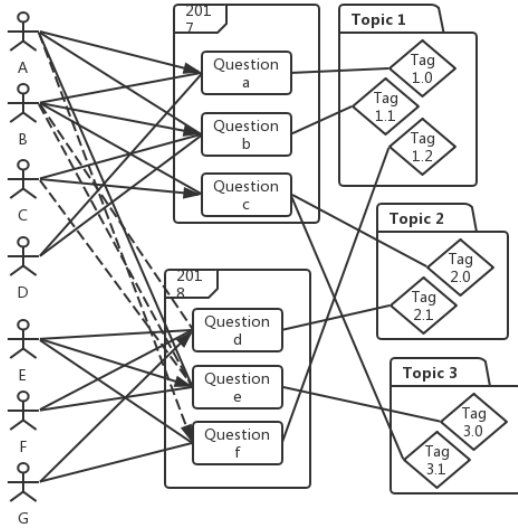


Fig. 1. example graph

has three tags like `java`, `apache` and `linux` which is asked by user A and answered by user B, and C, so A, B and C are participants of topic java apache and linux, by the way, We will discuss the participation of specific topics in several major categories (e.g. C++ and program language), so the three topics java apache and linux are also belong to major categories Program language, Framework and OS.

Magnet Magnets are those that attract a large proportion of new users. Thus, we calculate the magnetism of a topic as the proportion of users who ask or answer during the time period under research to all new users who registered their account at the year.

Sticky Sticky topics are those where a large proportion of the users will keep participating in discussion in the time period under research and the following. Thus, we calculate the stickiness of a topic as the proportion of the users who discuss within the topic in the time period under research to who have also discuss in the following time period.

Let us take the survey of magnet and sticky of some topics belong to a major category as an example. There are 6 questions (a, b, c, d, e, f) and 7 users (A, B, C, D, E, F, G), the Last Activity Date of question a, b, c is during 2017 and question d, e, f is during 2018. the register date of user A, B, C, D is during 2017 and the register date of user E, F, G is during 2018 [4].

Questions have different tags. We merge tags belonging to the same topic into one topic, for example tag "python-2.7" and tag "python-3.6" are merged into topic "python". So we get 3 topics in this case.

To calculate the magnet metric, we observe that there are four new users who register his/her account in 2017 (A, B, C and D), and all of them discuss in topic 1, while two users (B and C) discuss in topic 2 and two users participate in the discussion

of topic 3. In this case, Magnet value of topic 1 in 2017 is $4/4$, topic 2 is $2/4$ and topic 3 is $2/4$.

To calculate the sticky metric, on topic 1, there are three users participate in the discussion in 2017 (A, B and C) but only one of them also participate in the discussion in 2018 (A). Hence, the sticky value of project 1 is $1/4$. on topic 2, there are 2 users participate in the discussion in 2017 (B and C) but only one of them also participate in the discussion in 2018 (B). Even though new users E, F, G participate in the discussion in 2018, we still calculate the value of sticky as $1/2$. For the same reason, the sticky of topic 3 is $2/2$ in 2018.

III. STUDY RESULTS

We got research results and faced two questions against these results. We discuss the questions based on the results.

(RQ1) What are typical values of magnet and sticky in OSS?

We have calculated magnet and sticky values as defined in section 2. And we plotted the magnet value on the vertical axis and the sticky value on the horizontal axis. We classified the plotted points into 4 quadrants.

Attractive: Tag with high magnet and sticky value. By knowing attractive tags, we can find out what the engineers are interested in.

Fluctuating: Tag with high magnet and low sticky value. This tag attracts people but it is short-term. Excellent engineers will not continue to be interested.

Stagnant: Tag with low magnet and high sticky value. These tags are difficult to attract new users, but maintain existing users.

Terminal: Tag with Low magnet and low sticky value. This tag can neither attract new users' engineers nor keep them interested.

In this paper, the median of magnet and sticky values for each year is used for the threshold of the quadrant because the median value is not much affected by outliers. As we showed the sticky value definition in section 2, the sticky value depends on the number of tag users in that year and the following year. So in order to answer RQ1, we got 9 years' worth of sticky value from the information on the number of tag users from 2009 to 2018. The sticky value must depend on the number of new tag users but if the number of new tag users in the target year is too low, the sticky value will be too small. Therefore, in order to remove noise, we decided thresholds for each topic, and we set them all below zero.

