Max See Chin Woon
Trinity College, University of Cambridge, Cambridge CB2 1TQ, UK.
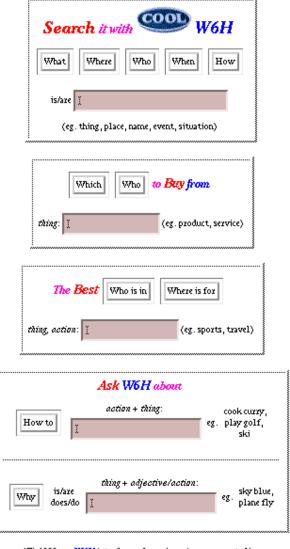Email: scw21@damtp.cam.ac.uk   Tel: (44) 1223 35 92 93

Search Engine Interface and Indexing Problem

Scenario:

Using Internet search engines to find relevant answers to a specific query or for data mining has been a frustrating information overloading experience for many Internet users. Most of the links in the search results returned by a search engine are links to Webpages with contents irrelevant to the question a user has in mind when posing the query to the search engine. To date, search engine companies are either just breaking even or making losses due to the limited revenue generated from the current Internet advertising business model, and they continue to search for a solution to address these pressing unsolved problems.

Solution:

The new W^6H search engine interface and indexing architecture.

Potential Clients:

The search engine development community in the search engine race for functionality and popularity, and Fortune 1000 media and publishing corporations currently seeking effective Internet portal strategy.

This solution is strong in substance for proposals and recommendations to top search engines (Excite, Infoseek, Lycos, Yahoo!) to solve their top problem in delivering "query-results relevance", and to media and publishing corporations seeking to establish more than Internet presence with effective Internet portal strategies.

Objective:

Increase the relevant information to junk information ratio in the results of a search query by using returned by a search engine.

Description of Solution:

Search W^6H is an intelligent context-based search engine interface and indexing architecture. W^6H stands for What, Where, Who, Which, When, Why and How.

The indexing approach of W^6H is specifically suitable for search engines serving the Internet masses such as Excite, Infoseek, Lycos and Yahoo! The approach gives a significant competitive advantage to search engines that adopt it.

According to general survey, when one wants to use a search engine, for most of the time one already has a particular question in mind and wish to look on the Internet for answers to the question.

We shall make use of the following observation:
All search queries can be broken down as questions that can be classified into 7 categories.

Each basic query is then a question of one of the following:
What, Where, Who, Which, When, Why and How.

(To verify this observation, try to see if you can think of an exception or a counterexample).

For example, a college student wants to learn more about the history of US Independence Day via the Web.

Coincidentally, a journalist is writing an online news article about and a brief historical account of the US Independence Day. The journalist, being the author of the Webpage, can think of questions that readers might ask which the Webpage can provide the answers, and then establishes the contexts for keywords which the author wish to be indexed in the following format:

"US Independence Day"
- When $\rightarrow$ indexed to the keyword "July 4" that appears on the Webpage
    - $\rightarrow$ answer to the possible search query of "When is the Independence Day?"
- What $\rightarrow$ indexed to the entire Webpage
    - $\rightarrow$ answer to the possible search query of "What is US Independence Day?"
- Who $\rightarrow$ indexed to the keywords "Thomas Jefferson" or a portrait image that appears on the Web page
    - $\rightarrow$ answer to the possible search query of "Who were the key people involved?"
- How $\rightarrow$ indexed to the section with the history of how independence was won.
    - $\rightarrow$ answer to the possible search query of "How did it happen?"

The questions of Where, Which and Why of US Independence Day may not be meaningful or may not be covered by the Web page and so may be left out.

When the college student keys in the search query keywords "US Independence Day" and clicks on one of the When, What, Who and How buttons, the search engine is able to return relevant and meaningful search results answering the question the student has in mind, and the

Web pages linked from the search engine results will be scrolled to the indexed keyword, block or image.

