1. **Hướng đi mới**

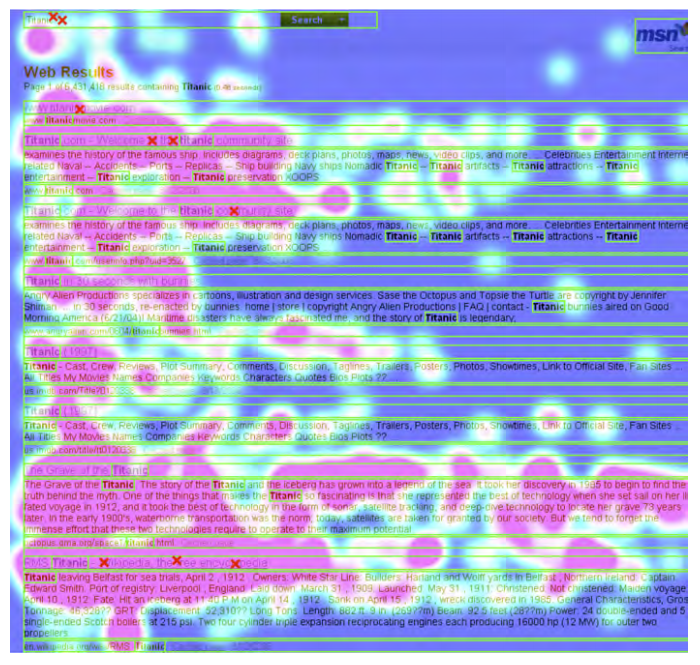
Tiếp theo chúng ta sẽ tìm hiểu về những vấn đề mở cũng như những thách thức trong việc khai phá query log đang được nhiều nhà nghiên cứu quan tâm. Rõ ràng phần này sẽ không đưa ra kết quả nghiên cứu mà các ý tưởng mới sẽ được phác họa một cách ngắn gọn.

* 1. **Theo dấu mắt (Eye Tracking)**

Như chúng ta đã thấy ở phần nâng cao tính hiệu quả của search engine, thông tin phản hồi là một phương tiện rất quan trọng để cải thiện trải nghiệm tìm kiếm của người dùng, bằng cách thay đổi thuật toán xếp hạng tương ứng với những lớp người dùng khác nhau hoặc tại các giai đoạn thời gian khác nhau.

Một loại phản hồi có thể thu được thông qua việc quan sát cách người dùng tương tác với trang tìm kiếm bằng cách theo dấu chuyển động của mắt.

Để ví dụ cho những thông tin bắt được qua kỹ thuật theo dấu chuyển động của mắt, hình 6.1 chỉ ra cách người dùng nhìn vào một trang kết quả tìm kiếm. Dễ nhận thấy người dùng thường đọc phần mô tả của các kết quả tìm kiếm, đặc biệt ở kết quả thứ bảy.



*H6.1 Biểu đồ nhiệt minh họa cho các khu vực của trang tìm kiếm mà ba người dùng khác nhau nhìn vào*

Thực tế, ta có thể nghĩ đến rất nhiều ứng dụng sử dụng loại thông tin phản hồi này, ví dụ như học cách xếp hạng kết quả tìm kiếm tốt hơn hoặc cách đặt quảng cáo để gây ra sự chú ý nhiều hơn tới người dùng, …

Minh họa cho điều này, Guan và Cutrell thực hiện thí nghiệm sau: với một câu truy vấn, các kết quả tốt nhất được thay đổi vị trí hiển thị (với những người dùng khác nhau) ở vị trí trên cùng, ở giữa và ở cuối trang tìm kiếm. Kết quả chỉ ra rằng người dùng càng mất nhiều thời gian và ít thành công hơn khi kết quả cần tìm được hiển thị ở vị trí càng thấp trong danh sách. Khi người dùng không tìm được kết quả cho truy vấn dạng điều hướng (*navigational query* - tìm URL của trang web chứa thông tin), họ thường chọn kết quả đầu tiên, hoặc chuyển sang câu truy vấn mới. Từ đó có thể kết luận như sau:

*“Bộ máy tìm kiếm có thể đưa ra nhiều kết quả tìm kiếm khác nhau với cách xếp hạng tốt nhất không thật sự rõ ràng do vậy người dùng cần đưa ra phán đoán về kết quả cần tìm để hành xử một cách hợp lý”.*

Cutrell và Guan mô tả trong nghiên cứu của mình cách sử dụng kỹ thuật theo dấu mắt để khám phá ra tác dụng trong việc thay đổi cách trình bày trang kết quả tìm kiếm. Một kết quả thú vị đó là nếu nội dung phần mô tả phong phú hơn thì sẽ tăng hiệu quả cho những câu truy vấn dạng tìm thông tin (*informational query*) nhưng lại làm giảm hiệu quả đối với những câu truy vấn dạng điều hướng (*navigational query*). Nguyên nhân có thể do khi độ dài đoạn mô tả tăng, người dùng sẽ chú ý nhiều hơn đến nó thay vì chú ý đến địa chỉ URL trang web..

