Testing Strategy for Reddit

Joseph David Bender

Overview of Reddit

The "front page" of a newspaper has a certain connotation to it. When displayed on a newsstand or plopped on the kitchen table with a cup of coffee, the front page is what really matters. It contains the news that people absolutely need to know. The precious space on the front page is given to informative headlines, captivating images, and the most significant facts and opinions. Unfortunately for the newspaper, the internet was invented and news quickly became another digital commodity. Then in 2005, a pair of twenty-two year old graduates from the University of Virginia decided to create a website they named Reddit with the slogan, "The Front Page of the Internet". It was an entertainment, social networking, and news site powered by registered users who had the ability to post text, links, and comments on others' submissions. The site, which was virtually a public online bulletin board, took off thanks to millions of nerds, geeks, intellectuals, professionals, and artists flocking to create accounts. Before long, the massive publisher Conde Nast purchased the company for somewhere around $20 million dollars (Fiegerman). Since then, it has been nothing but exponential growth for the site. They gave users the ability to create their own "subreddits", or specific mini-communities focused on a single topic. Today, there are over 900,000 different subreddits for practically every subject the public could think of (Redditmetrics). President Barrack Obama has conducted interviews on the site. The Reddit community has been a significant influence on rallying support to defeat anti-privacy laws such as SOPA, PIPA, or CISPA. There was even a subreddit devoted to web vigilantes tracking down the architects of the Boston Marathon bombing. Basically, if it is important or interesting, it's on Reddit. True to their tagline, if the internet was a newspaper, Reddit would be the front page.

Despite being the 24th most visited site on the internet, based on the Alexa rankings, Reddit's design and functionality is fairly simple (Alexa). All that someone needs to use Reddit is a computer or mobile device, internet connection, and web browser. The web service can even be used without an account in a read-only mode. Any registered user may submit a post to a specific subreddit. There are two types of submissions. Text posts are comprised solely of text. The text can contain links to images, videos, and other sites, but the original post is read in its entirety and is hosted on Reddit. Link posts consist of an internal or external URL that takes the user to an image, video, gif, or website. Once a submission is live, other Redditors (the nickname for registered reddit users) can either upvote or downvote the post to influence it's visibility to other visitors to the site. The posts from each subreddit that are upvoted the most are pushed to the coveted front page and become the most visible submissions on the site. This results in generally the most intriguing, newsworthy, and significant posts are shown to the public. The algorithm that decides which posts are worthy of the front page is the core of Reddit's functionality and decides what millions of people browse every day. Each post also has its own comment page, on which people can reply to the original submission and start discussions and conversations. Comments can also be upvoted or downvoted, and the most popular comments are displayed first. For individual users, every registered Redditor has their own personal profile. It compiles all of that person's submissions and comments from around the site, and can be sorted by various fields including date or number of upvotes. Another driving mechanism of Reddit is "karma". Karma is the accumulation of upvotes, and the total is displayed on a Redditors personal profile. Typically, the higher a user's link karma and comment karma are, the more respected they are around the community. A running joke on the site is that karma is also known as "internet points" because in reality, they are worthless. However, the pursuit of karma

is what keeps Reddit thriving and active. One final important detail are the default subreddits. These are the sub-communities that a user is automatically subscribed to when they create an account. They are typically the largest and most subscribed subreddits on the site such as r/funny, r/worldnews, or r/music. The crux of having an account, is that the front page is only populated with posts from subreddits to which a user is subscribed. Therefore, as a Redditor becomes more active and subscribes to subreddits in which they are interested, the front page becomes more tailored and specialized. Overall, these features are the foundation of Reddit. Users try to post the "best" stuff, each post has a comment page for discussion, and each post also belongs to a relevant subreddit. This simple formula propelled Reddit into being the primary source of news and entertainment for millions of internet browsers.

Structure of Reddit

In the very early stages, Reddit was written in Common Lisp but was quickly adapted to Python in order to utilize a wider array of code libraries. In 2008, Reddit became an open source project, meaning all of their code and programming is public and available for viewing on a hosting site named Github. This transparency helps to counter that they are a private company, and aims to retain the community's trust. In 2009, Reddit decommissioned their physical servers and migrated to Amazon Web Services, a cloud computing and hosting company provided by Amazon. Reddit stores its archives using SQL databases. This is a column based system that stores all of the posts, comments, and other data that populates content on the site. Reddit also makes its API (application programming interface) public for use by external applications. This allows developers to create mobile, desktop, and web applications that access Reddit's content and offer modified or expanded functionality. One of the most prevalent uses of the API is a browser extension called RES, or the Reddit Enhancement Suite. It offers numerous features that

the basic site does not have, such as user tags, keyboard navigation, and a subreddit shortcut bar. There is also Reditr, which uses the API in a desktop application, removing the necessity for a web browser to access Reddit. Since Reddit is a free service, they monetize traffic by using banner ads, sidebar ads, and a service called "Reddit Gold". Gold is a designation a user can purchase or gift that adds a limited set of expanded features. These features include removing ads, access to secret subreddits, special comment highlighting, and more. However, a lot of these features can be achieved by other cost-free extensions, and Reddit Gold is more so bought with the intention of contributing/donating to Reddit.

