

GIẢI PHÁP PHÁT TRIỂN VÀ NÂNG CAO VỊ THẾ CỦA CÁC CÔNG CỤ TÌM KIẾM VIỆT NAM TRÊN MẠNG INTERNET

Nguyễn Trần Hưng *

Công cụ tìm kiếm đã có một lịch sử phát triển và vị thế quan trọng ngay từ khi bắt đầu mạnh mẽ sự phát triển của Internet và Web. Tại sao lại vậy? Câu trả lời có lẽ nằm ngay trong tính năng tìm kiếm của chính các công cụ này. Với một kho tàng thông tin gần như vô tận trên Internet không được xử lý bằng một hệ thống hợp chuẩn nào và nội dung thông tin luôn luôn được cập nhật, bổ sung thì cách duy nhất để tìm kiếm, tập hợp các thông tin này đó là thông qua các công cụ tìm kiếm. Với thế giới, công cụ tìm kiếm đã có chỗ đứng vững chắc với tính năng vượt trội và thu hút người sử dụng. Ở nước ta, mặc dù đã dần khẳng định được chỗ đứng, tìm được hướng đi đúng đắn là tập trung vào các lĩnh vực tìm kiếm chuyên biệt song do những khó khăn về công nghệ tìm kiếm, về tài chính, về bản quyền... nên các công cụ tìm kiếm của ta chưa khẳng định giá trị của mình trên thị trường tìm kiếm trong nước và quốc tế. Do vậy, để có thể khắc phục những tồn tại, đáp ứng được khoảng trống của thị trường tìm kiếm trong nước mà các công cụ tìm kiếm mang tầm vóc quốc tế như Google hay Yahoo! chưa đáp ứng được thì không chỉ đòi hỏi những nỗ lực rất lớn từ phía các nhà cung cấp dịch vụ này mà còn rất cần đến sự hỗ trợ đặc lực từ phía Nhà nước và cộng đồng.

1. Khái quát về công cụ tìm kiếm

Internet và World Wide Web có đến hàng trăm triệu website chứa đựng các thông tin về nhiều chủ đề khác nhau. Khi người tìm kiếm cần biết về một chủ đề nào, họ sẽ sử dụng một công cụ tìm kiếm trên Internet. Công cụ tìm kiếm trên internet là các website đặc biệt, được thiết kế để giúp mọi người tìm kiếm thông tin được lưu trên các site khác nhau.

Công cụ tìm kiếm được cấu thành bởi ba bộ phận cơ bản có sự độc lập tương đối với nhau về mặt hoạt động nhưng lại có sự gắn kết với nhau, phụ thuộc lẫn nhau về mặt dữ liệu, đó là:

- Robot hay còn được gọi là bộ thu thập thông tin

Robot là một chương trình tự động duyệt qua các cấu trúc siêu liên kết để thu thập tài liệu một cách đệ quy, nó nhận về tất cả các tài liệu về một chủ đề. Robot được biết đến dưới nhiều tên gọi khác nhau: spider, web wanderer hoặc web worm,... Về bản chất robot chỉ là một chương trình duyệt và thu thập thông tin từ các site theo đúng giao thức web. Những trình duyệt thông thường không được xem là robot do thiếu tính chủ động, chúng chỉ duyệt web khi có

sự tác động của con người.

- Index hay còn được gọi là bộ lập chỉ mục

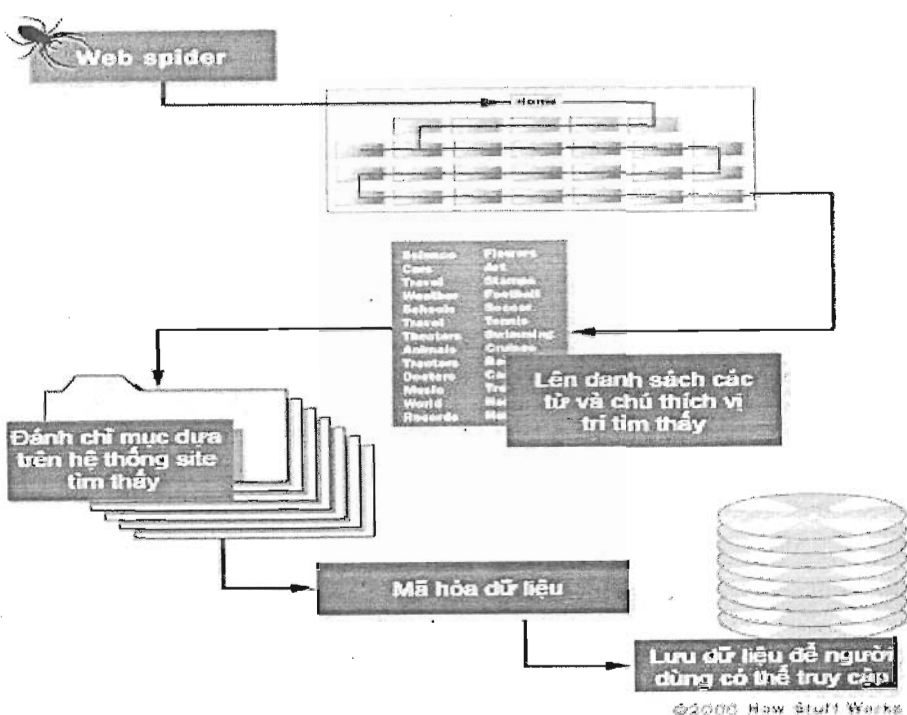
Hệ thống lập chỉ mục hay còn gọi là hệ thống phân tích và xử lý dữ liệu sẽ thực hiện việc phân tích, trích chọn những thông tin cần thiết (thường là các từ đơn, từ ghép, cụm từ quan trọng) từ những dữ liệu mà robot thu thập được và tổ chức thành cơ sở dữ liệu riêng để có thể tìm kiếm trên đó một cách nhanh chóng, hiệu quả. Hệ thống chỉ mục là danh sách các từ khóa, chỉ rõ các từ khóa nào xuất hiện ở trang nào, địa chỉ nào.

