Google, one of the most recognized and used search engines throughout the world, uses particular search algorithms in order to provide its customers the best possible search experience. Despite being one of the most used search engines in the world, Google has revealed little about search algorithms they use. There are, however, some algorithms that the public knows that Google uses. One of the algorithms is the Page Rank algorithm, which ranks pages based on the number and quality of links to a particular page in order to determine its importance. Besides using the PageRank algorithm, it is known that Google still currently uses, in some form, the Hilltop search algorithm.

The Hilltop algorithm was introduced in order to combat many problems inherent with providing quality search results; one problem that the authors of the algorithm contest was a problem was that the content of a page alone cannot determine how reputable a webpage is. Additionally, issues with spam encouraged its creation. Moreover, even with reputable websites, providing the best search results for a particular query is difficult if using an approach based on solely matching the keywords from the search.

The Hilltop algorithm favors pages that are specifically about a certain topic and contain many links to outside pages that are about that topic. These outside pages must be examined, though, because pages seemingly written by outsiders could abuse the system; thus some definition of "outside" source must be established. Google ranks the results through examining links that "expert" pages have linking to a particular website that appears in the search results as well as looking at how well the search terms used match with the page. "Authorities", pages that have links from a numerous amount of "expert" pages, will rank well solely using this algorithm. A relationship, then, is evident between the "expert" and "authority" pages: to rank high, pages must be "authorities" meaning that many "experts" must link to them.

When providing the results for a certain search, the algorithm determines the most suitable "expert" pages on the topic associated with the query topic. When examining "expert" pages only relevant links to the topic that is being investigated will be considered for the search; this demonstrates that the search results returned will be relevant to the query. More specifically, the links on the "expert" pages will have all the terms used within the search associated with them. The links that the "experts" provide must have at a minimum two unrelated or unconnected pages concerning the query. Note that hilltop will not produce results when no

"experts" link to the particular page. This fact illustrates that Hilltop only helps with providing more accurate search results.

In more technical terms, hosts are determined to be affiliated using the IP address and hostnames. The authors of this algorithm define affiliation to be "sharing the same first 3 octets of the IP address" or "[t]he rightmost non-generic token in the hostname is the same" (Bharat, Mihaila). An "expert" is more precisely defined to be a page that contains as many links that are associated with as many non-affiliated hosts. In other words, to be considered an "expert", supposing that the page contains a minimum amount of links, *x*, there should also be *x* non-affiliated hosts associated with those hosts. "Expert" pages that are associated with a particular search are linked together through an inverted index using keywords, in particular only text that is in important expressions within an "expert." Besides having an inverted index for the "experts," the links that are contained within every "expert" is also stored. When a search is conducted, a certain number of "experts" are compiled that are determined to be the most related. A score is then assigned to each "expert" through examining the links that these "experts" contain. An "expert" must have at the minimum a single link that has all the keywords from the search.

Regarding the mathematical basis of hilltop, the "expert" and target scores must be examined. An "expert" is defined to be a 3-tuple. If p is the number of number of words that appear in a search, and this search can be referred to as q, then one component of the 3-tuple can be calculated through only looking at important expressions that have exactly k subtracted by whatever index the score tuple component has i.e. for the third component, S_2, examines expressions that have two less terms than what was being searched for. The authors more precisely defined an individual component, S_i, as equal to the sum of $_{\{\text{key phrases } p \text{ with } k - i \text{ query terms}\}}$ *LevelScore*(*p*) * *FullnessFactor*(*p, q*). A phrase obtains its score based on what type of an expression a phrase is where the expressions could be of the following types: titles, anchor text or headings. On the other hand, the fullness factor considers how many terms in a particular key phrase is noted within the search query performed. As relevance is desired, the first component is given weight, since this first component represents having "experts" that have all the keywords matched within the document. The score is then changed from its tuple representation to a scalar through assigning weights to each of the components and performing a summation; the authors

assign the first component a weight of 2^{32} , the subsequent component is given a coefficient of 2^{16} and the last component is given no coefficient.

The target score, on the other hand, examines the pages that the "experts" link to. Through examining the targets, a ranking can be compiled. Targets must satisfy the non-affiliation principle. A score is assigned through examining how many "experts" are linked to it and the relevance that each "expert" has to that target, as well as the phrases associated with the links. The target score is constructed using three steps, and it involves graph theory: initially, for all the links that an "expert" contains, draw an edge connecting the expert node and target node; next, if a target has two edges with an affiliation between the two "experts," the edge that has a lower edge score is removed (the edge score can be noted in the original paper that authors wrote). The target score is then generated through creating a summation of the edge scores.

In the studies performed using Hilltop, the algorithm performed well against the competing search engines of the time. The authors observed that in the case of more narrow searches, such as the names of companies, "for about 87% of the queries, *Hilltop* returned the desired page as the first result, comparable with *Google* at 80% of the queries" (Bharat, Mihaila). In another test of Hilltop's capabilities, its efficiency with topics that had many resources readily available, were examined. In this test, the results demonstrated "for broad subjects our engine returns a large percentage of highly relevant pages among the ten best ranked pages, comparable with *Google* and *DirectHit*, and better than *AltaVista"* (Bharat, Mihaila). The algorithm's efficiency in a variety of situations is therefore evident.

After examining and thinking about the algorithm, to increase the likelihood of appearing first in the search results, it would be necessary to cater the website to queries that the user has made and have discussions with other similar content website owners. If the website involved a topic that customers would commonly query, then it would be suitable to cater the website towards the queries through featuring keywords from the query within the website. The discussions with other website owners would be necessary because it would make the website appear "authoritative" because if other reputable websites that are considered "experts" link to a website, the website will appear more authoritative and thus appear higher on the search results. Essentially, to ensure a higher search results position, mathematically determine the input values would generate the highest output values using the functions that determine the "expert" and target scores.

References:

https://en.wikipedia.org/wiki/Hilltop_algorithm

http://ftp.cs.toronto.edu/pub/reports/csrg/405/hilltop.htm