Channel Islands

Reddit Analytics Engine (CI RAE)

Nonfunctional Requirements

Phillip Porter, Preston Peterson, Zoltan Batoczki,

Mattias Huber, Heather Bradfield, Jesus Banford

Prevent Injection Attacks Requirement

**Category**: Security

**Purpose**: This requirement is needed to deter malicious agents from running injection code on our bot through reddit. An injection attack is a computer bug that allows for the processing of incorrect data which could be manipulated to make the program respond in unintended ways. An example would be an SQL injection attack in which the user would inject an SQL query into the bot’s input which would result in the database dumping the user table with usernames and passwords.

**Goal**: The goal of this requirement would be the software being immune to malformed input. If a user was detected trying to inject code into the bot, ban the user from interacting with the bot.

**Testable**: Attempt injection attacks by inputting malformed input, and respond by detecting injection attacks and banning the user.

**Validity**: The system will still be able to respond to the need of the user, just restrict what the user can do to not allow them to cause havoc.

**Consistency**: This requirement may conflict with the speed requirement because in scanning for injection attacks with have some overhead.

**Completeness**: All of the functions required by the user are still allowed, only incorrect input is disallowed.

**Realism**: This requirement could be simply implemented with just whitelisting only known good values, or could be more robust by modularizing and moving data out of an input string into data structures.

**Verifiability**: This requirement can be verified by testing the bot with injection attacks, and to detect if users are banned if injection attacks are attempted.

**Comprehensibility**: The requirement is clearly understood as a security vulnerability that must be handled appropriately.

**Traceability**: This requirement implementation can be documented to track the strategy of preventing injection attacks. This is in case there is a change of strategy responding to escalating attacks there will be documentation of strategies already implemented. This requirement deals with parsing input, and whenever there is input taken in by the system then the manipulation of the input must be documented so that the resulting format of the input is understood.

**Adaptability**: This requirement may be scalable based on the amount of threat to attacks the system experiences. However, increasing security features will result in increased overhead which would affect the speed requirement.

API Guidelines Requirement:

**Category**: External

**Purpose**: This requirement is needed so that we comply with the Reddit’s and Imgur’s API terms. Our entire bot is based around retrieving information from Reddit API and uploading comment reply images using the Imgur API. By conforming to their terms, we will be able to continue to use services provided to us through the API.

**Goal**: The goal of this requirement is to not be banned from using Imgur’s or Reddit’s API. In order to not be banned, we must implement their rules into our bot.

**Testable**: We are able to test to see that our bot conforms to the API guideline requirements using various methods in the official Reddit and Imgur API library. For example, the Reddit API library contains headers that allow us to track how many requests we’ve made, in order to prevent us from going over the limit.

**Consistency**: The API Guideline requirements directly conflict with speed requirements. Under heavy-load our bot will need to purposely slow down and queue requests in order to conform to the API’s request limit.

**Verifiability**: Both Reddit and Imgur have a development webpage with the clearly stated terms of service for API access. Therefore we are able to verify that our bot conforms to their requirements.

**Comprehensibility**: The requirements are clearly stated on the Reddit and Imgur development webpage.

**Traceability**: Documentation will provide reasoning as to why we provide data extraction limits, purposefully slow down our bot, or any other methods that would otherwise seem sub-optimal in order to conform to the API requirements.

**Adaptability**: Our bot heavily depends on the official API python libraries to extract data from the API. If the API guidelines change, it should be reflected upon the official library, which should allow us to adapt easily by updating to the latest version. If our bot uses a deprecated API, object, or method, the program will need to be rewritten depending on how much our bot relies on the specific deprecated entity.

Protect Supporting System Requirement

**Category**: Product requirements

**Purpose**: To ensure that the bot does not compromise the system running it. In addition, all related errors must be handled by the bot to ensure stability.

**Goal**: Successful implementation of server stability, bot stability, user error handling, API error and timeout handling.

**Testable**: Unit tests may verify each of these requirements by purposefully introducing bad input or submodule failure. The ability for the machine to survive the volatility of the testing environment is proof of successful implementation.

**Validity**: This feature serves the customer as a bot taken down by instability in design is of little use to the customer. The equity in an investment in the product is proportional to compliance with this requirement.