- Search Engine hay còn gọi là bộ tìm kiếm thông tin

Bộ tìm kiếm thông tin tương tác với người sử dụng thông qua giao diện web, có nhiệm vụ tiếp nhận và trả về những tài liệu thỏa yêu cầu của người sử dụng. Nói rõ hơn, tìm kiếm từ là tìm kiếm các trang mà những từ trong câu truy vấn (query) xuất hiện nhiều nhất, ngoại trừ các từ quá thông dụng như mạo từ a, an, the. Một từ càng xuất hiện nhiều trong một trang thì trang đó càng được chọn

* ThS., Trường Đại học Thương mại

để trả về cho người dùng, một trang chứa tất cả các từ trong câu truy vấn tốt hơn một trang không chứa một hoặc một số từ. Ngày nay, hầu hết các bộ tìm kiếm thông tin đều hỗ trợ chức năng tìm cơ bản và nâng cao, tìm từ đơn, từ ghép, cụm từ, danh từ riêng hay giới hạn phạm vi tìm kiếm như trên để mục, tiêu đề, đoạn văn bản giới thiệu về trang web,...



Ngoài chiến lược tìm chính xác theo từ khóa, các bộ tìm kiếm thông tin còn cố gắng hiểu ý nghĩa thực sự của câu hỏi thông qua những câu chữ do người dùng cung cấp. Điều này được thể hiện qua chức năng sửa lỗi chính tả, tìm cả những hình thức biến đổi khác nhau của một từ. Ví dụ: bộ tìm kiếm thông tin sẽ tìm những từ như speaker, speaking, spoke khi người dùng nhập vào từ speak.

Nguyên tắc hoạt động của công cụ tìm kiếm:

Search engine điều khiển robot thu thập thông tin trên mạng thông qua các siêu liên kết (hyperlink). Khi robot phát hiện ra một site mới, nó gửi tài liệu (web page) về cho server chính để tạo cơ sở dữ liệu chỉ mục phục vụ cho nhu cầu tìm kiếm thông tin. Bởi vì thông tin trên mạng luôn thay đổi nên robots phải liên tục cập nhật các site cũ. Mật độ cập nhật phụ thuộc vào từng hệ thống search engine. Khi search engine nhận câu truy vấn từ người sử dụng, nó sẽ tiến hành phân tích, tìm trong cơ sở dữ liệu chỉ mục và trả về những tài liệu thỏa yêu cầu.

2. Thực trạng tìm kiếm của các công cụ tìm kiếm Việt Nam

Từ tháng 11/1997 đến nay, đã hơn 12 năm internet có mặt tại Việt Nam. Với hoài bão đưa các websearch Việt vào hoạt động phục vụ nhu cầu tra cứu, những nhà cung cấp dịch vụ trực tuyến Việt Nam đã có nhiều nghiên cứu và cách đây 5 năm, sự ra đời của PanVietnam, Hoatieu hay Vinaseek... được xem là hình mẫu cho hướng phát triển các công cụ tìm kiếm trên internet tại Việt Nam thì đến nay, đã có khá nhiều những trang web kiểu này chính thức được công bố. Nửa cuối của năm 2007, chúng ta đã chứng kiến sự phát triển đến mức khó tin của các công cụ tìm kiếm thuần Việt. Theo số liệu không chính thức, hiện có hơn 30 công cụ tìm kiếm của Việt Nam đang hoạt động ổn định, song chỉ có khoảng 20% trong số này được xem là thuần Việt thực sự, trong đó phải kể đến: 7 sac.vn, xalo.vn, bamboo.com.vn, socbay.com, monava.vn,... Một cuộc đua mới trong lĩnh vực tìm kiếm mà tưởng như Yahoo!, Google đã thấu tóm toàn bộ thị phần Việt Nam đang hình thành. Điều đặc biệt, hầu như các nhóm phát triển công cụ tìm kiếm trong nước đều bao gồm những thành viên còn rất trẻ, các ý tưởng và

công nghệ hoàn toàn đi theo một hướng mới.