Thus, instead of just indexing the occurrence of certain keywords on a Web page, we have an indexing mechanism for attaching a corresponding context of a keyword or key phrase to not only another keyword but a particular block of text or images or an entire section on the Web page, or even the entire Web page itself, and for the search engine to directly deliver meaningful search results given a search query.

The "How" context is most useful for Web pages of operating manuals, guides to various activities (sports/games), and general instructions.

The "Why" context is most suited to Web pages of online encyclopaedia, Science, and general Q&A.

http://www.damtp.cam.ac.uk/user/scw21/w6h/   shows a sample layout.

In essence, the distinct architecture of this approach of indexing is its direct focus on answering possible queries given the keywords and items on a Web page so that search engine can deliver meaningful search results directly and thus effectively, rather than a focus on the attempt to build a conceptual knowledge base of the world and hope that Artificial Intelligence algorithms can make sense of the text on Web pages for indexing, and deduce correspondence between the resulting indices given a search query.

Strength:

W^6H will adopt a business model different from the banner ads to be described below.

The Strategy and Competitive Advantage:

The current META tag method allows Web authors to construct a list of relevant keywords but this keyword-listing mechanism does not capture the context of these keywords to make the search context-based and more intelligent.

The advantage of W^6H is that the category of "What, Where, Who, Which, When, Why and How" adds a context to a search query, making the search results more relevant and closer in context to what a search engine user has in mind when posing the query. The proposed standard XML is adopted in W^6H to allow Web authors to add context meaningfully.

Scenario:
Keying in "Company A" and clicking on "What" give links to Webpages with the corporate profile of Company A.
Clicking on "Where" instead gives links to Webpages with the address/map of Company A.

The "Which/Who to buy from" section is the prime revenue generator. Instead of having banner advertisement, it goes to a new Webpage with either Top 3 or Top 5 choices of products or services providers.

Scenario:
Keying in "book" in the "Which/Who to buy from" entry and clicking on "Who" give a Webpage with the top part showing the icons of the popular online bookshops, eg. Amazon.com, Noble & Barnes, etc. who spend significant amount of ads dollars at various websites. Instead of the conventional "per CPM hit" rate of charging banner ads, an ads model which does not generate revenues efficiently, Search W^6H will adopt a different business model for charging ads. The model will use a bidding mechanism: the online bookshop which bids and pays highest will get its icon listed first to the left, and so on.

Below these icons is an entry for title of book. Keying in the title and clicking, say, the Amazon.com icon in the first position runs a script in Search W^6H. The title of book will get

parsed into a proper format for searching Amazon.com catalogue transparently. The script then processes and extracts results, calls up the Amazon.com order form, and outputs the processed HTML onto the screen with title of book filled in as well as the delivery address, credit card info, etc (data registered when first using Search W^6H). This saves the hassle of searching the book and typing the ordering and delivery data manually.   If one is not satisfied with Amazon.com prices, one can click on the icon of rival online bookshop, e.g. Nobles & Barnes.

By adopting this strategy, one pitches the vendors against one another to be the first among the choices in grabbing the chance to get e-business through the Search W^6H, and this in turn will generate for the users a value-added service that will results in high satisfaction and approving recommendation to other users (a more effective form of marketing), and for Search W^6H many times higher ads revenue than would have been by using the conventional banner ads charging model.

Similarly, keying in "CD" and clicking "Which to buy from" will bring up a Webpage with icons of CDnow, Music Boulevard and a entry for album title/single title/name of performers. Keying in "airline ticket" and clicking "Which to buy from" gives a Webpage with icons of online travel agencies and airlines and an entry for departure location, single or multiple destinations, travel date and time. Keying in "car" and clicking "Which to buy from" gives different car models from different manufacturers, while clicking "Who" to buy from gives different car retailers with different offers. Thus, by adopting the strategy, one effectively pitches the online CD retailers against one another, the online travel agencies against one another, the car manufacturers against one another in advertising their cars, and the car retailers against one another in promoting their offers and services.

Some corporations may have bigger financial muscles or marketing aggressiveness than their rivals, and so will bid and spend significantly more to secure the first icon slot ahead of competitors to win e-business.

