

Lab 5: Information Access

TDT4275: Natural Language Interfaces

Jonas Myrlund

April 23, 2013

1 Written Assignments

A Using either web searches or a large corpus, devise a query using a polysemous word.

I want to find out more about large cranes – the bird. I start out with the following query: “large crane”.

See figure 1 for the initial results using Google¹.

B Then, devise two or three alternate queries which help to disambiguate the polysemous word and repeat the search.

The query “large crane” could be less ambiguous in the following ways:

1. “large crane bird”
2. “large crane feathers”

As figure 2 exemplifies, these queries result in only documents relating to the *bird sense* of the word *crane*.

C For each query, classify each of the top 10 search results as correct or incorrect.

Again, I will be using Google for the queries.

C.1 Query: “large crane”

The query yields only one correct document: the second search result. The others documents relate to the wrong sense of the word. For details, see figure 1.

¹<https://google.com/search?q=large+crane>

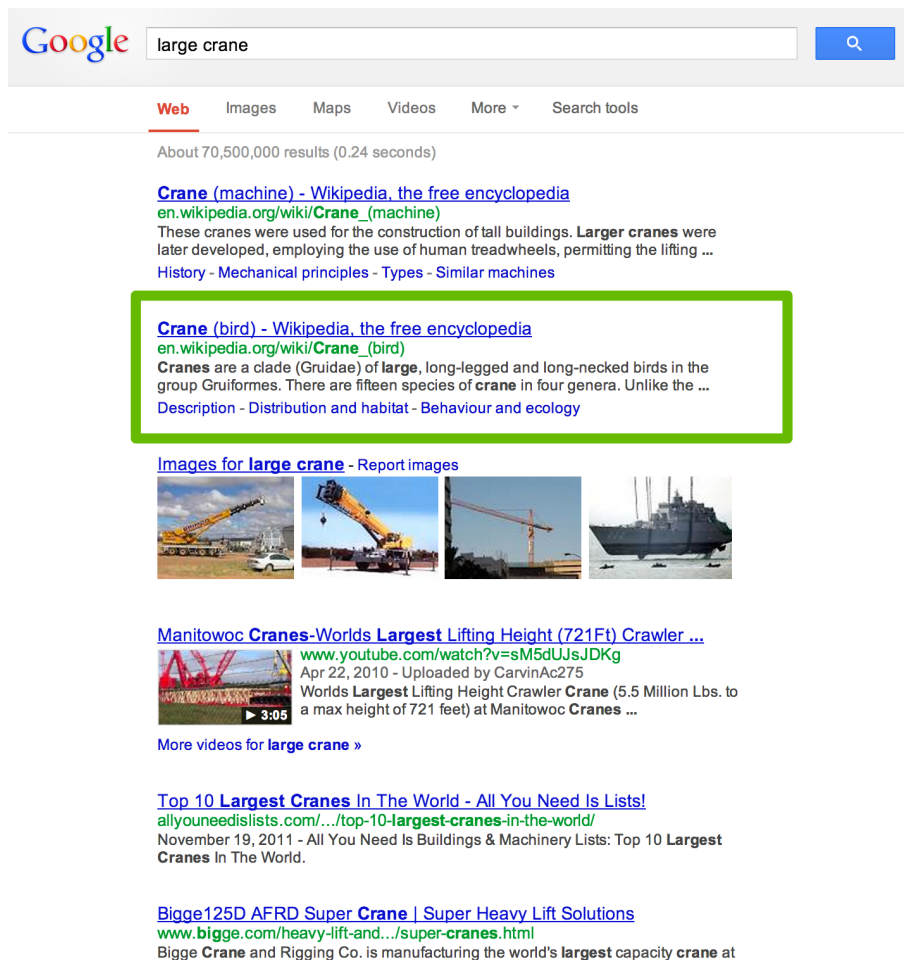


Figure 1: Google yields only one relevant search result when searching for “large crane”.

