

Red – Input to module
Green – Output from module

External

Reddit

Retrieve information from reddit.com by using their API, or by fetching HTML directly.

TextMining

Tool to mine features **F** from text **T**, name **N**.

RedditPostRankDiscoverer

Receives a particular Reddit post with URL **U**, and returns **N**, the rank of the post in the Subreddit .

RedditPostMiner

Mines a particular Subreddit **S** for new posts with age < **M** every **M** minutes. Outputs the text **T** of a post, the name **N** of a post, the URL **U** of the post, and the **K** karma of OP.

Web Layer

RedditPostPopularityPredictorControlUnit

Handles execution of the system. Determines where to route information. Commands what module to run and when, along with necessary callbacks. Starts all other python modules. Reads a central configuration file and passes parameters to modules. Config file includes ports to connect to for each module, how long to wait before mining Reddit, how long to wait before checking post rank, neural network parameters, Subreddit to mine, database credentials, time since last mine, time since last post rank check. Does logging.

Control Layer

TrainingArchive (MySQL DB)

Stores a post with URL **U**, features **F**, prediction **P**, with a timestamp **T** and an UUID. At request, it delivers **U**, **UUID**, according to **T**. After receiving actual rank **R** for UUID, it moves the record into another table, but with another UUID, timestamp, R, and a correctness Boolean.

Data Layer

NeuralNetwork

Receives **F**, a vector of features as input to the neural network. It predicts category/rank **N** of the post from the features. Alternatively, it can receive **F** and the correct output **C**, for training.

FeatureExtractor

Receives text **T** of a post, the name **N** of a post, and the **K** karma of OP. Extract a vector of features **F** as input for the neural network.

Learning Layer

Red – Request to module
Green – Response from module

Yellow circle – Separate thread on the control unit. Some operations depend on each other; some don't. Computations are grouped into threads by their mutual dependency.

