# Web Technologies

## Search Engines

- Billions of web pages, manually finding anything is nearly impossible.

- Software systems have been developed to find what users want quickly. These systems are search engines, with Google being the most well known.

- Search engines build up indexes of websites that can be searched quickly by various search algorithms.

- Early engines required site owners to notify the search engine sites bur later various robots, known as **spiders,** searched for sites by **crawling** over websites and indexing the words found there.

- All search engines now search the internet for various keywords then index these links to where they are found. The index is made available to users. Engines can cope with misspellings and provide various languages.

- As well as visible words, search engines also make use of meta tags - extra information that web designers add.

- With the web expanding search engines need to find the most relevant results for a user’s input even if they are worded in a clumsy or inaccurate way.

- A lot of search engines make money by being paid to display certain results by the company’s that own certain sites.

**Pagerank Algorithm**

- One of the most successful ways that search engines have used to produce meaningful results is the Pagerank algorithm.

- This has been a particularly successful process applied by Google to its web searches. This doesn’t just look at content to assess relevance; it ranks possible web pages according to external links. So at its most basic, if a web page has many links to it from other pages, these are considered votes, and it is deemed to be popular and more worthy of consideration.

- However, not all votes are equal. Some votes are deemed to be more significant than others and this is based yet again on the number of links into them. The process can be applied recursively to get a fairly good estimation of how important a page is.

- The original Pagerank algorithm is prone to abuse by those who set up **link farms** to artificially increase the number of links to favoured pages. Google continues to alter its algorithms to circumvent such problems.

**Client-side and Server-side Processing**

- Most web interactions involve two principal connected entities: the surfer or client and the web server that holds the resources that the client wants.

- These resources may be static data collections or often involve multiple interacts as, for example, when a customer is making a booking of some sort.

- Decisions have to be made about what processing occurs where. The basic issues are to do with performance and security.

- Code can be written to do processes on either the client or the server.

**Arguments in Favour of Client-Side Processing**

- Client-side processing reduces the load on the server. The server may be busy handling multiple transactions and if some of the processing can be off loaded to the client machines this can speed up the server.

- The user will have a better experience if data input is checked there and then without the delays for immediate sending of each item to the server for checking.

- Client-side processing also reduces the amount of web traffic. Reduces the likelihood of erroneous data being sent to the server to validate and process.

**Arguments in Favour of Server-Side Processing**

- Data validated by a client-side script may still have problems with it. It is still necessary to have validation on the server side.

- Server-side processing is essential for actually querying a database. It is vital to keep the data owned by an organisation secure, if not secret, and so any processing of that data must take place under the control of the organisation.

- So SQL processes will largely have to be located at the server end. No sensitive data should be sent to the client where it could be intercepted and manipulated.

**HTML**

- Web pages are interpreted and displayed by a browser.

- HTML is comprised of elements called tags which enclose other objects.

- Styled with CSS.

- Markup language, defines how things are displayed.

**CSS**

- Used to style web page elements.

- Styles can be applied to certain types of elements once.

- Also supports **media queries** which can determine things like screen size and adapt.

**Javascript**

- Interpreted language used in web browsers.

- Can be used for validating forms, animations and modifying the web page.

- Dynamically typed which means that data types don’t need to be defined (like python).