CSN-341 Computer Networks
Assignment 3
(Search Engines and Architecture of Social Media Network)

Group 20

Aniket Singh (16116005)
Harshal Mittal (16116020)
Harshit Bansal (16116021)
Samriddh Bhalla (16116057)
Vishal Sharma (16116076)

Q1) What is a search engine? How does is work?

A web search engine is a software system that is designed to search for information on the World Wide Web. It uses keywords to search for documents that relate to these keywords then puts the results in order of relevance to the topic that it was searched for. They filter the vast amount of information available on the internet and transform it into results(*in the form of a search engine results page (SERP)*) that each individual can easily access within a matter of seconds.

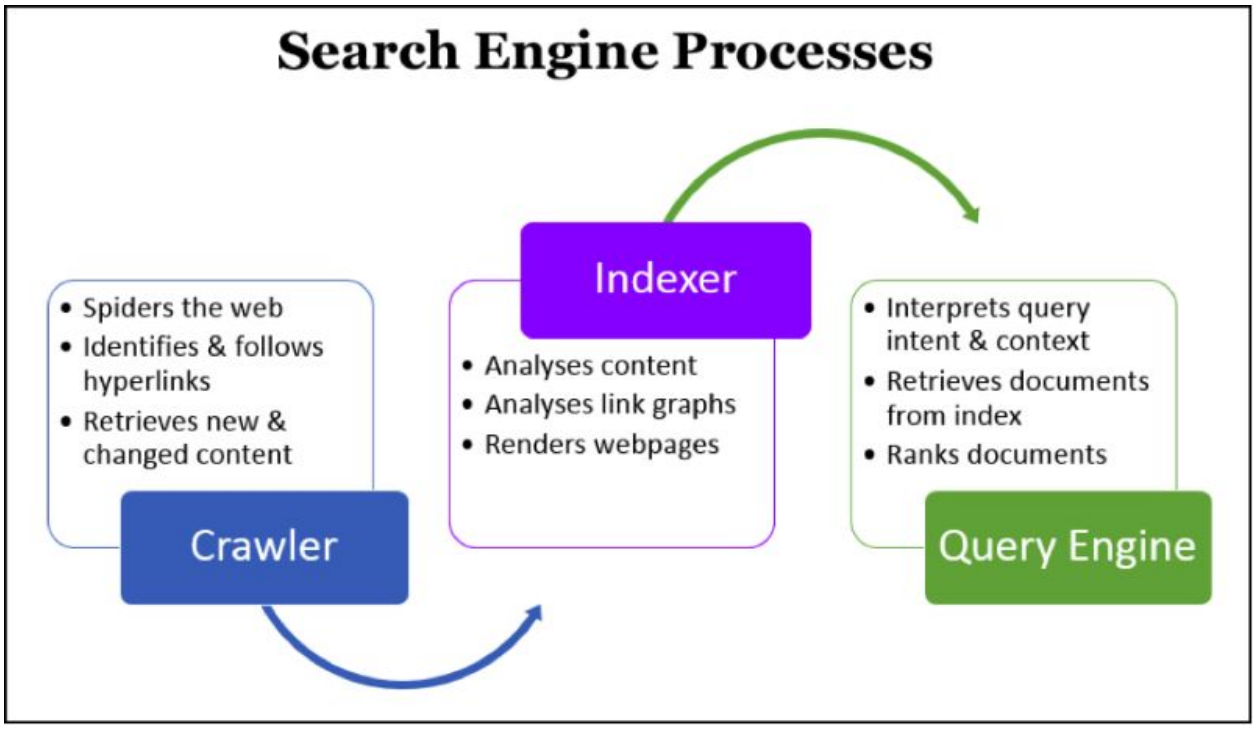
There are a variety of search engines present nowadays like Google, Bing, Yahoo, etc. Google is the dominating one of all the current search engines. The search engines can generate results in a variety of ways some of them being through the use of web crawlers, directories etc. I will be discussing the one that is most commonly used - crawler based search engine.

The basic working can be divided into 3 parts

1) Crawling - This stage involves scanning the sites and obtaining information about everything that is contained there: page title, keywords, layout, pages that it links to – at a bare minimum. This is done through “crawlers” or “spiders”. These robots access all the

popular websites first and then follow the links to find new interconnected documents. They keep returning to the original pages again and again to constantly update themselves.

2) Indexing - Once the data has been assimilated, selected pieces of it are stored in databases.
3) Ranking and Retrieval - Whenever we perform an online search, the search engines scour its database for the most relevant results. It also ranks them in order of relevance. The ranking order is different for different search engines. The ranks are given to these results based on their scores which is assigned as a sum of weights of some selected arguments.



Q2) Discuss about the working model of the search engine.

Every search engine has 3 main functionalities :

- Crawling(to discover content) : An automated bot (called a “spider”) visits page after page as quickly as possible, using page links to find

where to go next. When a web crawler visits a page, it collects every link on the page and adds them to its list of next pages to visit. It goes to the next page in its list, collects the links on that page, and repeats. Web crawlers also revisit past pages once in a while to see if any changes happened. Sites are crawled with different frequencies and depths.