E-Commerce: Drive-in Takeaway Shopping

Scenario:

Current e-commerce lacks economical logistics and incurs a high cost in the delivery of goods ordered online. The goods are often delivered by costly Fed-Ex, UPS and other local courier or parcel services. A minimal delivery charge makes ordering low-cost goods in small quantities a poor bargain since conducting e-commerce for such goods ends up yielding a low value/cost ratio. Perishable goods such as groceries find no low-cost same-day delivery service. The high delivery cost often makes it difficult to justify the convenience. Deliveries of perishable goods to home often take place at office hours when the customers ordering online are still in offices and not present at home to receive the deliveries. Consumers in general are still hesitant to key in credit card information over the Internet.

Solution:

Drive-in Takeaway Shopping — a Web-based drive-in takeaway retail solution.

Potential Clients:

Chain supermarkets and retail outlets, warehouses and credit payment processing centres.

Objective:

To address the problems of goods delivery in e-commerce, and make e-commerce work worldwide without the security drawbacks of online payment.

Description of Solution:

Step 1:
Order at home on a computer in the previous evening or the early morning before going to work, or the evening after work before heading home using a mobile email/web-enabled phone such as Nokia Communicator 9000. The order may be an email of a recipe/grocery list to a small or medium-size grocery shop in the neighbourhood, or an on-line purchase list built up through a Webpage of a chain supermarket or warehouse. The groceries or purchased goods will be gathered into carry bags.

Step 2:
Drive in after work or *en-route*. The carry bags will be loaded by the staff into the back of the car. Payments can be made at the drive-in counter rather than settled by credit card information over the Internet if preferred. In one stroke of a transformation and differentiating strategy for the retail industry, one takes advantage of online ordering in eliminating the hassle of locating goods, queuing at cashier counters and carrying goods into cars without the worries of the security online transactions.

Spin-off Advantages:

- Use pattern recognition algorithms on the database of the history of on-line ordering info to predict usage and remaining supply of mundane grocery items (Coke, coffee, milk, sugar, etc) and suggest and prioritise items needed regularly for the next on-line order.
- Option of in-house delivery vans for city and suburb deliveries, and partnership with courier services for remote deliveries --- free delivery to home for orders above a minimum payment --- for buyers who commute by public transport (bus or underground tube). 90% of Singapore and Hong Kong population live in condos and centralised flats, making such delivery services commercially viable.

Additional Revenue Source:

Commissions over transactions for order-payment processing centres (at very low rate to promote wide-usage but deriving substantial revenue by large volume of transactions).

Surfies and Webbies

Scenario:

A consultant at a remote client site needs some files and data from headquarters office. The consultant connects the mobile phone to the laptop and dials up but finds that there is no mobile phone transponder within range at the remote site for making a connection call. Similar scenarios may occur for an executive or a field engineer at a remote site.

The advent of the satellite phone is a solution. However, one has to carry both the satellite phone and laptop separately. The solution can be taken one step further in improvement.

Solution:

Convergence of Mobile/Satellite Phone and Internet-enabling Appliances.
The idea is to integrate mobile/satellite phone capability into laptops and palmtops.

Objective:

To create an entirely new and large market segment for mobile/satellite phone operators (Iridium, ICO, Globalstar) and OEM partners to substantially increase the revenue generation, and take the leadership of innovation in telecommunications throughout the next Millenium.

Cost-effective Manufacturing Advantages for OEMs:

- Since a laptop or palmtop has battery, display screen, software and input features, integration of mobile/satellite phone capability into a laptop or palmtop eliminate the hardware duplication of battery, display screen, buttons, and software duplication of memory functions of a standalone mobile phone.
- OEMs can choose to subsidise such integration cost by soliciting commissions from mobile/satellite phone network operators for OEMs to pre-configure default subscription choices of the built-in mobile/satellite phones in laptops and palmtops before shipping. The net effect is that this will predetermine most customers' first choice of mobile/satellite phone network operators. If a mobile/satellite phone operator initiates this proposal to its partners of mobile/satellite phone manufacturers, it will enjoy the goodwill of being pre-selected as the default choice of mobile/satellite phone network, and thus establishing a secured market share in this segment over competitors.

