

The Anatomy of a Search Engine Ranking Algorithm

A search engine's ranking function or more accurately it's sorting function; takes into account a variety of factors when ordering documents returned for a search query. It is often assumed (incorrectly) by SEO's that a search engine's ranking methodology is a linear evaluation of pre-defined criteria against which a given web page is evaluated. Like an itemized checklist. The problem with this type of conceptualization is that it can lead to the development of potentially harmful search optimization techniques which are based upon incorrect assumptions regarding search engine ranking technologies.

The following map is greatly over-simplified; however it is built upon a simple taxonomy which is designed to illustrate the intricate self-similar nature of the modern search engine sorting algorithm.

A Ranking Function is built upon two distinct concepts: Relevancy and Importance.

Each search engine defines these concepts in a slightly different fashion, however there are certain commonalities that are nearly universal.

Relevancy can be seen as a product of local and non-local measures of query/topic relatedness. Local relevancy can be further defined as a product of three distinct types of analysis:

- Hit Analysis: The matching of query terms with document terms in an attempt to determine document relevancy. Specifically this can include evaluation of term frequency, term location, term proximity and term emphasis (such as bold or emphasized instances of a given term) to name a few.
- Document Level Analysis: The evaluation of context, sentence and passage semantics, etc.
- Domain Level Analysis: The evaluation of a given document in context with the other documents from the same domain.

Non-Local Relevancy commonly involves two types of analysis:

- Inbound Anchor Text: The evaluation of anchor text associated with inbound citations.
- Citing Document Text: The textual evaluation of the citing document.

Importance can be seen as a product of local and non-local measures of a given documents inherent value. Local importance is commonly defined as a product of:

- Document Content: The evaluation of the relative quality of a given document and its architecture.
- Temporal Factors: Evaluation of a given document based upon historical data associated with the document.
- Topical/Community: Evaluation of a given document using a Hubs and Authorities scoring method.

Non-Local Importance is almost always an assessment of inbound citations or backlinks associated with a given document. For example Google's PageRank and Yahoo's short lived WebRank are examples of a modern importance evaluation used by search engines.

To sort the list of documents returned for a query, the ranking function must employ both the relevancy and importance scores assigned to each document in an attempt to furnish the results of greatest value to the user.