- Indexing(to track and store content) : The processing and placing of data in the database is called indexing.
- Retrieval(to fetch relevant content when it is demanded by the user) : The processing of user's query by the search engine and returning of the most relevant pages by the search engine is retrieval. Ranking algorithms the search query against billions of pages to determine each one's relevance.

Following are the steps that are performed by the search engine:

1) The search engine looks for the keyword in the predefined database instead of directly searching the web for the keyword.
2) Web crawler is used to search for information in the database.
3) The pages found by the web crawler are shown as results by the search engine as results.
4) User may select any of the search results to open it.

Q3) Discuss the general architecture of social media network.

A Social Media Network can be represented as a graph where each node can be represented by a single person and the edges describe his/her relationship with other beings. A natural method of growth of a social network graph is when friends of friends are added to the graph.

There are four commonalities unique to current social networking services:

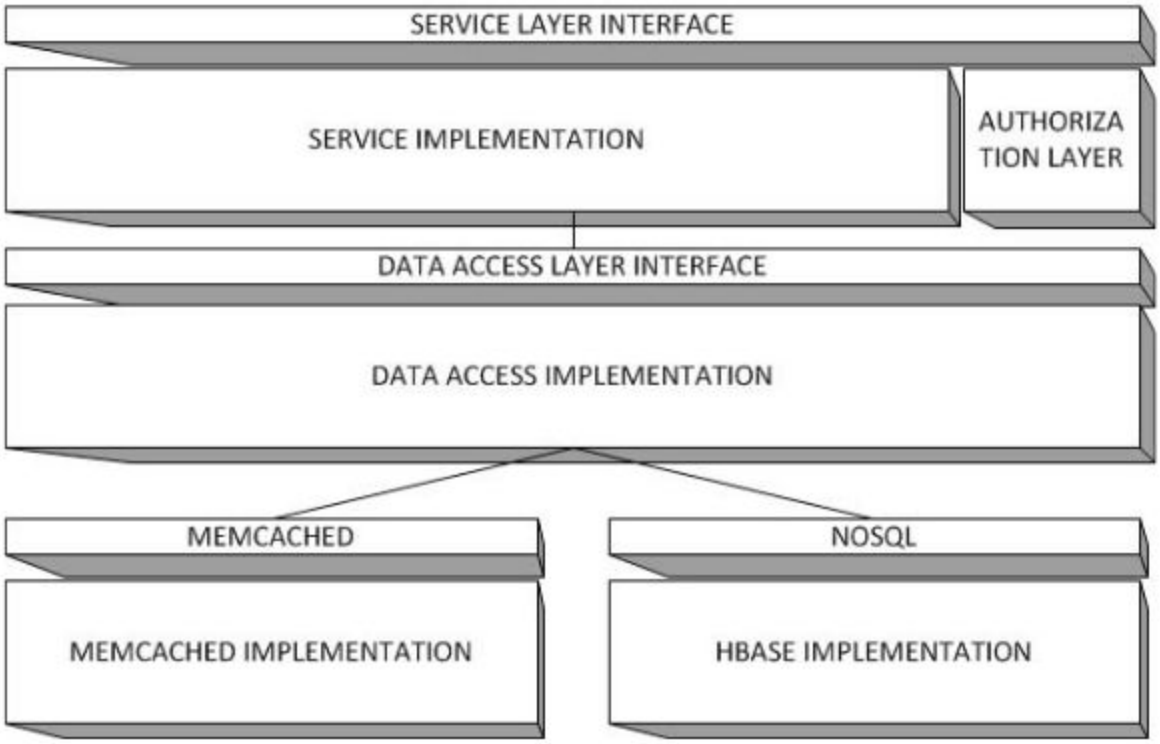
- Social networking services are interactive Web 2.0 Internet-based applications
- User-generated content (UGC), such as user-submitted digital photos, text posts, "tagging", online comments, and diary-style "web logs" (blogs), is the lifeblood of the SNS organism,
- Users create service-specific profiles for the site or app that are designed and maintained by the SNS organization, and
- Social networking services facilitate the development of social networks online by connecting a user's profile with those of other individuals or groups.

Network Architecture:

The system has three basic layers connected by interfaces.Service layer can call data access layer that access the storage using the data transmission layer. Service layer serves as the interface for direct or remote callings from the application. It processes data and authorizes operations. Data access layer fetches or stores dat a from/in database and transforms it into business objects and back The last layer consists of basic operations like add, delete, get, replace on database or cache and creates a simple framework over any storage.

There are four basic services currently in the application:

1. Follow service is responsible for creating and managing relations. It also enables to get data about followers and following (complete or random list which can be paginated).
2. Wall service performs actions related to posts and walls. It provides methods for creating, deleting and viewing either the entire post or only minimized one (for example on wall we do not want to show all comments connected with the post).
3. Like service allows entities to like and unlike the posts.
4. Discussion service manages the comments of specific posts.


SERVICE LAYER INTERFACE
SERVICE IMPLEMENTATION
AUTHORIZA
TION LAYER
DATA ACCESS LAYER INTERFACE
DATA ACCESS IMPLEMENTATION
MEMCACHED
NOSQL
MEMCACHED IMPLEMENTATION
HBASE IMPLEMENTATION