Sam Hewitt
Ssh5d

Answer Retrieval in Search Engines

For the most part, search engines can't answer questions, or for that matter actually show us where in a document the answer to a question is located. Search engine results mainly return a list of documents deemed relevant to a query and it is then up to the user to find the answer or information they are looking for within those documents. This method works well for users who have time to delve into multiple documents to find the answer, but many of today's users are looking simply for a quick answer to a question, or only need a small amount of information. Going through document(s) might hamper their ability to get the answer or information that they are looking for.

Some examples of queries for this type of small information retrieval are: "how do you make mac and cheese", "write a for-loop in java", "when did humans land on the moon", and "who won the Nationals game last night". These searches do not imply a need for document retrieval, but rather to display only the information requested to the user in an easily digestible way.

Going back through the examples, if the user wants to know how to make mac and cheese, the user could be shown a recipe, an ingredient list and cook time. For programming or math help, the user could be shown a step by step procedure and an example. For a query about when something happened, the user could be shown a date and a brief snippet of information about the event. For sports information, the user could be shown a box score for the game.

Despite the simplicity of the returned information, the process of a search engine creating that information from a multitude of different sources is undoubtedly difficult. Most domains present information in different formats and figuring out what information lives where on a given domain would be a costly and time intensive endeavor. Also, if information is present in this way, as an answer, to the user, the user will be likely to rely on its truthfulness and completeness. If the information were pulled from a poor source, then the user's need for an answer would not be fulfilled.

As a starting point to determining where and how to compile this type of information, a search engine could first identify which domains they want to provide answers, in addition to documents for. Once those domains have been established then a set of reputable sources needs to be found and processed. Ideally, if the search engine determined that there were multiple reputable sources for a given domain, they could display the answers from both to the user, either split out or combined in some fashion. It would also serve the user's information need to see the sources for the information

being shown to them so that they can quickly find more information on the topic if the search engine's "answer" did not sufficiently answer their request for information.

Currently Google is providing "answers" to two of the domains mentioned above: date questions, and sports. After doing some sample searches, I found that Google would consistently serve the date and a portion of a Wikipedia article for questions such as "when did man land on the moon?". While Wikipedia is a decent source for common information, it will not suffice well for current or contested events due to its open editing nature. Also, for more specific questions, it might not have a relevant article. For sports queries, a box score for the last game is shown as well as some information about the teams and upcoming games. Google does not reveal its source for this information

An answer retrieval system would be able to quickly fulfill a user's information need, but would require a lot of up front investment and research into pulling and displaying information back to the users. Google and Bing do seem to think this is a useful way to display information however, and are doing it for selected domains. I would expect to see this expand across more domains as their algorithms grow more complex, and user's demand more and more of this information.