**Consistency**: This requirement conflicts with the speed requirement as the protection of system resources involves rationing system resources to the bot. The speed requirement is best addressed by allocating more system resources at it.

**Completeness**: This feature is a standard product feature as it involves managing sub features meant only for the programmer. Failure of such features is not of relevant concern for customers.

**Verifiability**: These requirements can be checked by examining historical evidence, standardization bodies, and other related products for design standards.

**Comprehensibility**: This requirement is comprehensive because the expectation of failure is made clearly real by real example and theory regarding dysfunction.

**Traceability**: This requirement was brought to the attention of the development team by unit testing that produced erroneous results due to exceptions or resource overuse. Application of this requirement is complete throughout all modules and classes. Satisfaction of this requirement is universally expected in all of the project.

**Adaptability**: This requirement may be adapted using by modifying error handling behavior as additional errors are specified by changing submodules. In addition, the old behavior of error handling functions can possibly apply to new errors given similar overall problems.

Speed Requirement

**Category**: Efficiency

**Purpose**: The nature of the modern internet has conditioned users to be very impatient in their day-to-day internet activity. Furthermore, when used for internal analytics, the Reddit Analytics Engine must sometimes process very large quantities of information. If this information cannot be analyzed quickly, it will not be useful in any practical sense. Thus, it is not enough for the program to work. It must work quickly.

**Goal**: The goal of the speed requirement is to ensure that the engine executes in a minimal amount of time.

**Testable**: The speed requirement is testable because it is directly measurable by the amount of time the engine takes to process a request. It is also non-functional because the primary purpose of the engine is data analytics, and given a static array of data the concept of data analytics does not depend on time or speed.

**Validity**: On Reddit.com, threads (comment topics) rarely remain active for more than 24 hours. This means that an individual’s interest or involvement in any particular thread is a highly temporal and volatile thing. Also, because comments are the medium of communication between users and CI RAE, and comments exist inside of threads, the system will best support the customer’s needs if it can respond to a user’s request comment in as little time as possible - ideally within one minute.

**Consistency**: The speed requirement will not hinder any other requirements - although we may implement a maximum processing time for a given user request (a time that should never be reached in theory). However, there may be analyses and API calls that could hinder the program speed. Of course, this is unavoidable. After all, he fastest program in the world is one that does nothing.

**Completeness**: Speed is a function that will be included in the product. However, speed is not a binary thing, it is a spectrum. The Reddit Analytics Engine will always seek to minimize request processing time.

**Realism**: Speed is a realistic goal. Using Python 3’s robust threading libraries, the Reddit Analytics Engine will utilize multithreading to its full potential so that it is capable of both processing multiple requests simultaneously. Additionally, rather than scanning or scraping all of Reddit to find user requests, it will use Reddit’s Mentions feature in order to instantaneously receive user requests from any corner of Reddit at any given time. This combination of quick request retrieval and fast analytics processing means that the requirement of speed is very realistic and attainable.

**Verifiability**: The requirement of speed is verifiable. We can keep track of the time it takes to retrieve a user request from Reddit’s servers, process the request and the associated analytics, and submit a response back to Reddit’s servers. We can also keep track of how much time each analysis function takes, in order to figure out where speed bottlenecks exist in the engine.

**Comprehensibility**: This requirement is more-or-less self explanatory. Speedy execution is of the utmost importance, so long as it does not come at the cost of correct execution.

**Traceability**: The requirement of speed originates from a need to respond the user within a reasonable window of time, and for the need of analytics to be generated quickly so as to be usable in a functional way. The requirement of speed leads to the implementation of certain program logic that would not otherwise be necessary such as multithreading.

**Adaptability**: This requirement can be changed, but if we implement a maximum limit on the execution time of a given process, other requirements could be affected.

Operational Requirement

**Category**: Organizational

**Purpose**: Defines how the system will be used. Most interactions on Reddit is done through a comment section of a submission thread. Using the comment section, users will be able to submit a comment following a specific format in order to call in the CI\_RAE reddit bot. The original purpose of the bot is to allow users to display information regarding their activity on reddit.

