
# SOCIAL NETWORK SEARCH

Michael Huben -15396501
Josh Malone - 15357971
Matthew Farrelly - 15366246
Maksims Kompanijecs - 15309671


GoSocial

# GoSocial

Universal search engine across multiple social media platforms (Facebook, Instagram, Twitter, Reddit)

Users given an option on to search for:

- Images, text
- blogs, articles
- trends, hashtags

GoSocial scans social media networks based on the search query

Returns results based on most relevant, popular and up-to-date results

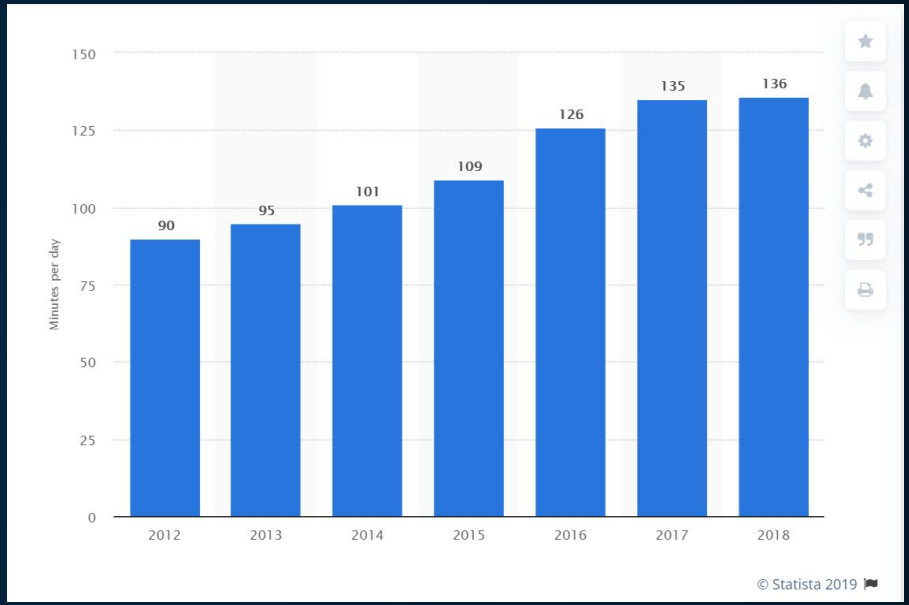
# Users & Motivation

Users spend a lot of time browsing multiple different social media platforms

Most of the content is the same - news, current events, sports, memes

A universal platform will save users time and allow them to consume relevant content more efficiently

E.g. A user interested in a general election can search and find the most popular/relevant articles, posts & media relating to it




Average amount spent on social media in the recent years

# User Analysis

User demographic will be of ages 18 - 29 as they are the most active social media users

Expected user knowledge:

- Non-specialised with no experience in search technologies
- Proficient in general search & social media usage

Expected type of search queries:

- Informational queries
- Multimedia search

# Problems with Social Media Content

**Content variety and style -** information on social media has different format, length, quality, language

**Poor quality content** - content with typos, poor choice of words

**Misinformation** - deliberate spam, to attract attention

**Duplicate content** - very similar or duplicate information published on the same site

**Inaccessible content** - private, not accessible content

# Functional Description - Web Crawler

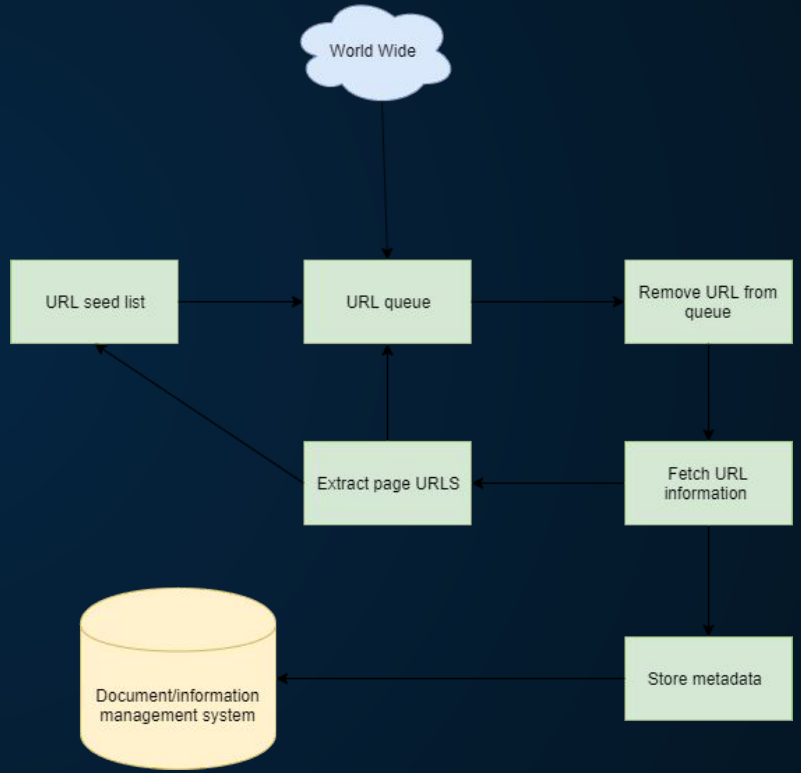
Navigate through the web and collect HTML page data/information on multiple social media platforms

It is important for web crawler to follow a policy to:

- How to know what pages are relevant
- To know when to check back again the same pages
- Be more sensitive about what pages it should crawl and in what order

And respect robots.txt file

- Don’t degrade website performance
- An identify website being crawled with contact information



# Functional Description - Document Indexing

Efficient information retrieval will be achieved by document indexing

This feature allows search engines to access the data easily that it requires

We will index documents by:

- Stopwords needs to be removed such as ‘and’,’in’ etc
- Break up sentences/strings/text into tokens - tokenization
- Transformation of words to their root form - stemming
- Normalization - transformation of text into its simplest form

# Functional Description - Ranking

Last component of our system will be the ordering of search results by most relevant to least relevant based on the search query

We will most likely opt in to implement PageRank algorithm

- It’s more feasible as it computes rank score during indexing rather then query time
- Returns relevant content as the rank is calculated based on popularity of the page

# Evaluation

To accurately evaluate our technology, we must calculate both precision & recall values

To fine-tune our search technology at the beginning of its life-cycle, we intend on implementing test collections

These collections will begin with standard ‘b’ and ‘k’ values

The product is heavily user-focused, so their values and preferences will be taken into account often

# Conclusion

Search Technology USP:

- Universal one-stop social media browser
- Ability to search popular social media artifacts (text, multimedia, hashtags)

Target demographic:

- Active social media users
- No specialised skill-set or queries
- General social media & search proficiency

GoSocial Function:

- Web crawler that parses the data from social media sites
- Documents indexed & ranked
- Results displayed to end user


GoSocial

# Any Questions?