Marketing Campaign:

- When a potential customer using both mobile/satellite phone and Internet is presented with the myriad choices of mobile/satellite phone network operators and Internet service providers, a one-stop gateway will attract the customers to sign up or switch from their existing operators. A partnership between satellite and mobile phone network operators is in a position to aim to be that one-stop gateway.
- '*Surfies*' and '*Webbies*' shall be adopted as the catchy, trendy, easy-to-remember tradenames with potential popularity and acceptance by general public to promote the laptops and palmtops with mobile/satellite phone capability respectively.

Spin-off Advantages:

- Prime Internet-via-satellite competitor Teledesic currently offers Internet-in-the-sky with a large and bulky receiver box to be installed at the rooftop. However, it lacks mobility, convenience and universal portability. A partnership between satellite and mobile phone network operators is in a position to offer mobile Internet-via-satellite with *Surfies* and *Webbies*.

Action Plan for Strategy Adopter:

Establish partnership with computer and laptop OEMs (Acer, Dell, Compaq, Toshiba, Gateway, NEC) to manufacture *Surfies* and *Webbies* based on JavaOS and Windows CE.

- Introduce merchandise models of *Webbies* (casing designed with golf embedding for executives, Disney or Warner or Japanese cartoon-characters for children, Pop/Movie stars for mass-markets, Star Wars or Star Trek characters for niche-markets)
- Add a pair of Infra-red transmitter and receiver diodes to *Webbie* for the functionality of a universal remote control which simulates all the remote controls for home entertainment appliances (TV, VCR, Hi-fi — one *Webbie* replaces all remote controls) and interacts with a computer server in an intelligent home.
- Study feasibility of using LEP (Light Emitting Polymer — a low-cost plastic that emits light) screen display from Cambridge Display Technology-Seiko-Epson for *Webbies*.

Intuitive financial market news delivery system: Java applet for equity, bonds, options, forex news

Scenario:

Survey the existing equity, bonds, options and forex time series charts on the Reuters, Bloomberg and Bridge terminals and on the Internet (e.g. quote.yahoo.com, www.eschwab.com, www.webstreetsecurities.com, www.etrade.com, www.datek.com). The charts show the peaks, troughs, up-trends and downtrends – how the prices of equity, bonds, options and rates of forex vary with time, but these peaks/troughs and up/down-trends are meaningless patterns without the information in relevant news. In all the existing systems, for instance, news is itemised chronologically in a list separate from the time series charts. Analysts and traders have to check the date corresponding to a peak or trough, scroll through the list of news to find a news item with the corresponding date, and click on the item before they can read the relevant news (e.g. news of successful merger reflecting the up-trend to the peak, or stock-split reflecting the downtrend to the trough). Such tracking of news relevant to the time series charts is time-consuming, tedious and non-intuitive.

Solution:

Use the interactivity of a Java applet to provide a solution that will make equity, bonds, options and forex news delivery on trading terminals or on the Internet top-of-the-breed.

Potential Clients:

Investment banks spanning from Wall Street, Chicago and London to Tokyo and Singapore, online brokerage firms, Internet portal Website and financial market research firms providing information to fund managers, institutional investors and private investors.

Description of Solution:

The key elements of the design and implementation are the following:
Embed a Java applet showing a time series chart with a news window on a Webpage. When the mouse pointer is moved near to a peak or trough of a time series chart, the news relevant to the peak or trough will appear in the news window. As the mouse pointer is moved away from the peak or trough, the news relevant to the downtrend or up-trend will appear in the news window. Such interactivity provides analysts, traders, sales, and even corporate and private clients (through Internet) an intuitive and instantaneous access to the tracking of relevant news on any specific equity/bonds/options/forex or indices, giving an edge over the existing state of services offered by competitors.

http://www.damtp.cam.ac.uk/user/scw21/java/news.html   shows a sample demo.