Thành công quan trọng nhất đối với các công cụ tìm kiếm Việt Nam đó là đã tìm được hướng đi đúng đắn - tập trung vào các lĩnh vực tìm kiếm chuyên biệt. Theo ông Phạm Trung Dũng - Sáng lập viên kiêm Giám đốc Công ty cổ phần 7 Sắc (<http://www.7sac.com>) thì "sẽ chẳng khác gì chuyện "trúng chọi đá" khi phải đương đầu với Google, Yahoo!. Nhưng nếu chỉ nhắm vào một phân khúc thị trường cụ thể, hoàn toàn có lý do để các công cụ tìm kiếm Việt Nam tự tin".

Nổi trội trong các công cụ tìm kiếm nhạc là Bamboo, một hệ thống tìm kiếm với hai chức năng tìm kiếm chính là mp3 và video. Bamboo tìm kiếm video khá tốt với khả năng lưu trữ file video nguồn về máy chủ của mình nhằm gia tăng tốc độ xem video clip, cũng như hỗ trợ nhiều các tiện ích đi kèm dành cho người dùng như giỏ video, hỗ trợ lưu trữ bản ghi. Nhưng điểm thu hút người dùng nhất của Bamboo là nhờ giao diện và cách hiển thị kết quả cực kỳ đơn giản, dễ hiểu.

Tuy nhiên chức năng tìm kiếm nhạc phổ thông của Bamboo lại không ấn tượng bằng 7 Sắc (<http://7sac.com>) - một hệ thống tìm kiếm với 4 chỉ mục là tìm kiếm tin tức, nhạc, video và blog. 7 Sắc được nhiều blogger cũng như người dùng yêu nhạc và phim lựa chọn bởi kết quả tìm kiếm vô cùng phong phú, với giao diện thiết kế khoa học và kết quả tìm kiếm được tổng hợp từ rất nhiều nguồn. 7 Sắc cũng hỗ trợ người dùng có thể tải bài hát trực tiếp, nhúng vào blog, chia sẻ với bạn bè hay lưu trữ kết quả tìm kiếm ưa thích...

Thế nhưng, để có thêm thông tin chi tiết về một ca khúc đang nghe hay nghe các ca khúc trong cùng một album, Zing Mp3 (<http://zing.vn>) của Công ty Vinagame lại là hệ thống tìm kiếm tốt nhất với cơ sở dữ liệu nhạc được biên tập và tổ chức, khả năng tìm kiếm theo nhiều tiêu chí, sắp xếp theo dạng album, ca sĩ...

Công ty Cổ phần Dịch vụ Công nghệ thông tin Naiscorp vừa chính thức công bố đã ký kết đầu tư với hai quỹ đầu tư mạo hiểm lớn trên thế giới là IDG Ventures Vietnam và Softbank, đồng thời chính thức ra mắt cổng thông tin tìm kiếm tiếng Việt trực tuyến www.socbay.com. Công cụ tìm kiếm Socbay chạy thử nghiệm từ năm 2002, đến nay được khá nhiều người tiêu dùng tin nhiệm. Hiện tại, Socbay thu hút số lượng truy cập mỗi

ngày khoảng 600.000 người. Socbay cung cấp các dịch vụ tìm kiếm theo chiều dọc như: nhạc, từ điển, hình ảnh, tin tức, web, rao vặt... Socbay ngoài khả năng tìm kiếm video và mp3 miễn phí, còn có thêm chỉ mục tìm kiếm lời bài hát và tìm kiếm địa điểm.

Mặc dù các công cụ tìm kiếm Việt Nam đã đạt được những thành công quan trọng ban đầu, tạo ra các tiền đề về bản sắc, dấu ấn riêng cho người sử dụng Việt, nhưng đó mới chỉ là một chặng đường rất nhỏ trong công cuộc khai thác thị phần tìm kiếm cũng như hoàn thiện các công cụ tìm kiếm này cho phù hợp hơn nữa, tiện ích hơn nữa với đòi hỏi của người sử dụng trực tuyến. Trong giai đoạn hiện tại, khi đang chuyển sang việc hoàn thiện và thương mại hóa, đa phần các công cụ tìm kiếm thuần Việt đều có rất nhiều tồn tại cần được giải quyết, cụ thể:

- *Tồn tại đầu tiên là về vốn mở rộng dự án.* Để có một trang web tìm kiếm thực thụ, cần phải có nhiều server. Người làm công cụ tìm kiếm cần phải có một mạng lưới server lớn, thiết kế được một phương thức lưu trữ dữ liệu và kết hợp dữ liệu trả về thật tốt. Chính vì vậy chi phí là vô cùng lớn. Trước thực tế hầu hết các công cụ tìm kiếm thuần Việt đều rất khó khăn về vấn đề tài chính để có thể nâng cấp phiên bản tìm kiếm của mình thì thử thách ngày càng chồng chất hơn khi những người khổng lồ như Google và Yahoo! bắt đầu mua lại các công cụ tìm kiếm tốt nhất của Việt Nam như họ đã từng làm với rất nhiều công ty triển vọng khác trên khắp thế giới để có thể thầu tóm cả công cụ tìm kiếm lẫn người dùng chỉ trong một thời gian ngắn.

Theo dự báo của các chuyên gia, với một khoản tiền để nghị hấp dẫn và đang trong tình thế tiến thoái lưỡng nan do bị eo hẹp về nguồn vốn cũng như đầu ra, sẽ có không ít doanh nghiệp Việt Nam bị dao động mạnh về tâm lý.

- *Tồn tại thứ hai là sự yếu kém về công nghệ.* Đa số các công cụ tìm kiếm trong nước sử dụng công nghệ nước ngoài hoặc phát triển trên nền cơ sở dữ liệu API của Google, Yahoo!, hay các phần mềm mã nguồn mở Lucene, Sphinx... Điển hình nhất đó là vụ công cụ tìm kiếm Monava chỉ là một trang meta search (lấy kết quả tìm kiếm từ các công cụ tìm kiếm khác) sử dụng cổng lập trình API đưa từ khóa cần tìm

lên Google và Picsearch. Sau khi hai trang này trả kết quả về Monava sẽ bóc tách rồi cho hiển thị kết quả. Monava không dám đặt máy chủ ở Việt Nam vì tình trạng nghẽn cổ chai ở cổng ra Internet quốc tế của Việt Nam sẽ khiến việc kết nối đến máy chủ Google và yêu cầu tìm kiếm không thực hiện được. Sau một thời gian chời bỏ mình là một meta search, cuối cùng công ty Monava đã phải thêm vào kết quả tìm kiếm sau khi trả về của mình dòng chữ: "Power by Google". Đây là một bài học lớn cho các công cụ tìm kiếm Việt, khi sử dụng công nghệ trên nền của người khác mà không có sự đầu tư tạo điểm độc đáo riêng thì sẽ khó có được chỗ đứng đối với người sử dụng trực tuyến chứ chưa nói đến thị trường tìm kiếm thuần túy.

- *Tồn tại thứ ba là vấn đề bản quyền.* Theo khảo sát ban đầu, rất nhiều trang trong số các công cụ tìm kiếm nhạc ở Việt Nam có thể tự upload nhạc cho người dùng tìm kiếm, chứ không phải thực sự xây dựng các công cụ tìm kiếm mp3 vì phần lớn các bài hát tìm kiếm được trên các trang này khi download về máy không có dữ liệu hoàn chỉnh - tức là ngoài tên bài hát còn phải có tên album, tên ca sỹ... Một đại diện trong nhóm phát triển nhacso.net cho rằng hầu hết các bài hát trên nhacso chỉ up lên đoạn nghe thử, tức là chỉ khoảng 45 giây/bài hát nhưng khi tìm kiếm trên một vài trang tìm kiếm nhạc, các trang này cho biết nguồn lấy từ nhacso, mà lại có toàn bộ bài hát kéo dài đến khoảng 5 phút/bài, điều này cho thấy một số trang tự upload nhạc lên nhưng nói là tìm kiếm để né tránh vấn đề bản quyền. Trong tương lai, bản quyền sẽ trở thành vấn đề lớn đối với các trang tìm kiếm âm nhạc mới xuất hiện do họ có lưu trữ và cho người dùng download miễn phí về máy tính.

- *Tồn tại thứ tư là khả năng thu hút quảng cáo của các trang web tìm kiếm bằng tiếng Việt.* Hàng loạt trang web tìm kiếm Việt được đầu tư lên tới hàng triệu USD nhưng đến nay hoạt động vẫn chưa có doanh thu. Nguồn thu của các trang web tìm kiếm này chủ yếu là từ quảng cáo nhưng hiện chỉ có các doanh nghiệp nước ngoài mới có khái niệm về quảng cáo trên các công cụ tìm kiếm, còn các doanh nghiệp trong nước chưa quan tâm tới mảng quảng cáo này. Bởi nhận thức của đại bộ

phần các doanh nghiệp trong nước chưa cao và chủ yếu vẫn tập trung vào hình thức quảng cáo trên một không gian web cụ thể mà ít biết hoặc chưa biết đến lợi ích của quảng cáo thông qua các công cụ tìm kiếm.

