



Web search services, including Google search, follow a What-You-See-Is-What-We-Index policy. When users click on a search result, they expect to view the part of the document or web page their query matched. This is what we all depend on when we do our search queries: e.g., hotels in a small town in Italy for a vacation, the kings of England in 1400s for a high school exam, what a politician said ten years ago to check for flip-flopping, the precise form of an oft-quoted statement to settle a bet and so on.

In scholarly publishing, for most articles, just the abstract is accessible to all users, full-text access is limited to subscribed users. As a result, Google.com search currently matches just the abstracts of subscription articles. However, a lot of the magic in science arises from serendipity - connections made by researchers without planning ahead. Abstracts are of course what the authors of an article considered the most significant part of the article. Actual impact of an article can come from many other components - it could be the methodology used, it could be the relationship with a different technique, it could be the intermediate results, it could be the lemmas used on the way to proving the theorem and so on. Over the years we have found that a large fraction of scholarly queries we see on Google Scholar (which indexes full-text of all scholarly articles), match in the body of articles.

We, the Google Scholar team, have worked with the Google web search group to consider an extension of the What-You-See-Is-What-We-Index model to improve discoverability of scholarly articles. The key idea is to take institutional subscriptions into account so that we can match the user's query against subscribed articles as well. This would allow institutional users to discover articles they have access to and then to seamlessly access the copy available to them. To implement this, the Google web search system would need to know which URLs on a journal web site are accessible to users at an institution. This is beyond the link resolver based mechanism used for Library Links - which does not have information about individual URLs. We are currently working with many of our publisher partners to enable this.

Anurag Acharya
Distinguished Engineer
Google