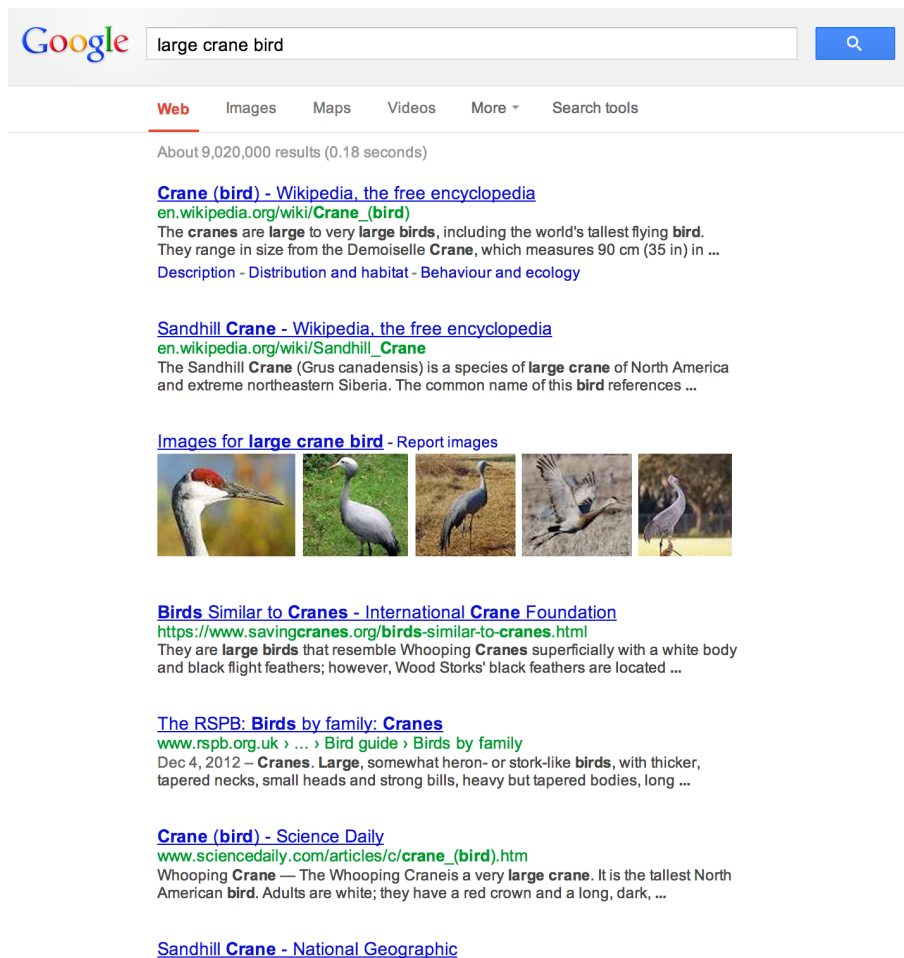


Figure 2: When adding the term “bird” to the query, all the top results are relevant.

C.2 Queries: “large crane bird” and “large crane feathers”

For both these queries, all the first 10 search results relate to the right sense of the polysemous word in Google.

Yahoo! and Bing, however, both give three documents relating only to feathers, which is not what is wanted.

D Create an interpolated precision-recall curve comparing the accuracy of each of the queries.

The recall values for the first query, “large crane”, stagnates in that it gives only one correct document. Thus, the scales of the precision-recall curves do not overlap to a degree that makes it viable to plot them together.

However, the curves quite clearly show the improvement adding a disambiguating term yields, when the precision axis is normalized to one. See figure 3, 4 and 5.