- *Tồn tại thứ năm là khả năng tương tác toàn cầu so với các công cụ tìm kiếm đã có chỗ đứng vững chắc trên thị trường tìm kiếm như Google, Yahoo!, MSN...* Có một thực tế sử dụng Google đã trở thành thói quen của hơn 2/3 người dùng công cụ tìm kiếm trên toàn thế giới và ở Việt Nam là 80%. Nhìn chung, Google được xem là công cụ tìm kiếm tốt nhất, mặc dù kết luận này chỉ mang tính cảm quan. Công cụ tìm kiếm thuần Việt có thể tốt ngang với Google hoặc thậm chí tốt hơn mặc dù Google liên tục cải thiện và bổ sung thêm các chức năng, tuy nhiên có một thứ văn hóa đã hình thành xung quanh Google. Vấn đề nổi cộm trong lĩnh vực kinh doanh công cụ tìm kiếm là tương lai của lĩnh vực này hầu như không có gì mới để khám phá dù các kết quả tìm kiếm có chính xác hơn. Những thông tin mà Google, Yahoo!... đem lại đã quá đủ cho phần đông số người sử dụng muốn tìm kiếm thông tin trực tuyến, điều này cũng giống như đối với hành khách đi máy bay, việc bay ở độ cao 10.000 m và 11.000 m cũng chẳng có gì khác nhau, độ cao có tăng thêm nữa đối với họ cũng chẳng có ý nghĩa gì, họ chỉ cần biết là họ bay đủ cao để tới được nơi cần tới.

3. Một số giải pháp nhằm phát triển và nâng cao vị thế của các công cụ tìm kiếm Việt Nam

Với những thành công ban đầu đã đạt được, để có thể khắc phục được những tồn tại kể trên và đáp ứng được khoảng trống của thị trường tìm kiếm trong nước mà các công cụ tìm kiếm mang tầm vóc quốc tế như Google hay Yahoo! chưa đáp ứng được, nhằm khẳng định chỗ đứng cũng như tận dụng các cơ hội thương mại hóa, các công cụ tìm kiếm trong nước cần phải tập trung vào một số giải pháp như sau:

a. *Giải pháp về vấn đề tài chính*

Bất kể một công cụ tìm kiếm nào, dù được xây dựng trên nền công nghệ nào đi chăng nữa, một yêu cầu bắt buộc là phải thiết lập được một hệ thống cơ sở dữ liệu khổng lồ ngừng được cập nhật và gia tăng số lượng chỉ mục. Muốn vậy, phải có một hệ thống máy chủ đủ mạnh, đường truyền, tốc độ, các giải thuật trả về kết quả nhanh và chính xác,

điều này đòi hỏi chi phí đầu tư ban đầu rất lớn. Thực tế cho thấy rất nhiều công cụ tìm kiếm thuần Việt mặc dù được đầu tư với số tiền lên tới 2 hoặc 3 triệu USD nhưng tốc độ tìm kiếm vẫn chậm, số lượng chỉ mục, cơ sở dữ liệu chỉ mục vẫn còn rất ít, như vậy chi phí để cho một công cụ tìm kiếm hoạt động hiệu quả là rất lớn. Do đa phần công cụ tìm kiếm thuần Việt đều tập trung vào lĩnh vực tìm kiếm chuyên sâu và tận dụng lợi thế ngôn ngữ tiếng Việt để khai thác những kẻ hở thị trường mà Google, Yahoo!... chưa bao phủ hết nên mỗi công cụ chỉ tập trung vào một lĩnh vực riêng và cố gắng đáp ứng tốt ở lĩnh vực đó, trong quá trình phát triển, muốn nâng cấp hoặc mở rộng tìm kiếm sang lĩnh vực khác đòi hỏi chi phí rất lớn. Nếu như các công cụ tìm kiếm thuần Việt có được sự hợp tác với nhau thì đương nhiên sẽ phát huy được thế mạnh riêng có trong từng lĩnh vực của mình và khoản chi phí bỏ ra đầu tư cũng nhỏ hơn rất nhiều so với khi nâng cấp hệ thống.

