



reddit



Amanpreet Singh & Anirudh Ganesh

Reddit











- Reddit is a social content sharing and discussion platform which was founded by Steve Huffman and Alexis Ohanian in 2005 now worth \$1.8 billion.
- Users submit links, discussion topics, images etc which earn reputation or “Karma” and are upvoted or downvoted by other users.
- Divided into communities that share common interests called subreddits. ‘r/history’ or ‘r/machinelearning’ or ‘r/Keratoconus’
- 542 million monthly visitors, 73.15 million submissions, 725.85 million comments, over 1.1 million subreddits, 5 million comments daily.

Interface

MY SUBREDDITS ▼ HOME - POPULAR - ALL - RANDOM - USERS | ASKREDDIT - FUNNY - TODAYILEARNED - WORLDNEWS - PICS - AWW - VIDEOS - SOCCER - GIFS - MOVIES - INTERESTINGASFUCK - MILDLYINTERESTING - TELEVISION

  **hot** new rising controversial top gilded wiki

↑ trending subreddits r/Emuwarflashbacks r/Unity3D r/blackfriday r/PictureGame r/FunnyandSad 14 comments

- ↑ 47.3k ↓  **Backpack seller answers a question** (i.imgur.com)
submitted 4 hours ago by haemaker to r/funny
947 comments share save hide report crosspost
- ↑ 16.2k ↓  **Bringing down a couch the proper way** (gfyca.com)
submitted 4 hours ago by DSR to r/gifs
520 comments share save hide report crosspost
- ↑ 22.9k ↓  **"It's only when we wake up that we realize something was actually strange" Digital, 3800x3800** Artwork (i.imgur.com)
submitted 4 hours ago by buttscracha to r/Art
602 comments share save hide report crosspost
- ↑ 13.9k ↓  **Preparing a deli sandwich** gifv (i.imgur.com)
submitted 4 hours ago by JF_112 to r/woahdude
371 comments share save hide report crosspost
- ↑ 12.4k ↓  **Machine Learning** (i.redd.it)
submitted 5 hours ago by imeriem to r/ProgrammerHumor
104 comments share save hide report crosspost
- ↑ 17.8k ↓  **Cat Hat** (i.redd.it)
submitted 5 hours ago by Subterfug3 to r/aww
253 comments share save hide report crosspost
- ↑ 15.6k ↓  **Fake News Is Only the Beginning. The FCC Votes to Let Monopolies Decide What Local News You See** Politics (commondreams)
submitted 6 hours ago by LoyalAsFuck to r/technology
1013 comments share save hide report crosspost
- ↑ 7096 ↓  **Marvel's 'Ant-Man and the Wasp' Has Officially Wrapped Production** (cbr.com)
submitted 4 hours ago by BunyipPouch to r/movies
541 comments share save hide report crosspost
- ↑ 27.4k ↓  **As far as the eye can see. Dyrholaey Iceland. [3601x4646][OC]** (i.redd.it)
submitted 6 hours ago by Slam-Cannon to r/EarthPorn
763 comments share save hide report crosspost
- ↑ 3830 ↓  **Billions of dollars stolen every year in the U.S. (from Wage Theft vs. Other Types of Theft) [OC]** Based on 3 cities,... (i.redd.it)
submitted 2 hours ago by palmfranz to r/datavisbeautiful
303 comments share save hide report crosspost

Community

- The large community of “Redditors” supports interesting discussions on wide-ranging topics and ideas.
- Power of this community can be seen in the “Reddit Effect” also called the “Reddit Hug Of Death” or in “Mr. Splashy Pants”.
- r/fatpeoplehate and several other controversial subreddits are widespread.
- The same community is also seen engaging in anonymous altruism in the form of collective donations, talking people out of suicides, re-uniting them with parents etc.

Reddit for Developers

- Reddit was fully open source up until September 2017.
- Nonetheless, the Reddit API is extremely robust making it possible for so many Reddit clients to be user's first choice for browsing Reddit content.
- OAuth2 based API with several wrappers available in a multitude of languages.
- High use of object orientation.
`thing` is the Reddit base class.
A `thing` (say a `comment`) may implement `votable` or `created`.
- PRAW, a Python wrapper for Reddit is the most widely used wrapper, especially to create Reddit Bots.

Bots

- Reddit API support makes it possible for users to create Reddit bots which are bot accounts that have specific triggers associated with them.
- Some are helpful like u/MetricConversionBot which converts any units mentioned in comments.
- Some are silly like u/PirateCaptainSparrow which just corrects any mention of 'Jack Sparrow' to '*Captain* Jack Sparrow.