Quantitative results: Figure 2 is the magnet vs sticky value quadrant plots of the 2010 framework tag. It shows that there are points with extreme magnet and sticky values. The attractive tags are in the first quadrant, the stagnant tags are in the second quadrant, the fluctuating tags are in the third quadrant, and the terminal tags are in the fourth quadrant. From the figure, it can be seen that the value of magnet is considerably lower than the value of sticky. In other words, tags used for stack overflow are generally sticky instead of

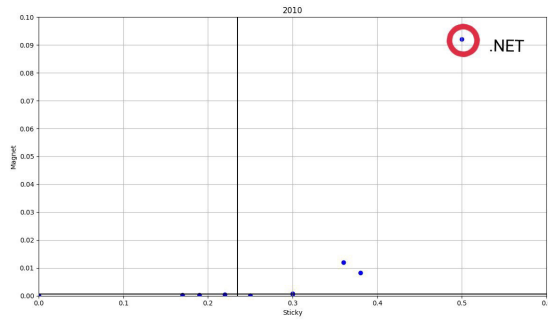


Fig. 2. Distribution of Magnet and Sticky values in framework

magnet. This also applies to the investigation of the Pew Research Center mentioned earlier. Engineers tend to continue to develop using the tags of the same content so that those who continue living on the same land rather than citizens living in the same place will be able to spend more time.

Manual analysis: From the figure2 .NET is high sticky and magnet value. It can be said that it is very good to be a high magnet and sticky value. The .NET Framework was published by Microsoft in November 2006. This is the foundation system for building applications on the network. Microsoft .NET aims to connect all digital information equipment to the network, as well as electronic equipment such as personal computers, servers, mobile phones, smart phones, home appliances, etc. and to be accessible anytime via the Internet. It is the .NET Framework that framed it using the .NET environment. On this framework, it is possible to develop and execute software supporting the .NET environment. The reason why the .NET magnet value is high is beginner engineers can develop somewhat advanced software. The reason why .NET's sticky value is high is its convenience.

Sticky is a more general attribute than magnet. Tags with high magnet value are easy to use even for beginners. Especially magnet tags like .NET are relatively easy to use, so it is widely used for beginners and others. Machine learning gradually became fashionable around 2015 and as it is now, sticky values such as machine learning library like torch are high.

(RQ2) How do magnet and sticky values change over time?

Approach: We investigated how the tags move with time over the quadrants of Figure3.

Quantitive results: Figure3 shows the quadrant transition likelihood using the state transition diagram. The values on the arrows indicate the possibility of transition from each quadrant to quadrant. For example, Figure 3 shows that 25% of programming language tags moved from Stagnant to Attractive from 2017 to 2018. We use * to represent year of less than 100

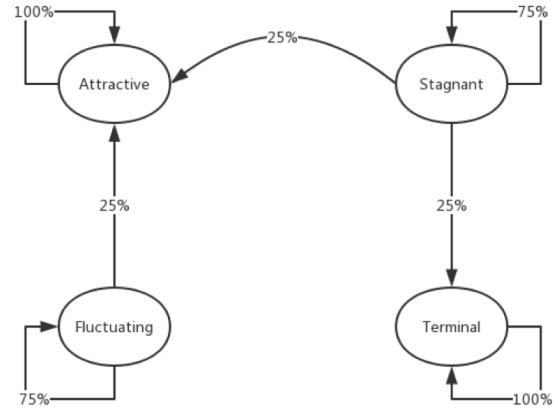


Fig. 3. The Transition of Programing Language from 2017 to 2018

users of tags. From the figure3, you can see that only Stagnant tags moves to Terminal. Fluctuating that attracts new people even if there are at least a few long-term users has a possibility of transitioning to Attractive, but Stagnant that many people are using for many years cannot tie up in trends and no one can use. Also, you can see that all tags often stay in the same quadrant. Tags that were Attractive in 2017 remain Attractive for 2018, and Stagnant or Fluctuating tags become popular and sometimes become Attractive tags.

