History of web search engines-:

The beginning of it all-:

The need for search engine first surfaced in July 1945. Vannevar Bush published an article "As We May Think" in Atlantic monthly. Few lines from his article were -: "The difficulty seems to be, not so much that we publish unduly in view of the extent and variety of present-day interests, but rather that publication has been extended far beyond our present ability to make real use of the record" which shed light on a greater issue i.e. we need an index to be built on the data we have as the data is growing so much so we are unable to make real use of it. So he proposed the idea of a system called Memex. Memex is a blend of words memory and index and is essentially a device in which people would compress and store their books, communications and records, it is done in such a manner that it will allow easy, fast and flexible access to content in it and this idea of memex influenced the development of early hypertext systems and eventually lead to creation of WWW(world wide web).

Then came the era of the father of information retrieval systems Gerard Salton (1960s - 1990s), his work greatly contributed towards the development of modern web search technologies. He and his team at Cornell and Harvard developed the SMART(Salton's Magic Automatic Retriever of Text). SMART exhibited the concepts like Term Frequency(TF), vector space model, term discrimination values, Inverse Document Frequency(IDF) and relevancy feedback mechanisms. He also wrote a book called A theory of Indexing which explains many of his tests upon which search is still largely based

The first Search Engine-:

Archie was the first web search engine to surface in 1990 created by Alan Emtage, a student at McGill University in Montreal. The initial name was "archives," but it was shortened to Archie.It was a tool to index FTP archives and allowed people to find specific files, it made use of a script based data gatherer and a regular expression matcher for finding file names queried by a user.

Before WWW files were shared using FTP but it was only effective for small groups and then Tim Berner Lee introduced WWW and created the virtual library which is a catalogue of the web.

Archie gained popularity and University of Nevada System Computing Services developed **Veronica**. Veronica worked on plain text files which were sent via **Gopher** an alternative to Archie. Another user interface appeared and it was **Jughead** which performed the same task as Veronica but only searched a single server at a time.

World Wide Web Wanderer - The web's first bot which was created by Matthew Gray in 1993. At first, he wanted to measure the growth of the web by counting the active web servers but he upgraded the bot to capture actual URLs and the database he created was called **Wandex** but the problem with this bot was the fact that it caused lag because it was accessing same pages hundred of times a day. Though he fixed the bot but people became sceptical of bots and started questioning their value.

Around december 1993 there were three bot based web search engine -: **Jumpstart, World Wide Web Worm** and **RBSE**(Repository-Based Software Engineering) spider. Jumpstart collected the info about a page's title and header using a simple linear search which eventually lead to its downfall as web grew. The WWW Worm indexed titles and URLs but it showed the results in order it found them without any ranking as it was the case with jumpstart. This problem was took care of in RBSE spider as it had a rating system but the problem with these all was the fact that you only get results if you know the exact name of what you were looking for.

Then in 1994 came the **Infoseek(**webmasters could submit pages to it in realtime and became default search engine for Netscape.**)** a popular search engine which came around in 1994..

The ElNet Galaxy was a web directory and it was organized in a similar manner to how today's web directory are organized and it utilized different web features but the web size in 1994 doesn't called for a web directory even then soon more web directories were followed.

In may 1994 **ALIWEB**(Archie-Like indexing of web) was launched and it crawled meta info and allowed users to submit there pages they wanted indexed along with a description what this meant was there was no need for a bot which collects data and no excessive usage of bandwidth but the thing was users didn't know how to submit their sites.

Then came the **Yahoo Directory** created by David Filo and Jerry Yang in april 1994 and it was a collection of their favorite web pages but increasing number of pages motivated them to become searchable directory. It also included a man-made description associated with each URL. They started to add informational sites for free but commercial sites were charged. But by 2014 end Yahoo Directory was closed.

WebCrawler was the first crawler that indexed whole pages and gain too much popularity that it cannot be used during day time. It was followed by other search engine such as **Lycos(**went public with a huge catalog of 54k, and used ranked retrieval, prefix matching and word proximity**), LookSmart(**gave tough competition to Yahoo!**), Excite.**

Then **AltaVista** was launched and it offered many new features such as unlimited bandwidth(first one to do so), first one to allow for natural language queries, advanced search techniques, users can add or delete their own URL within 24 hrs and it also provided search tips.

The Dawn of Modern Web Search- Google(initially BackRub)

In 1996, Larry page and Sergey Brin students at Stanford University developed a search algorithm- BackRub named so because of it uses BackLinks(the incoming links to a web page) to rank its search results and this ranking scheme was called Citation notation. In their PageRank algorithm all incoming links to a web page counted as vote towards the reliability of that page but some links counted more than other based on certain criteria. The modern name for this type of search is Social Search.

Between 1996 BackRub and launch of Google in 1998 the other search engines that surfaced were **Inktomi:HotBot (**used paid inclusion model**), AskJeeves (**natural language search engine which used human editors to match search queries, used clustering to find local web communities**), MSN** Search.

Then there were other search engines which were launched after launch of Google like **AllTheWeb**, **Teoma**, **Snap**(too complicated for general web surfer),**Cuil, LiveSearch(**by microsoft**),Bing(**a rebranding of LiveSearch/msn**), Schema**(created in collaboration of Google, Microsoft, Yahoo).

During this period Google changed SERP(search engine results page) with new features like news, videos, images, local and others. It released "Google Suggest"- dropdown of suggested topic related to search. It introduced new web indexing system named Caffeine which provided 50% fresher search results. Updates likes Google Instant(real-time search results), Panda(major change in ranking algorithm),Penguin, Hummingbird(first search algo that has ability to parse intent behind a query not just the language), Pigeon, Mobilegeddon,RankBrain(google revealed that machine learning has played important role in ranking algo), Possum.

Google in 2014 started prioritizing websites which uses HTTPS which was seen as google effort towards web security.

Google is still dominating the market with approx 90.46% and other search engine like bing(3.13%),Yahoo!(2.21%) and etc. These statistics show how Google has continuously dominated and revolutionized modern web search.