

To make Medium work, we log user data. By using Medium, you agree to our [Privacy Policy](#), including cookie policy.



To make Medium work, we log user data. By using Medium, you agree to our [Privacy Policy](#), including cookie policy.



Query Segmentation



Daniel Tunkelang [Follow](#)

Apr 28, 2017 · 4 min read

The previous two posts focused on using query rewriting to increase recall. We can also use query rewriting to increase precision — that is, to reduce the number of irrelevant results. While increasing recall helps us avoid small or empty result sets, increasing precision helps us avoid large result sets that are full of noise.

In this post, we'll discuss the first of two query rewriting strategies used to increase precision: query segmentation. We'll first talk about how to perform query segmentation, and then about how to use query segmentation to increase precision through query rewriting.

Semantic Units

Query segmentation attempts to divide the search query into a sequence of semantic units, each of which consists of one or more tokens. For a single-token query like *machine*, there's only one possible segmentation. Multiple-token queries, like *machine learning framework*, admit multiple possible segmentations, such as “*machine learning*” *framework* and *machine* “*learning framework*”. In theory, the number of possible segmentations for a query grows exponentially with the number of tokens; in practice, there are only a handful of plausible segmentations.

The goal of query segmentation is to identify the query's semantic

To make Medium work, we log user data. By using Medium, you agree to our [Privacy Policy](#), including cookie policy.



Discover Medium

Welcome to a place where words matter. On Medium, smart voices and original ideas take center stage - with no ads in sight. [Watch](#)

Make Medium yours

Follow all the topics you care about, and we'll deliver the best stories for you to your homepage and inbox. [Explore](#)

Become a member

Get unlimited access to the best stories on Medium — and support writers while you're at it. Just \$5/month. [Upgrade](#)

[About](#)[Help](#)[Legal](#)