

Matching Strings

There are two major approaches to matching a document with a query: (1) exact match, and (2) best match. The former is typically realized by the means of Boolean matching and retrieval, and the latter is implemented with different models that account for approximations and probabilities. Let us explore these two methods in detail.

Exact Match (Boolean)

This is a model where the query specifies a very precise matching criteria [Manning et al., 2008]. A typical example of a boolean query is *java AND coffee*. This query indicates that a matching document has to have both *java* and *coffee*. It means if a document has only one of these terms, it will not be matched. This can be an advantage or a disadvantage depending on the situation. Also, in the pure exact match situation, the result for a given query is a set of documents without a ranking. This is because for an exact match situation, a document either matches to the query or it does not (binary relevance).

Best Match

In this model, a query specifies what a good or best matching document could be. Every document has some chance of matching the query. This “chance” could be measured using a probability or expectation and used to rank the results. Thus, with this model, we get some kind of score for each document with respect to the query that can be used to rank or organize the matching documents.

Each of the above approaches have their pros and cons. They are summarized in Table 1.

Table 1: Comparison of exact match and best match retrieval.

EXACT MATCH	BEST MATCH
Easy to understand and explain.	More difficult to convey the cognitive model of “relevance”.
More useful to an advance searcher.	Quite usable even for a novice searcher.
Can be efficiently implemented.	Less efficient. Cannot reject documents early.
Difficulty increases with collection size.	Scales well.
Higher precision comes at the cost of recall.	Getting higher precision does not result in harming the recall much.

[Manning et al., 2008] Manning, C. D., Raghavan, P., and Schütze, H. (2008). Boolean retrieval. Introduction to Information Retrieval.