Như vậy, để giải quyết vấn đề tài chính và tận dụng được lợi thế riêng có của từng công cụ tìm kiếm, doanh nghiệp - chủ nhân của các công cụ này cần phải thành lập một hiệp hội các công cụ tìm kiếm thuần Việt, trong đó các thành viên có sự chia sẻ, hợp tác tích cực và phối hợp tìm kiếm chéo với nhau. Chẳng hạn như Bamboo mạnh về tìm kiếm file video, Zing Mp3 mạnh về tìm kiếm nhạc số, có thể có sự giao lưu tìm kiếm với nhau bằng cách thiết lập một cổng tìm kiếm chéo, khi người sử dụng tìm kiếm nhạc số trên Bamboo thì Bamboo có thể gửi yêu cầu tìm kiếm tới Zing và cho hiển thị kết quả tìm kiếm. Ngược lại, người sử dụng có thể tìm kiếm video hoặc các đoạn clip trên chính Zing cũng bằng cách tương tự với yêu cầu được gửi về Bamboo và Zing sẽ cho hiển thị kết quả theo sắp xếp của mình. Với sự hợp tác này, các công cụ tìm kiếm thuần Việt có thể giải quyết được khó khăn về tài chính, vốn là một bài toán nan giải đối với các công cụ tìm kiếm Việt Nam.

b. Giải pháp về vấn đề công nghệ

Mặc dù thị trường các trang web công cụ tìm kiếm trong nước phát triển khá mạnh, có sự khởi sắc mạnh mẽ cả về số lượng và chất lượng nhưng hầu hết đều yếu kém về công nghệ so với các công cụ tìm kiếm nước ngoài.

Đa số các công cụ tìm kiếm trong nước sử

dụng công nghệ nước ngoài hoặc phát triển trên nền cơ sở dữ liệu của các công cụ tìm kiếm sẵn có như API của Google, Yahoo! hay các phần mềm mã nguồn mở Lucene, Sphinx... chứ chưa tự xây dựng được nền cơ sở dữ liệu cho mình. Tuy nhiên, rất nhiều chuyên gia nước ngoài cho rằng nền cơ sở dữ liệu này đều dừng lại ở mức đại cương, phần khung bên ngoài bởi vì những công cụ tìm kiếm hàng đầu thế giới không bao giờ phổ biến hết bí quyết và kinh nghiệm của họ. Nếu như các công ty Việt Nam làm công cụ tìm kiếm chỉ biết sao chép thì chẳng thể nào cạnh tranh với các công cụ tìm kiếm hàng đầu như Google hay Yahoo! được.

Như vậy, để có thể bù đắp được khả năng yếu kém về mặt công nghệ, công cụ tìm kiếm của Việt Nam muốn thành công trên nền công nghệ của các công cụ tìm kiếm nước ngoài khác thì trước hết phải tạo nên nét riêng, độc đáo của mình, tiếp theo là phải đầu tư vào những mảng, lĩnh vực mà Google, Yahoo!... khó xâm nhập vào Việt Nam. Việc mà các công cụ tìm kiếm Việt cần thực hiện ngay là tiếp tục củng cố cơ sở dữ liệu tìm kiếm và xử lý bảng câu hỏi truy vấn tiếng Việt, tận dụng ngôn ngữ bản địa là lợi thế mà các công cụ tìm kiếm toàn cầu như Google hay Yahoo! không thể khai thác hết được. Bên cạnh đó, tập trung vào các lĩnh vực tìm kiếm chuyên sâu hoặc tìm kiếm theo chủ đề, song cũng cần phải lưu ý dù đó là chiến lược đúng đắn nhưng sẽ hạn chế khả năng cạnh tranh với các trang web lớn nên tránh việc đối đầu trực tiếp với các công cụ tìm kiếm toàn cầu.

c. Giải pháp về vấn đề bản quyền

Bản quyền là vấn đề lớn đối với tất cả các công cụ tìm kiếm đặc biệt là đối với các công cụ tìm kiếm thông tin, âm nhạc, video. Rất nhiều chuyên gia tin học trong và ngoài nước cho rằng, nhiều công cụ tìm kiếm nhạc ở Việt Nam có thể tự upload nhạc cho người dùng tìm kiếm, chứ không phải thực sự xây dựng các công cụ tìm kiếm mp3 bởi nếu là hệ thống tìm kiếm thật sự sẽ tốn rất nhiều tài nguyên hệ thống và tài nguyên internet vì phải luôn luôn đi Index các site về lưu trong bộ nhớ của hệ thống, nhưng thực tế đa phần các công cụ tìm kiếm thuần Việt trong lĩnh vực tìm kiếm âm nhạc, video đều có cơ sở dữ liệu nhỏ hẹp, như trường hợp của Monava.

