**Common Algorithms in Action**

(Source: [Oleksii Trekhleb’s GitHub Repository on Algorithm Implementations](https://github.com/trekhleb/javascript-algorithms))

Reference worksheet explaining algorithms in action within common interfaces. Note that many of these algorithms are proprietary, but they have a foundation in common algorithms that we are calling “first-principle” algorithms.

**Search**

**(Google)**

**Term frequency-inverse document frequency (**[**TF-IDF**](http://en.wikipedia.org/wiki/TF_IDF)**)**

Term frequency-inverse document frequency ([TF-IDF](http://en.wikipedia.org/wiki/TF_IDF)) is a technique used to retrieve relevant documents from a collection of documents. It normalizes the importance of a term across many documents of varying sizes. TF-IDF is calculated by multiplying the term frequency by the inverse of the document frequency of that term. Term frequency is how often a particular term appears in this document, and document frequency is how many documents contain at least one instance of that term. The term frequency represents how important a term is to a document. The inverse document frequency acts to prevent very common terms from affecting the search results. A term that appears in all documents is not a good term to differentiate documents, so very common terms across all documents are penalized and contribute less to a particular document’s score.

**Binary Search**  
In computer science, binary search, also known as half-interval search, logarithmic search, or binary chop, is a search algorithm that finds the position of a target value within a sorted array. Binary search compares the target value to the middle element of the array; if they are unequal, the half in which the target cannot lie is eliminated and the search continues on the remaining half until it is successful. If the search ends with the remaining half being empty, the target is not in the array.

Uses and Common Applications: Ranking documents for a search term; Identifying search term value matches

**Entertainment**

**(Netflix)**

**Merge Sort**

In computer science, merge sort (also commonly spelled mergesort) is an efficient, general-purpose, comparison-based sorting algorithm. Most implementations produce a stable sort, which means that the implementation preserves the input order of equal elements in the sorted output. Mergesort is a divide and conquer algorithm that was invented by John von Neumann in 1945.

Uses and Common Applications:

Comparing User’s video ranking and preferences

**Social Media**

**(Facebook)**

**Dijkstra's Algorithm**  
Dijkstra's algorithm is an algorithm for finding the shortest paths between nodes in a graph, which could represent, for example, road networks. The algorithm exists in many variants; Dijkstra's original variant found the shortest path between two nodes, but a more common variant fixes a single node as the "source" node and finds shortest paths from the source to all other nodes in the graph, producing a shortest-path tree.

Uses and Common Applications:

Determining the shortest path between friends and associates within Facebook network; Creating “weights” that define closeness of a friendship

**ECommerce**

**(Amazon)**

**Collaborative Filtering (E.g., a Bloom Filter)**A bloom filter is a space-efficient probabilistic data structure designed to test whether an element is present in a set. It is designed to be blazingly fast and use minimal memory at the cost of potential false positives. False positive matches are possible, but false negatives are not – in other words, a query returns either "possibly in set" or "definitely not in set". A bloom filter can be used on a blogging website. If the goal is to show readers only articles that they have never seen before, a bloom filter is perfect. It can store hashed values based on the articles. After a user reads a few articles, they can be inserted into the filter. The next time the user visits the site, those articles can be filtered out of the results.

Uses and Common Applications:

Setting up user recommendations based on current and past browsing or purchasing patterns.