Bots

u/ColorizeBot

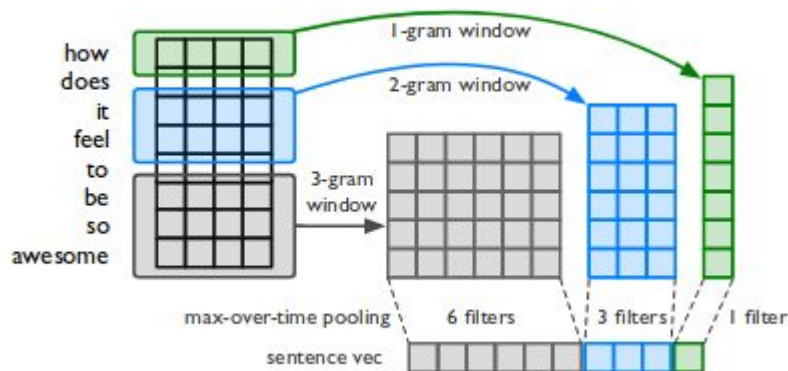


Source:
<http://richzhang.github.io/colorization/>

A Study of Question Effectiveness Using Reddit “Ask Me Anything” Threads

- Reddit AMA is a more candid and open interview where a person, generally a famous personality answers all sorts of random questions from users.
- Since this is the internet, and reddit users are generally anonymous, it can result in some hard hitting questions that the interviewee fails to address deliberately.
- Paper aims to determine what an “answerable” question is.
- They form a reddit question answer corpus in which they have collected the top 1500 posts per AMA.
- They train a specialized model of CNN called context-aware CNN with specialized POS tags to learn answerable questions.
- The model's effectiveness was evaluated by comparing its agreeability with human test subjects and it was found that there is a high agreeability with the human test subjects.
- However it was also found that both human and the model had lower than expected agreeability with the AMA host himself. This is largely due to AMA host prioritizing certain questions over others for promotional reasons or contractual obligations.

A Study of Question Effectiveness Using Reddit “Ask Me Anything” Threads



System	Mean	Min	Max
Logistic Regression	0.512	0.512	0.512
Baseline CNN	0.516	0.509	0.525
Context-aware CNN ($x=5$)	0.523	0.503	0.537

Table 3: Results on 100k test set as evaluated by AUC scores.

	2009	2010	2011	2012	2013	2014	2015	2016	Total/Avg
# of threads	5,055	11,987	21,381	11,555	7,517	5,514	4,054	660	67,723
Avg # of posts per thread	87.55	97.04	94.12	214.08	308.35	395.46	338.98	447.02	247.83
Avg # of sentences per post	3.19	3.03	2.87	2.79	2.62	2.57	2.65	2.65	2.80
Avg # of words per post	49.82	46.44	43.49	41.98	38.07	37.02	39.02	40.37	42.03
# of question posts	697	2689	27,835	90,522	109,969	109,569	62,874	16,064	420,219
% of question posts w/ answer	44.76	30.83	30.03	20.48	19.66	19.77	22.02	18.56	25.76

Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach

- Mental health is a niche and poorly understood domain that a lot of people overlook.
- The large amounts of progress in the last few decades have help alleviate some of the social stigma associated with consulting for mental health, however despite this a lot of people are hesitant to get any sort of consultation.
- Because of the anonymity of reddit users, people are a lot more forthcoming with their mental health problems, which has resulted in an active and thriving community that helps people openly discuss and share possible solution and offer some limited consultation.
- This paper analyzes the trends and patterns that exist in these communities by taking in a large corpus of posts from Anxiety, Depression and PTSD subreddits.
- They discover through clustering that all communities are overall positive PTSD and Anxiety communities focused more on treatment and help motivate users while Depression community generally leaned towards discussing their scenario.

General Trends in Research: Reddit

- Overall a lot of research that uses Reddit uses it mainly as a platform (for eg, bot)
- The research that focuses on Reddit as a social media generally involves Reddit as a corpus for conversational data to test out their existing techniques.
- This type of research uses the threaded structure and generally context sensitive conversation among users to better model NLP systems that take advantage of such data, like question answer system.
- The other kind of research trend that focuses on Reddit explores primarily the actions of Reddit's niche communities.
- Research involving reddit's communities takes advantage of a high retention rate of the user-base to derive trends in such communities which can be generalized.
- Another benefit of analyzing such communities is that the top posts have an intrinsic property of being extremely appealing to the community. This information can be used to create a more robust qualitative assessment models for tasks like sentiment analysis.



Questions?