There are many stakeholders that are invested in Reddit. The largest group are the users and visitors to the site. They want the site to remain active so they can continue to be entertained and informed by the content. The employees and administrators of Reddit and its affiliates are also invested in the company. They are being paid by the organization and have a financial obligation to maintain the web service. Reddit is also an immensely powerful promotion and marketing tool. The Slashdot effect is a phenomenon that occurs when a mega-popular site links to a smaller webpage and therefore increases its traffic exponentially. There are thousands of examples of Reddit helping to endorse other web pages on the internet. If a submission externally linking to a smaller website gains popularity on Reddit, that smaller website will inevitable receive increased visitation. For example, when a man was facing insurmountable costs to adopt a newborn son, he took to Reddit to sell his prized comic book collection. The post promoting his comic books achieved a massive amount of exposure. His online comic book store not only sold out, but most of the purchases were sent right back to him because his cause was noble (Morris). This generosity creates indirect stakeholders and allows Reddit to be quite the versatile tool. It has also been nicknamed the "Reddit Hug of Death". Many of the smaller sites

that blow up due to redirects from Reddit don't have the server capabilities to handle the exponential jump in traffic, then ultimately crash because they're overstressed.

Types of Tests

Every user will operate Reddit differently. Every person has their own unique interests and level of tech savviness. As a result, some Redditors will simply browse the front page of the default subreddits without having an account. However, others will fine-tune their subscriptions, add extensions, and push Reddit to its limits. Exploratory testing would be extremely useful for simply simulating what various people do on the website. Traveling to different subreddits, visiting various link, submitting/deleting links and comments, subscribing and unsubscribing from subreddits, and other features could all be investigated by exploratory testing. Reddit is a living breathing beast, and a rigid test plan could miss some important defects or shortcomings.

Reddit is definitely a segmented service and can be broken down into easily identifiable parts and functionalities. This is an indicator that unit testing could be effective in assuring quality across the site. The front page is far and away the most important aspect of the site. A vast majority of users will come to the front page first and foremost to browse. When a user looks at the front page, they want to see content from the subreddits they are subscribed to. A unit test would check to see what subreddits a user is subscribed to, and then test to ensure that all of those subreddits are represented somewhere on the front page. If, for some reason, a particular subreddit is missing from the front page after a significant number of posts, there could be a defect in the algorithm generating content.

Reddit is a web application hosted on the internet and accessible from any web browser. Therefore, web testing would be a viable way to ensure that functionality remains nominal. Web Application Performance Tools (WAPT) typically simulate virtual users and test expected

outcomes and performance usage. A WAPT such as Selenium would be a practical way to ensure Reddit does what the developers think it should do. To confirm that posting works correctly, Selenium could submit a text or link post, and then search Reddit for a keyword from the recently created post to guarantee that it is public and searchable on the web. Reddit also has a feature which prevents users from excessive posting. If a user has posted many posts in a row very quickly, then a time delay is activated which forces the user to wait a certain amount of time before posting again. The timer and limit are determined by the amount of karma a user has reaped from the subreddit to which they are attempting to post. A Selenium script could attempt to post numerous submissions in a subreddit and then test for the time delay feedback string to make sure that this prevention is functioning. This is important to deter spammers from using automation to troll Reddit with advertisements. This script could then be tested on various subreddits and on users with various levels of karma. In general, web testing with a WAPT is highly versatile and flexible, and can be used to examine virtually any functionality that Reddit has to offer. It doesn't necessarily test the backend code and programming, but is useful in assessing the resulting functionality that the developers are trying to accomplish.

Security is moderately important on Reddit. Granted, it is not an online banking interface or hospital database. If something of that magnitude was hacked a lot of money or lives could be at stake. However, Reddit has the power of the public's attention and in modern day that can be unfathomably valuable. Having all of their code be open source is beneficial for allowing developers to improve and modify it for external applications. Although, it could also be used for malicious reasons. Understanding how the backend of Reddit works would be critical in exploiting it. Any area the user can insert text should be an element of concern. Injection attacks are one of the most common ways of commandeering a website for malevolent purposes. The

search bar and submission test field should both disable the ability to use javascript to manipulate Reddit.

