From Internet Information Searching to Information Summarizing

Yu-Hui Tao
National University of Kaohsiung
ytao@nuk.edu.tw

Abstract

The paper describes an incremental advance in search engine from Internet information searching to information summarizing service. This incremental innovation requires only existing text summarization technologies and can be achieved within a shorter time frame than other radical innovations that may require years of effort on to-be-developed or to-be-validated technologies. The potential business models based on this Internet summarization service are briefly discussed with some imperfections that may occur at the initial stage.

1. Introduction

Internet technologies have been maturing and now it has come to a stage where many innovative, insightful or interesting applications are more feasible than ever. Each technology not only has its original strengths and applications, but is also maximizing its potential for creating new emergent applications with other technologies. While Internet search services, such as Google search engine, have been changing business rules and peoples’ culture [1], the journey of exploration of its unlimited potential has just begun.

Although the future perfect search applications have been best summarized in Battelle’s book [1], they may still require many years of technology development and business-model maturing processes to achieve. Davila et al.’s. [2] have classified innovations into incremental, semi-radical or radical innovations. One possible incremental innovation of Internet search is to extend its search capability to summarization capability by further summarizing the search results into an article format with the ranked URL links embedded. With the integration of only existing technologies, new IT-enabled services may be quickly created to meet the criteria of this incremental innovation, which is further discussed next.

2. Current and future search engine

How does a search work? Figure 1 gives a high-level brief sketch of the search engine process. In short, a search engine must be able to automatically search though the Internet contents, index the search results into its own database, and return a user’s query with results arranged in a ranking order based on it’s relevance or importance to the query.

While Battelle [1] described that future search engines may provide the answers to the query, some issues still need to be resolved, including searching everywhere instead of just networked computers, accumulating the wisdom of search engines, local and personalized search, search with a new interface and the impacts of Web semantic that includes the increasing Blog structured contents, and labeling the documents for perfect query approach.

3. The concept of internet search with summarization

Wouldn’t it be nice that when we want to know a certain subject, a few clicks on Google search engine
would compile a human-readable article, with the search results summarized according to the user’s preferences, as shown in Figure 1? The subject can be as simple as a terminology, or as complicated as an R&D report of specialized domain knowledge.

A simple Yahoo Taiwan tool called Minipen can illustrate part of the simple scenario (http://tw.mini.yahoo.com/). The Minipen is visible in the Web browser, and can be used as a real pen by using the mouse to underline the word in the document displayed by the Web browser. The selected word can serve as a keyword to search for Web pages, Yahoo!dictionary, Yahoo!Answer Blogs, language translation or even its pronunciation. During the search process, document formats of IE, Word, Excel, PowerPoint, PDF, Outlook, and Yahoo! Messenger are acceptable to be processed. Yahoo! Minipen is appropriate for use as a dictionary with pronunciation. But it does not go further to summarize the linked documents for the users, which can be hours or days of effort for some time-shortage users.

The summarization concept can further integrate other information technologies or systems. One good and practical example is an automobile travel planner that produces an itinerary with road map that is accessible from one’s mobile phone or PDA with GPS, as shown in Figure 1. In this case, the computer agent program takes a user’s travel requirement, submits the queries to Google search engine, and compiles different alternative itineraries for the user. The user can compare the travel schedules with the summarized text description and graphical map, and can click the link to the original sources for detailed information if needed. When the user finalizes the itinerary, all the information can be downloaded to his personal PDA or mobile phone that has a GPS and map system for driving guidance. On the road, the original sources can be accessed via the mobile device that is equipped with Internet connection.

Since there is already a high usage of Search Engines on a daily basis, we can imagine how big the demand will be when users can use their mobile devices to quickly access Internet knowledge in an automatic summarized Q&A format without screening the tremendous linked information themselves. The WebFountain project in IBM has its own solution for more advanced search result categorization, which has an option to re-label tag structures based on customer label structure in their indexed databases to maximize search efficiency. However, the query of WebFountain is complicated and not user-friendly for individual users, and is used to serve business customers as of now.

4. What is required for summarization

Since text summarization is a research domain with a large volume of research studies, therefore the major effort is to couple it with search engine and fine-tune the integrated product and service to achieve a user acceptable level. One of the biggest issues in text summarization is that the automatic generalized article is far inferior to human-written article. Therefore, the aim is to provide a user acceptable summary with accurate information despite of its un-human writing style.

In order for the user to customize the summarized search results, a regular article structure with preferred outline maybe an option through a user-friendly interface even though it may not be as convenient as the Minipen. Another potential improvement is that the search base should not be limited to certain products or formats, such as in the Minipen case. In other words, a truly universal and cross-platform Internet data source is the ideal scope. Regarding mobile device, an everlasting battery power is needed, especially for the above-mentioned case of the user being on the road, as shown in Figure 1, too. Even though solar-based system usually replaces only 30% of the original power source, the solar-based battery should still be incorporated into the design of the mobile devices to address this availability issue.

5. Business models

Since the Internet search result summarization depends on the search engine, the business model can be considered for search engine companies and non-search engine companies. For search engine companies, they can develop their own search result summarizing capability, such that a vertical integration of technologies such as information gathering, indexing, retrieval, aggregation, summary etc. can be streamlined. Similar to IBM’s WebFountain, insightful Web semantic can be integrated into the overall technology development. The profit model may remain unchanged or expand into the pay per use or membership value-added services.

For non-search engine companies, they have to rely on existing search engines for providing the Internet search functionality. However, a new intermediate business supplier can develop a system interface for users to customize the search and summary parameters for a summary instead of the ranked search results. The profit model can be serving its own Web or selling a search engine plug-in program to individuals and business corporations for advertisement, member, or usage fees.
Like the business model of iPod by Apple computers, other peripheral products or services can be packaged together to make the automatic summarizing of search results a popular product or service. As in the example above, the variations of the above-mentioned business models can be claiming one interface and multiple search engines, including a solar-based mobile phone that is always powered, or emphasizing instant knowledge everywhere and anytime. The potential of this IT-enabled service may be of no boundary with well-thought business models.

6. Conclusions

In this article, an incremental innovation of search engine has been described as achievable short-term technology integration for potentially profitable business opportunities. Although automatic summarization of Internet search results may not be as readable as human-written article, it does extend the Internet search engine capability for saving user screening effort. Accordingly, search engine companies or new intermediaries can base on this incremental innovation to develop their strategies under various business models. The market may be expandable if different companies compete with each other using diversified IT-enabled services for discovering hidden consumer wants.

7. References