**Goal**: Users will be able to issue multiple commands with the bot at any time in any subreddit.

**Testable**: The bot will respond to the comment with the requested information. For example, a user comment of “/u/ci\_rae help” would result a reply comment from the bot with a list of commands that the user can submit.

**Validation**:With the flexibility of the bot, functions that may not be available to a user now can be added into further enhance a user’s experience with the bot.

**Consistency**:Based on which function a user requests and how many users are requesting to the bot, a user may experience a late response from the bot, which the speed requirement discusses.

**Completeness** Since no actual users have requested functions, the bot will come with a default set of functions. Once users begin to use a bot, they can suggest a function and can be added if it meets [Reddit’s API Terms of Use](https://docs.google.com/forms/d/e/1FAIpQLSezNdDNK1-P8mspSbmtC2r86Ee9ZRbC66u929cG2GX0T9UMyw/viewform).

**Realism**: Since reddit’s API is based around Python, it allows us to make use of Python libraries to make unique features such as a word cloud or a graph with relative ease. Again, requirements can be met as long as it meets [Reddit’s API Terms of Use](https://docs.google.com/forms/d/e/1FAIpQLSezNdDNK1-P8mspSbmtC2r86Ee9ZRbC66u929cG2GX0T9UMyw/viewform).

**Verifiability**: Requirements can be checked without the use of public threads on reddit by using a private forum (known as “subreddit”). With it, developers can create test submissions to post comments that can test the bot’s functionalities.

**Comprehensibility** - Given simple documentation, users will see that using the bot is a simple comment submission following a specific format:

/u/ci\_rae <command> <(optional) parameter>

**Adaptability** - Each function the bot can run is independent from each other. Errors in function will not affect another function. However, depending on future functions that could be added in, it might affect the speed at which the bot cn process that request and reply back to the user.

Scalability Requirement

**Category**: Development

**Purpose**: This requirement is needed to ensure that the Reddit Analytics Engine is capable of handling a growing amount of Reddit users calling our bot, or has the potential to be enlarged in order to accommodate that growth.

**Goal**: The goal of the scalability requirement is to ensure that the engine can handle being used by a large number of Reddit users and function as intended.

**Testable**: Load testing, putting demand on the Reddit Analytics Engine and measuring its response, will determine its behavior under both normal and anticipated peak load conditions. It will identify the maximum operating capacity of the engine as well as any bottlenecks and determine which elements are causing degradation. The ability for RAE to survive a large amount of successive calls from simulated Reddit users within the testing environment is proof of successful implementation.

**Validity**:Together with the “Protect Supporting System” and the “Speed” Requirements, the system should be able to handle and respond to the needs of all 234 million unique Reddit users at once (the rarest of cases).

**Consistency**: Based on which function a user requests and how many users are requesting to the bot, a user may experience a late response from the bot, which the speed requirement discusses.

**Completeness**: Scalability is a function that will be included in the product. The Reddit Analytics Engine will always seek to minimize request processing time, ensure server stability, and expand storage space.

**Realism**: Scalability is a realistic goal. The Reddit Analytics Engine will utilize multithreading so that it is capable of processing multiple requests simultaneously. Additionally, rather than scanning or scraping all of Reddit to find user requests, it will use Reddit’s Mentions feature in order to instantaneously receive user requests from any corner of Reddit at any given time. Also, by default, our SQL Server automatically expands the database according to the growth parameters defined when the database was created.

**Verifiability**: This requirement can be verified by load/stress testing RAE and detecting whether or not the system crashes when a large amount of simulated Reddit users use bot.

**Comprehensibility**: This requirement is clearly understood. Speedy execution, server stability, and a large amount of available storage are of utmost importance when dealing with a large, growing internet community.

**Traceability**: The requirement of scalability originates from the growing number people who know about our bot as well as the growing number of Reddit users in general. RAE needs to be able to handle a lot of traffic, to respond to a large number of user within a reasonable window of time, to generate analytics quickly, and in general not crash under too much stress.

**Adaptability**: This requirement can be changed, but if we implement a maximum limit on the execution time of a given process, other requirements could be affected (as mentioned in the Speed Requirement).