Performance testing could also give developers integral information on how to proceed with future development of the website. The most important key performance indicator (KPI) would be page load time. Reddit is all about content, and has a massive database and archive that stores all of the posts and comments that have ever been submitted. Populating the front page, or a subreddit's main "Hot" page, could take some time if the programming is not efficient and effectual. Using a timer to measure the load duration of various pages around the site would help guide improvements to Reddit. Another KPI would be average ad income per page. Since Reddit is a free service that operates on an ad driven model, the placement of advertisements is critical. Income is important for a private company, and much of the development of Reddit is powered by paid employees. Optimizing advertisements should be a priority. Is the front page the most effective place to host advertisements? Or are specific subreddits with targeted advertisements more advantageous?

Property testing is a great way to ensure that observed outcome matches expected outcome. A simple property test could use arithmetic to confirm the displayed karma total on a user's profile is the total of their upvoted posts and comments. A user's submission archive could be queried for the individual upvote counts. Once all of the upvote counts are obtained, they would be tallied and totaled. A property test would then compare this total to the displayed total. The test would fail if there is a discrepancy between the calculated total and displayed total.

Prioritization

The front page is the most important aspect of Reddit. There are a lot of users that don't have an account with the site but simply visit to browse the most popular posts of the day.

Therefore it should be given the most attention when it comes to quality assurance and development. The amount of time a post remains on the front page should be fine-tuned to be efficient and informative. If a post remains on the front page for too, long a frequent user may be exposed to it multiple times and feel that the content is not refreshed enough. If a post doesn't get enough time on the front page users who access the site less frequently may miss important submissions. The algorithm that controls front page visibility should be tested rigorously with all types of users.

The search box is also vital in accessing the material on Reddit. What would be the use of spending massive amounts of storage space on archives if there isn't an effective way to access them? A user should be able to effortlessly dive into the history of Reddit to find posts that match what they are searching for. Testing to ensure that search results are relevant and sorted properly should be another high priority for the developers.

In general, Reddit should follow an ice cream cone anti-pattern of testing. User interface tests should be the vast majority of the focus. The Reddit graphical user interface is the basis of its operation. Without a functional and efficient website, the desire to spend time on Reddit as a user would drop significantly. Rigorous testing of the front page, subreddits, user page, search bar, and other aspects of the site should be examined and improved as much as possible. Quality assurance should coincide with development. Continuous integration is a great way to ensure that the product is usable every step of the way. Taking a backseat to UI tests are service and unit tests. While also important, these are testing more of the back-end, behind the scenes factors. However, UI, service, and unit tests all working in conjunction with one another can result in an airtight user experience.

External Help

Reddit should absolutely employ external help to assist in software quality assurance. They are a small, private company that could easily become exceedingly incestuous and overlook the shortcomings of their website. Outsourcing their software testing to a third party organization would eliminate bias and increase their credibility. They could also look to employ some of the moderators of the default subreddits. This could increase accountability for the most popular communities, and bring in an external perspective. Turning power users into employees would help to bridge the gap between administration and public.

Management Structure

Reddit is a privately owned company, which means they have no stock available for trading. For example, when Conde Nast purchased them the exact figure was not released to the public. They simply declared that it was somewhere between ten and twenty million dollars. This can be dangerous because the people they are financially responsible to are internal and usually administrators. In a public company, there is a board of directors that governs the inner workings of the corporation. The whole basis of Reddit is freedom of speech, which slightly contradicts its management structure. One of the largest controversies surrounding Reddit arose in the summer of 2015. The CEO at the time, Ellen Pao, was in charge of attempting to rid the site of harassment. However, censorship is a sensitive matter on Reddit, due to the web service historically taking pride in its edgy content. Pao made the decision to ban five subreddit communities, such as r/FatPeopleHate, for promoting hate and discrimination. She soon found that banning material on the site is a slippery slope, and there was a public outcry asserting that it was unjust to remove some communities and not others. There was a petition started that received 200,000 signatures calling for her resignation. After much pressure, Ellen Pao eventually did decide to leave the company. This is just one example illustrating that it can be

dangerous to have such an open and community sustained website run by a private company. Diversifying the management structure and bringing in third party consultation could help to diffuse incidents such as this.

