Analyse Badges in StackOverflow

Peter Baumgartner

2018-05-21

Contents

```
## Loading required package: tidyverse

## -- Attaching packages ------------------------------------------------- tidyverse 1.2.1 --

## √ ggplot2 2.2.1      √ purrr   0.2.4
## √ tibble  1.4.2      √ dplyr   0.7.5
## √ tidyr   0.8.1      √ stringr 1.3.1
## √ readr   1.1.1      √ forcats 0.3.0

## -- Conflicts ---------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## Loading required package: lubridate

##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date

## Loading required package: stackr

## Loading required package: jsonlite

##
## Attaching package: 'jsonlite'

## The following object is masked from 'package:purrr':
##
##     flatten

## Loading required package: dlstats

## Loading required package: printr
```

0.1   Categorization of Badges

I need to download all types of badges with their description and for further investigation to categorize them.

0.1.1 Badges for different activities

The badges are grouped according to the necessary activities to get them. The different types of badges can be seen on help page for badges:

1) Question badges
2) Answer badges
3) Partication badges
4) Moderation badges
5) Other badges
6) Documentation badges

0.1.2 Badges from different sources

Another differentiation is the point of origin the badges come from. Even though all badges are generated automatically by the system, there are two principal sources to get them. There are two different entities responsible for awarding badges.

1. There are badges you can get by just doing a certain action once or a needed number of times.
2. And there are badges you get only after some community action.

I call the first type `user` and the second type `community` badges.

Examples for user badges:

Name: "Civic Duty" (Vote 300 times). Name: "Enthusiast" (Visit the site each day for 30 consecutive day)

Some of the `user` badges are only available after some threshold value is reached and a certain privileges is granted. For instance voting up needs 15 and voting down 125 reputation points.

Examples for community badges:

Name: "Favorite Question" (Question favored by 25 users) Name: "Popular Question" (Question with 1000 views).

There are also badges tied with certain events, like moderator election, working with the beta version, meeting employees at an event etc. Mostly these badges are in my categorization `user` badges, but you cannot get them at any time, even if you would have the required privileges.

The badge system of Stack Overflow is a dynamic system. When the website was launched (Spetember 15, 2008) it did not start with all 91 badges which are listed today (May 21, 2018). Some of the badges were added during the history of the website but other are not awared anymore. Some of those badges which are not functional anymore are retired badges (e.g. "Analytical", for visiting all sections of the FAQ). This means you will still see some veteran users with these badges. Other badges (all three documentation badges for instance) are withdrawn, e.g. they were even deleted from those user accounts they were awarded earlier. I call these two groups of badges `event` and `dead` badges.

To sum up there are 4 point of origins for badges:

1) Badges from user action = `user` badges
2) Badges from community action = `community` badges
3) Badges from some action during a specified occurence = `event` badges
4) Badges where the point of origin is not interesting anymore as they are now non-funtional = `dead` badges

0.1.3 Groups of badges with different degrees of difficulty

Badges are ranked by their difficulty. According to their level of difficulty users are awared with bronze (relatively easy), silver (moderately difficult) or gold (difficult) badges. User can see those groups of badges

separated by lines on the Stack Overflow website. On this page one can also inspect how many time each badge was awarded.

There is also a special group of badges intimately linked with reputation of the user's expertise. These `Tag Badges` also come with different degrees of difficulty.

I will focus in my research on those badges where users are mainly responsible. Those are the badges where I can show if users are motivated to strive for them.

Visit the site each day for [...] consecutive days. (Days are counted in UTC.)

- 30 (Enthusiast)
- 100 (Fanatic)

Ask a well-received question [...] and maintain a positive question record

- on 5 separate days (Curious)
- on 30 separate days (Inquisitativ)
- on 100 separate days (Socratic)

Complete at least [...]. This badge is awarded once per review type

- one review task (Custodian)
- 250 review tasks (Reviewer)
- 1000 review tasks (Steward)

Edit and answer [...] (both actions within 12 hours, answer score > 0)

- 1 question (Explainer)
- 50 questions (Refiner)
- 500 questions (Illuminator)

Edit

- first post (Editor)
- 80 posts (Strunk & White)
- 500 posts (Copy Editor)

Edit [...] that was inactive for 6 months

- first post (Excavator)
- 100 posts (Archeologist)

The `user` badges are of utmost importance for my research. Most of these badges are granted for quality assurance work on the website and are not linked with reputation points. They are – from a systemic perspective – important for the dynamic of the website and maintain respective raise the qualitiy of the platform and their ecological sustainability.

When I can show that people are striving to get `user` badges – even when they are not linked with reputation points – than it is evident that badges are some additional (motivational) factors for the community development.

```
> badges <- readRDS("data-processed/badges.rds")
>
> ###############################################################################
> # Count the number of who badges and how many there are per type
>
> by_activity <- badges %>%
+         group_by(activity) %>%
+         summarise(n = n(),
+                   total = sum(award_count)) %>%
+         ungroup()
```

```
>
> my_table <- knitr::kable(by_activity, caption = "Badges in Stack Overflow per Type of Activity")
> cat(my_table)

Table: Badges in Stack Overflow per Type of Activity  activity          n       total -------------   --
> ###############################################################################
> # Count the number of badges types and how often they are awarded by their type
>
> by_origin <- badges %>%
+         group_by(origin) %>%
+         summarise(n = n(),
+                   total = sum(award_count)) %>%
+         ungroup()
>
>
> ###############################################################################
>
> # knitr::kable(badges, caption = "Type of Badges in Stack Overflow")
> # knitr::kable(by_activity, caption = "Main Types of Badges in Stack Overflow")
> # knitr::kable(by_origin, digits = 3, caption = "User and Community Badges in Stack Overflow")
```

0.2 Infos on Tags

What are those tags related with the r-tag?

0.3 Get info for specified user