Study: Google ranking vs keywords

Obejective:

This study aims to find out whether there is an relation between the position of google ranking for a certain webpage and the keyword this page ranked for.

Data:

We crucnhed the keywords and backlinks data of around 50 webpages in backlinko.com, and we do a counting of keyword appearance.

More specifically, each webpage has many backlinks directing to this page, and this page also has many keywords it ranks for(in top10). Therefore for a certain webpage and a keyword this page ranks for, we will count:
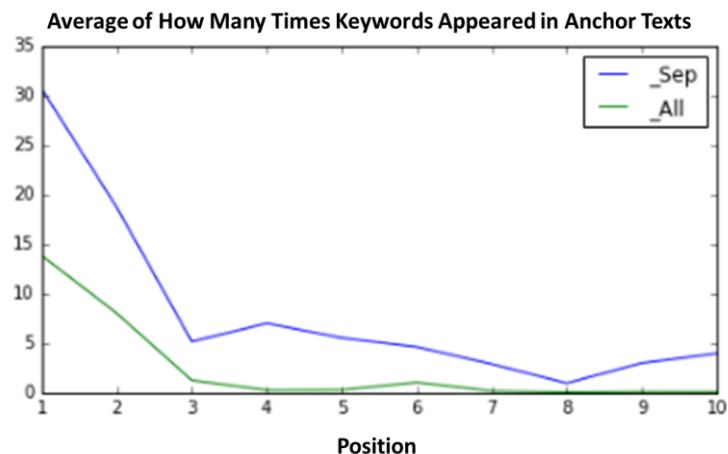
1) how many backlink anchor/surrounding texts contains this keyword.
2) Whether this keyword is contained in the page or not.

Note that we use two versions of definition for "contain":

1) _All: Only increase the count if the keyword as a whole is contained in anchor/surrounding text or the page content. Strictly contains.
2) _Sep: Increase the count if every word in the keyword is contained in anchor/surrounding text or the page content. The keyword as a whole may or maynot be contained. Thus it is a more relaxing condition than _All. _All is a subset of _Sep.
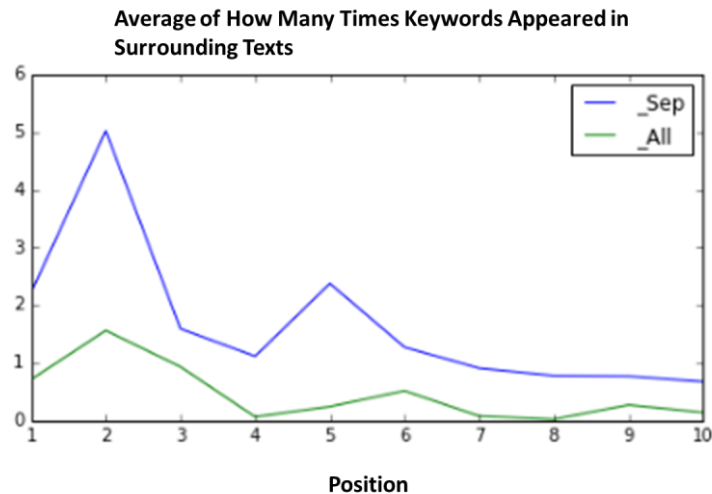
Result:

1. *Does google rank a page based on how many backlinks' anchor text contain the keyword you are searching for?*