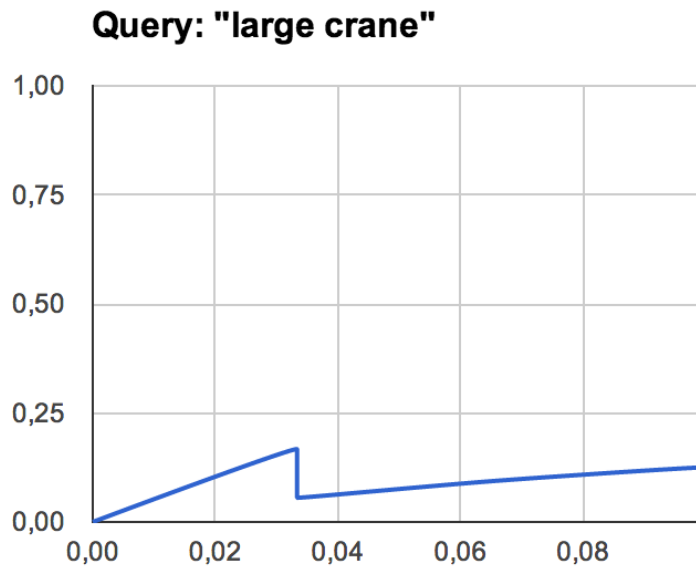


Figure 3: Precision-recall curve for “large crane”.

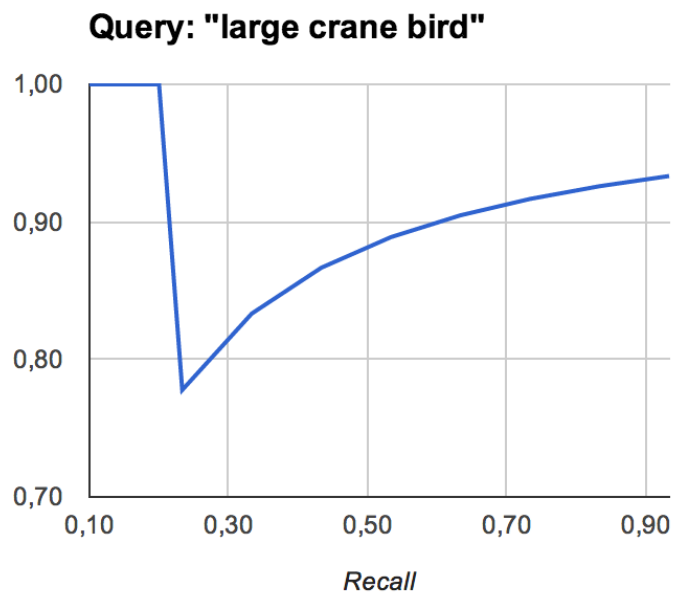


Figure 4: Precision-recall curve for “large crane bird”.

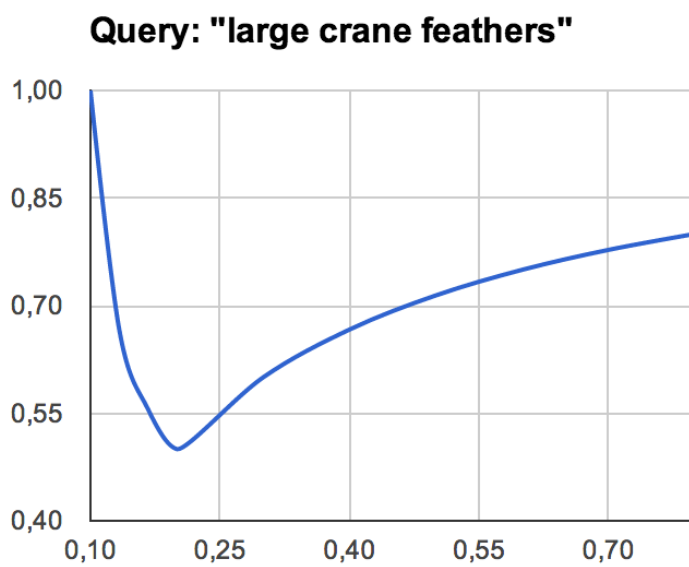


Figure 5: Precision-recall curve for “large crane feathers”.