

What is an inverted index?

Inverted index is a hash that maps words to places where they appear, it features most prominently with search engines where it matches the search query with the pages where the query exists.

What is the Fourier transform?

The Fourier transform is an algorithm that breaks up data into its ingredients. An example of this would be music compression, MP3 compression works by separating the notes and/or frequencies of a song then eliminates the least important ones and keeps the important ones. It's a great algorithm for processing signals, probably being the centerpiece of an app like Shazam, and also applies to images such as the JPG format, it also applies to predicting earthquakes and analyzing DNA.

What are parallel algorithms?

To make a quicksort algorithm run at $O(n)$ instead of $O(n \log n)$, you'd need to run the algorithms in parallel across all cores of a computer. They're tricky to design and the time gains are not linear, meaning that doubling cores, won't double algorithm speed gains. Thus load balancing needs is crucial, how to split up loads properly between cores, as well as the overhead of managing parallelism, how to divide work done by different cores. The *distributed algorithm* is a kind of **parallel algorithm** that is gaining popularity, because it runs on hundreds of cores on different machines. An example is the **MapReduce** algorithm. The components of **MapReduce** are simple enough, they are the array **map** and **reduce** functions.

What is a bloom filter?

To determine if a page's crawled, a **bloom filter** offers a solution. **Bloom filters** are probabilistic data structures in that they give an answer that could be wrong but is probably correct. As a result, *false positives* are possible, might say it's crawled but it's not. But *false negatives* are not possible, it's always right when it says it's not crawled. They're useful because as opposed to hash tables (if they're used), they take up so little space.

What is HyperLogLog?

HyperLogLog determines unique search queries by users. It keeps a log where every time a search queries, it checks the log and adds it if it doesn't find it. It approximates the number of unique elements like **bloom filters** and also takes little space.

What is the SHA algorithm?

The **secure hashing algorithm** is one that instead of generating an array index from a string like a regular hashing function, it returns a hash from string to string. It's a useful algorithm to determine if two files are the same specifically if large in size, by putting them through the **SHA** and comparing hash results. It's also useful when checking passwords or other data that you don't want to reveal. **SHA** is also locally insensitive, meaning that if you change one character of a string, the hash result is totally different from the original string. A locally sensitive hash would be a *Simhash*, which changes the hash result if the original string is slightly altered.

What is **Diffe-Hellman key exchange**?

The **Diffe-Hellman algorithm** solves a problem of how to encrypt a message so it's read by the receiver. Agreeing on a cipher would be the simplest and most obvious solution. Its solution is having both a public and private key, the public key encrypts a message when it's sent and the private key decrypts when it's received.

What is **linear programming**?

Linear programming is an approach that concerns itself with optimization, by picking the most efficient course of action.