This graph shows that the more anchor texts contains a certain keyword, the higher this page ranks for this keyword.
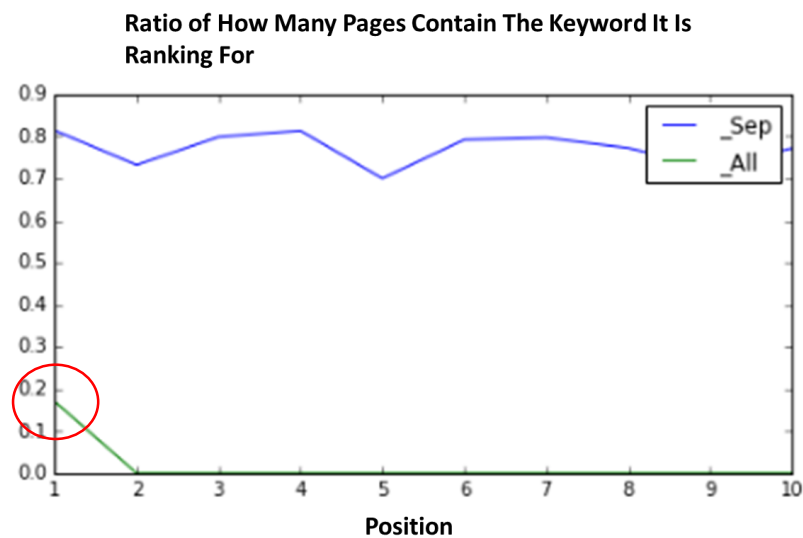
For example, on average a #1 keyword of a certain page will appear in 30 anchor texts of this page's backlinks, while a #3 keyword of a page will only appear in around 6 anchor texts of this page's backlinks.

2. ***Does google rank a page based on how many backlinks' surrouding text contain the keyword you are searching for?***

Average of How Many Times Keywords Appeared in Surrounding Texts

Position

Although surroudning text shows weaker trending compared to anchor text, we can still see trending that the more surrounding texts of a page's backlinks contain a keyword, the higher this page ranks for this keyword.
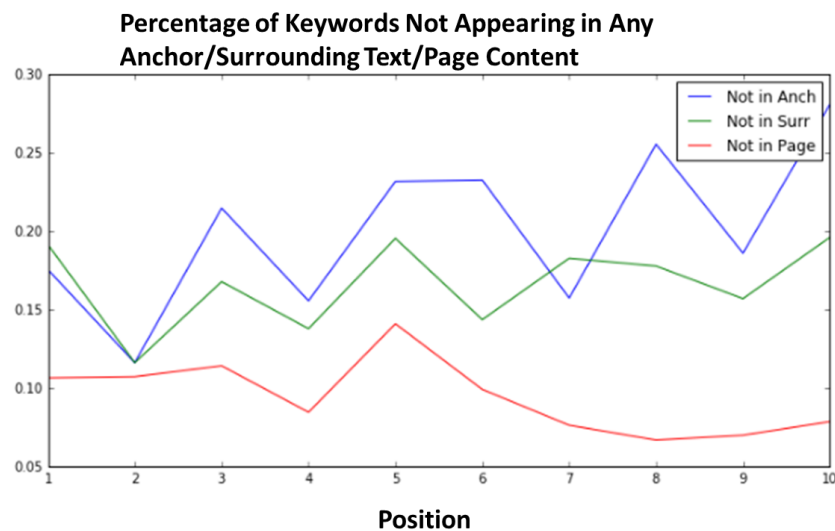
3. ***Does a page ranks higher for a keyword if this keyword appeared in the page?***

Ratio of How Many Pages Contain The Keyword It Is Ranking For

Position

We can see that there is no obvious trending that higher ranked pages are more likely to contain the keyword it ranks for.

However, if we use the stricter definition of "contain" (_All, meaning only increase count if the keyword as a whole is contained in the page content), ony position#1 has non-zero value of aounrd 0.18. This implies that all pages containing keywords it ranks for will be ranked at #1. This conclusion is drawn based on the above graph, and it seems to be too ideal. It could be due to the fact that not enough data are used for this study. However, it still give us an implication that if the page strictly contains a keyword, it is more likely to be ranked higher for this keyword (it may even be ranked at #1!).

4. *How many percent of keywords does not appear in anchor/surrounding texts/page content?*

Percentage of Keywords Not Appearing in Any Anchor/Surrounding Text/Page Content

Position

This graph shows the percentage of keywords not appearing in anchor/surrounding texts/page content. (Not appearing means not even a single word in keywords appeared in any anchor/surrounding text/page content). There is slight trending that higher ranked pages has less keywords not present in anchor/surrounding texts.