Hiệp hội công nghiệp ghi âm Việt Nam (RIAV)

đã tuyên bố sẽ áp dụng các thủ tục tố tụng cần thiết để yêu cầu các trang web nhạc số, kể cả các công cụ tìm kiếm nhạc vi phạm bản quyền bồi thường thiệt hại. Theo ông Phạm Long Minh - Chánh văn phòng RIAV, hiện có rất nhiều trang web tìm kiếm nhạc số không chịu chi trả tiền bản quyền, sự phát triển của các trang tìm kiếm nhạc có thể sẽ vô tình "giết chết" nền công nghiệp ghi âm Việt Nam. Một tác phẩm âm nhạc vừa phát hành đã có ngay trên Internet và người dùng có thể dễ dàng tìm kiếm, nghe, tải miễn phí chất lượng cao tại các trang tìm kiếm nhạc như: Zing, Bamboo, Socbay, Monava... khiến các hãng băng đĩa thất thu một khoản khá lớn, các hãng đều ngại khi sản xuất các chương trình mới, việc giao lưu hợp tác với quốc tế trong sản xuất, ghi âm cũng khó khăn vì điều kiện tôn trọng bản quyền không được thực thi.

Giải quyết được vấn đề bản quyền là tương đối khó và hiện đang gây rất nhiều tranh cãi ngay cả đối với các website cung cấp âm nhạc. Thực tế hiện nay đã có một số website cung cấp âm nhạc có được bản quyền khai thác thực sự của các đối tác như nhacso.net với cơ sở dữ liệu hiện có trên 20.000 bản ghi và đơn vị này khẳng định đang là đối tác duy nhất có được sự ủy thác của hầu hết các hãng băng đĩa như Việt Tân, Saigon Vafaco, Thế giới giải trí, Hãng phim Trẻ, Lạc Hồng, Trùng Dương, Quang Cường, Tuấn Nguyễn, Tuấn Trinh, Golden fish... Và ở thời điểm hiện tại, doanh nghiệp này chỉ mới cấp phép cho duy nhất Công ty Yahoo! Việt Nam quyền sử dụng các bản ghi trên môi trường Internet, chưa thấy một công cụ tìm kiếm thuần Việt nào của Việt Nam có ý muốn hợp tác, chia sẻ các bản ghi âm nhạc có bản quyền này. Như vậy, để có thể giải quyết vấn đề bản quyền một cách triệt để, các công cụ tìm kiếm thuần Việt trong lĩnh vực tìm kiếm âm nhạc hay video cần phải có sự hợp tác với các website cung cấp nhạc uy tín như nhacso.net, trong đó các kết quả trả về nên có sự trích dẫn đầy đủ nguồn, tên bài hát, tên tác giả, tên ca sỹ... Hơn nữa, sự hợp tác này đối với cả hai bên website cung cấp âm nhạc uy tín có bản quyền như nhacso.net và các công cụ tìm kiếm âm nhạc đều có lợi. Đối với website cung cấp âm nhạc uy tín thì có thể thông qua các công cụ tìm kiếm

để quảng cáo và nâng cao vị trí của mình trong các bảng xếp hạng trên Internet và có thể được một phần phí từ các công cụ tìm kiếm, còn các công cụ tìm kiếm có thể giải quyết được vấn đề bản quyền tác phẩm âm nhạc, thu hút thêm nhiều người sử dụng trực tuyến, đồng thời hệ quả của nó là cải thiện được vị trí tìm kiếm chuyên sâu của mình, do đó sẽ có thêm nhiều nguồn thu từ hoạt động quảng cáo, liên kết từ các website khác.

d. Giải pháp về vấn đề thương mại hóa các công cụ tìm kiếm

Tất cả các công cụ tìm kiếm được xây dựng không phải chỉ nhằm thực hiện các mục đích mang tính chất cộng đồng, xã hội mà mục đích cuối cùng vẫn là lợi nhuận, khả năng thu hút đầu tư, hợp tác, lưu lượng truy cập... hay nói cách khác đó là khả năng thương mại hóa.

Để tìm kiếm một nội dung hay một website, phần lớn người lướt web sẽ nhờ đến công cụ tìm kiếm Internet. Muốn tìm kiếm được thì website đó phải được các công cụ tìm kiếm lưu trong cơ sở dữ liệu chỉ mục của nó. Để tạo ra được điều này, website đó phải nổi bật trong các kết quả tìm kiếm hợp lệ, tốt nhất là ở trang kết quả đầu tiên hoặc ở một trong ba trang đầu (phần lớn người xem kết quả tìm kiếm sẽ không thích mở quá 3 trang). Do đó, nếu như muốn quảng cáo một website nào đó thì có thể đăng ký mua các vị trí trên công cụ tìm kiếm và tốt nhất nên xuất hiện ở các vị trí đầu tiên trong các kết quả trả về để thu hút được sự chú ý của người sử dụng trực tuyến và các khách hàng tiềm năng trên internet.