Reddit is also a relatively small company employing under 100 people (Reddit). The careers fall under a handful of categories such as engineering, business development, communications, human resources, legal, product, and sales. The entire company is based in San Francisco. They used to have satellite, remote offices in Salt Lake City and New York. However, at the end of 2014 they decided to shut them down and require all employees to relocate to San Francisco (Truong).

Quality Improvement Recommendations

There are many areas of concern pertaining to the operation of Reddit. It is such a massive site that prides itself on being customizable and promoting freedom of speech, that they are bound to run into some obstacles. One problematic aspect of Reddit is moderation. It was a brilliant idea to allow users to create their own subreddits. This permitted thousands of particular communities to spring up and foster incredible and specific content, discussion, and opinion. However, this begs the question of who is in charge of these mini communities? Moderation has always been a point of contention on the site. Power on the internet is a fickle business, and many moderators enjoy being in charge of massive groups of faceless web surfers. Moderators can also alter the CSS (Cascading Stylesheets) of a subreddit to alter the appearance. Much of the time this customization can hurt or break the functionality of a subreddit. Buttons can be covered up by banners, spacing and highlighting can become wonky, or colors can make content unreadable. Another way that moderators could exploit an unskilled user is by disguising malicious links. The sidebar of a subreddit is typically an informational pane that contains

relevant links relating to the topic of the community. A hacker or moderator with ill intentions could have a seemingly helpful link actually redirect to a url containing malware. A system of checks and balances for moderators would greatly improve accountability. Subscribers of a subreddit should be able to impeach moderators, call a vote or a new addition to the moderation team, or contribute to the modification of the stylesheet.

Reddit is also known to crash when it experiences increased user traffic. The witty and funny 404 screens are entertaining, but not as entertaining as the content that should be displaying instead of an error. This is simply a physical infrastructure problem. They should expand their hardware and server load to be able to deal with random spikes of activity. In addition, nearly all of the image posts on the website link to an external image hosting platform named Imgur. Occasionally, Imgur goes down and renders a majority of the link posts on Reddit useless. Having such dependence on an external source can be dangerous. Recently, Reddit launched a new first-party hosting option where the user uploads the image directly to Reddit. The company should devote resources to improving and promoting this new method to decrease dependency on external hosts.

The search functionality could also be greatly improved. As it stands, there are no predictive suggestions and no spell check. Google search has become incredibly intelligent in guessing what a user is searching for. It corrects slight misspellings, and provides feedback on the change at the top of the results page. It also has intuitive predictive search offering numerous suggestions as a searcher types their query. Reddit could employ some of this functionality to make combing through the archives a little more user-friendly.

Conclusion

Reddit is an extremely unique web service that has amassed an enormous following in the last decade. It has transcended its home on the internet and had significant repercussions in the real world. It is a community that prides itself on freedom of speech, embracing the avant-garde, and celebrating the human condition. There is a subreddit for all walks of life, and virtual friends can be found everywhere. Although it is managed by a private company, the foundation of Reddit could not be more public based. Since it receives such a large amount of traffic, a lot of effort has been put in to making its functionality efficient and airtight. However, there isn't a website in existence that doesn't have room for improvement. Reddit has numerous aspects that can be customized or modified, and this can lead to both benefits and vulnerabilities. Comprehensive testing would only benefit the web service, and should occur in tandem with development for optimal operating capacity. Reddit has worked hard to earn a spot in the top 50 frequented sites on the web, and it isn't going anywhere. With a diversified plan of quality assurance, the future of Reddit will be successful and constructive for its users. Upon waking up, everyone should pour a bowl of cereal, grab a cup of coffee, and enjoy the front page...of the internet.

Works Cited

Alexa. The top 500 sites on the web.   Retrieved from http://www.alexa.com/topsites

Contributors, R. (2016). About Reddit.   Retrieved from https://about.reddit.com/

Fiegerman, S. (2014). Aliens in the Valley: The complete history of Reddit, the Internet's front
        page. Retrieved from http://mashable.com/2014/12/03/history-of-reddit/#PsrjcCUGOkqw

Metrics, R. New Subreddits by Date.   Retrieved from http://redditmetrics.com/history

Morris, K. Reddit rallies behind comic artist's adoption struggle. Retrieved from Daily Dot
        website: http://www.dailydot.com/society/karl-kesel-comics-adoption-crowdtilt/

Truong, A. Reddit Gives Remote Employees Until End Of Year To Relocate To San Francisco.
        Retrieved from Fast Company website: http://www.fastcompany.com/3036660/fast-
        feed/reddit-gives-remote-employees-until-end-of-year-to-relocate-to-san-francisco