Manual analysis: Figure4 shows the transition of the quadrant of each tags. From this figure, we can learn how tags move quadrants. As can be seen from figure4, the programming language tags often stay in the same quadrant. It turned out that tags frequently move between Attractive and Stagnant, Terminal and Flactuating.

As a result of investigating from 2009 to 2018, in the programming language, the average transition from Attractive to Attractive was 91.5%, the transition from Fluctuating to Fluctuating was 77.8%, the change from Stagnant to Stagnant was 80.6%, and the change from Terminal to Terminal was 87.2%. In the programming language, it turned out that the classification of tags was difficult to change.

IV. CONCLUSIONS

Whether it's a programming language or a program framework or an operating system, keeping the community alive and attracting more people to participate in discussions is critical to its development. Especially on the stack overflow, the world's largest program Q&A platform, having more questions and answers on a topic means that customers of the product are more likely to solve their own problems, which is even more tedious than that developers rack their brains to write a lengthy development document or Q&A This paper applied the magnet

Programing Language Name	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
JavaScript	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive
HTML.*	Stagnant	Attractive	Attractive	Attractive	Attractive	Stagnant	Stagnant	Stagnant	Stagnant	Stagnant
CSS.*	Terminal	Terminal	Stagnant	Stagnant	Terminal	Terminal	Stagnant	Stagnant	Terminal	Terminal
C++	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Stagnant	Stagnant	Stagnant	Stagnant
SQL	Attractive	Stagnant	Stagnant	Stagnant	Stagnant	Stagnant	Stagnant	Stagnant	Stagnant	Stagnant
Java	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive
Bash	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal
Python.*	Attractive	Attractive	Fluctuating	Fluctuating	Attractive	Attractive	Fluctuating	Fluctuating	Attractive	Attractive
C#	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive
PHP	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive	Attractive
C	Fluctuating	Fluctuating	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal
TypeScript	*	*	*	*	Terminal	Terminal	Terminal	Fluctuating	Fluctuating	Fluctuating
Ruby	Terminal	Terminal	Fluctuating	Fluctuating	Fluctuating	Fluctuating	Fluctuating	Terminal	Terminal	Terminal
Swift	*	*	*	*	*	Terminal	Fluctuating	Fluctuating	Fluctuating	Fluctuating
assembly	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal	Terminal
Go	*	*	*	Terminal	Terminal	Fluctuating	Fluctuating	Fluctuating	Fluctuating	Fluctuating

Fig. 4. Quadrant Transition of Programing Language 2009 - 2018

and sticky population concepts to a set of topics in Stack Overflow. We find that:

1. The number of topics that people participate in is exploding with the development and popularity of computer technology. Even the most popular themes can't attract the high percentage of people involved in the discussion like what they did ten years ago.

2. Under their respective major categories, the most popular topics are still very popular after ten years, and only a small number of languages or frameworks can stand out and become one of the most popular topics.

3. This research can provide some reference for enterprises to choose their own main technology stack, and can also be used as a reference for computer science students to learn new technologies, because it (1)predicts the trend of computer technology in the next few years, (2)points out which technologies are easier to access and the questions can be easier to get answers to.

ACKNOWLEDGMENT

This research was partially supported by ???

REFERENCES

- [1] K. Yamashita, S. McIntosh, Y. Kamei, and N. Ubayashi, "Magnet or sticky? an oss project-by-project typology," in *Proceedings of the 11th working conference on mining software repositories*. ACM, 2014, pp. 344–347.
- [2] C. for Community and C. E. Economic Development University of Wisconsin-Extension, "Community & economic development update," 2011, <https://myemail.constantcontact.com/News-from-the-Center-for-Community—Economic-Development.html?soid=1104293309477&aid=FnOfZtbhFgo>.
- [3] X. Liu and H. Zhong, "Mining stackoverflow for program repair," in *2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 2018, pp. 118–129.
- [4] K. Yamashita, Y. Kamei, S. McIntosh, A. E. Hassan, and N. Ubayashi, "Magnet or sticky? measuring project characteristics from the perspective of developer attraction and retention," *Journal of Information Processing*, vol. 24, no. 2, pp. 339–348, 2016.