Như vậy, giải quyết vấn đề thương mại hóa cho các công cụ tìm kiếm thuần Việt, trước hết cần phải thay đổi được nhận thức của các doanh nghiệp kinh doanh trên internet về cách thức quảng cáo website. Cần có sự hợp tác tích cực giữa các công cụ tìm kiếm trong việc viết bài trên các báo hoặc tổ chức các chương trình riêng hay các buổi hội thảo với thành phần tham dự là đồng đẳng các doanh nghiệp kinh doanh trực tuyến để giới thiệu về cách thức quảng cáo trên các công cụ tìm kiếm, phải làm rõ được sự khác nhau cũng như tính hiệu quả của quảng cáo trên công cụ tìm kiếm với quảng cáo trên các website có số lượng người truy cập lớn, chẳng hạn như quảng cáo trên: vnexpress, hoặc dantri.com, vietnamnet.vn... thì khoản

chi phí bỏ ra để mua không gian trên các site này rất lớn, hơn nữa có một thực tế là những người sử dụng khi truy cập vào các website này đa phần chỉ lấy tin tức, ít khi quan tâm đến các quảng cáo. Trong khi quảng cáo thông qua các công cụ tìm kiếm là quảng cáo hướng trực tiếp tới khách hàng có nhu cầu thực sự và chi phí đương nhiên tiết kiệm hơn rất nhiều so với quảng cáo trên các website khác, mà người sử dụng trực tuyến hoặc khách hàng có nhu cầu lại dễ tìm đến hơn.

Điều thứ hai mà các công cụ tìm kiếm thuần Việt cần làm đó là phải thu hút được đông đảo người sử dụng khiến cho các doanh nghiệp kinh doanh trực tuyến quan tâm và muốn có được vị trí tốt trên công cụ tìm kiếm. Để tạo ra sự thu hút này, các công cụ tìm kiếm phải hoàn thiện, nâng cấp cả về độ chính xác trong kết quả trả về khi xử lý các câu hỏi truy vấn, tốc độ tìm kiếm, sự đơn giản trong sử dụng,... nhằm tạo ra sự tiện dụng tối đa cũng như các giá trị gia tăng trong các dịch vụ cung cấp cho người sử dụng trực tuyến.

Kết luận

Với sự ra đời của hàng loạt các website tìm kiếm thuần Việt trong năm 2007 đã đánh dấu một bước phát triển mới trong thị trường tìm kiếm của Việt Nam. Các công cụ tìm kiếm trực tuyến thuần Việt đang dần khẳng định được chỗ đứng và phát triển theo một hướng mới. Tuy nhiên, trong các giai đoạn tiếp theo của việc nâng cấp và hoàn thiện hệ thống tìm kiếm chuyên sâu, có thể nói rằng các công cụ tìm kiếm nội địa gặp rất nhiều khó khăn đến từ cả sự yếu kém về mặt tài chính và công nghệ tìm kiếm. Hơn bao giờ hết, các công cụ tìm kiếm này đang rất cần sự động viên và hỗ trợ kịp thời của cả cộng đồng trong nỗ lực khẳng định giá trị của mình trên thị trường tìm kiếm trong nước và quốc tế. ♦

Tài liệu tham khảo:

1. Jim Davis, Gloria J. Miller, Allan Russell, *Information Revolution 2006*, John Wiley & Son, Inc.
2. Soumen Chakrabarti, *Mining the web*, Team LinG - Live, 2005.
3. Ian Witten & Eibe Frank, *Data mining: Practical Machine Learning Tools and*

Techniques with Java Implementation, Team LinG - Live, 2006.

4. Jiawei Han & Micheline Kamber, *Data mining: Concept & Techniques*, Team LinG - Live, 2006.

5. Ian Witten, Marco Gori, Teresa Numerico, *Web Dragon*, Team LinG, 2006.

6. Vũ Thị Nha, *Internet Searching version 4*, Trung tâm Thông tin phát triển Việt Nam, 2008.

7. Lê Thúy Ngọc, Đỗ Mỹ Nhung, *Tìm hiểu về Search Engine và xây dựng ứng dụng minh họa cho Search Engine tiếng Việt*, Đại học Quốc gia Thành phố Hồ Chí Minh, 2004.

8. <http://www.wikipedia.com>

9. Một số các websearch:

<http://www.google.com>; <http://www.yahoo.com>
<http://www.socbay.com>;

<http://www.bamboo.com>

<http://www.zing.vn>; <http://www.monava.vn>

<http://www.xalo.com>; <http://www.7sac.vn>

Summary

Search engines have long developed and affirmed their important positions along with the development of the Internet and Web. Why? The answer comes from the searching functions of these engines. With a seemingly endless information store which is not appropriately processed by a standardized system while constantly updated, the only way to search and gather the information is through search engines. In the world, search engines have held a firm position in users mind whereas in Vietnam, despite gradually affirming their status by focusing on specialized searching fields, our search engines have still failed to gain a foothold in domestic and international market due to problems related to technology, finance or talent, etc. To solve these problems and better meet the domestic market gap which international search engines like Google or Yahoo! cannot satisfy, not only great efforts from service providers but also supports from the Government and community are needed.