* 1. **Quảng cáo dựa vào tìm kiếm**

Quảng cáo các sản phẩm thương mại trên nền web ngày càng phát triển. Đây là một lĩnh vực nơi mà các thông tin phản hồi từ việc tìm kiếm giúp ích rất nhiều. Rõ ràng vì mục đích của quảng cáo trực tuyến là thu hút người dùng click vào trang quảng cáo cho nên việc phân tích các đặc điểm của các trang quảng cáo được click nhiều nhất có thể giúp ích cho việc đặt vị trí quảng cáo hợp lý hơn.

Để chính xác hơn, một quảng cáo có thể được mô tả gồm bốn thành phần với là tiêu đề của quảng cáo; là mô tả - chỉ dùng vài từ ngắn gọn súc tích nhất để chỉ ra nội dung được quảng cáo; là URL của trang quảng cáo; là tập các từ khóa liên quan đến nội dung quảng cáo và có khả năng tối ưu hóa lợi nhuận cho nhà kinh doanh nếu như người dùng click vào quảng cáo đó.

Tổng số tiền lớn nhất mà một quảng cáo có thể mang lại cho nhà kinh doanh trên một lần click được gọi là *bid*, và nó thường là một phần của quá trình xếp hạng quảng cáo. Thực tế quá trình tìm *bid* thường được gọi là “*to bid for a keyword*”.

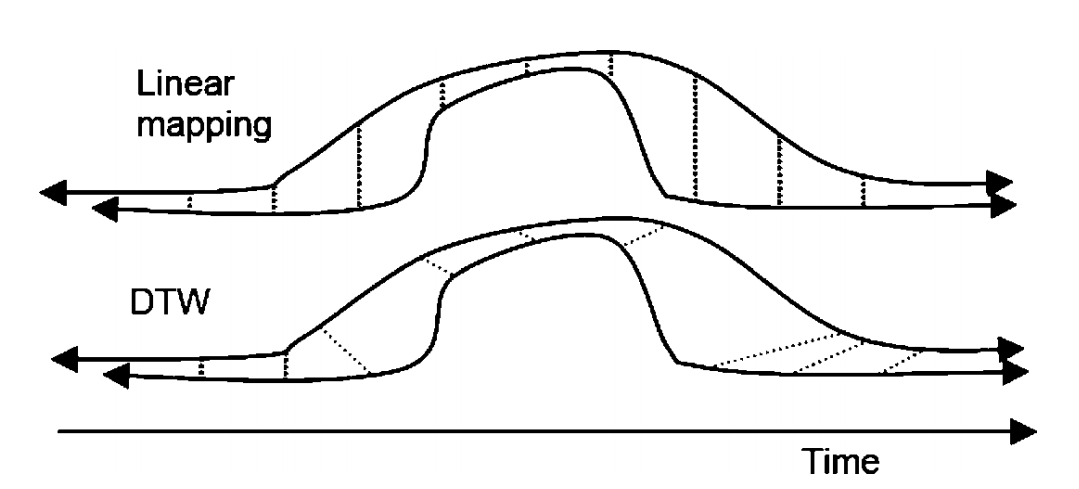
Một nhân tố quan trọng trong việc lựa chọn quảng cáo phù hợp đó là sự liên quan giữa nội dung được quảng cáo và câu truy vấn của người dùng. Trong trường hợp đơn giản, có thể chọn những quảng cáo mà các từ khóa của nó giống với ít nhất một từ khóa của câu truy vấn. Ví dụ nếu người dùng truy vấn *“first aid”* thì những quảng cáo có bid cho *“first”* hoặc *“aid”* hoặc cả hai sẽ được chọn.

* 1. **Phân tích các câu truy vấn theo chuỗi thời gian**

Sau đây chúng ta sẽ trình bày một góc nhìn khác về query log. Các câu truy vấn đều có thể được xem như là tín hiệu trên miền thời gian. Trên mỗi đơn vị thời gian chúng ta sẽ ghi lại số lần xuất hiện của câu truy vấn.

Mục đích của việc sử dụng chuỗi thời gian là để dự đoán (i) tại sao người dùng truy vấn những từ khóa này và (ii) cách mà người dùng phản ứng và tạo ra sự lan truyền của tin tức. Một thuật toán để so sánh tín hiệu từ các nguồn thông tin khác nhau là *Dynamic Time Warping* (DTW). DTW là một cách so sánh hai chuỗi thời gian bằng cách ánh xạ các điểm tương ứng giữa hai chuỗi. Hình 6.2 là một ví dụ của DTW, được so sánh với phương pháp ánh xạ tuyến tính đơn giản. Đường nét đứt nối giữa từ đường cong này đến đường cong kia chỉ ra cách các sự kiện được ánh xạ với nhau.

Tính toán DTW sử dụng một giải thuật quy hoạch động nhằm mục đích tối thiểu hóa khoảng cách euclide giữa hai chuỗi thời gian.



*H6.2 Sự khác nhau giữa giải thuật DTW và ánh xạ tuyến tính khi so sánh hai chuỗi thời gian*

Mục đích của hướng đi này là để khám phá ra mối quan hệ giữa các chuỗi thời gian, từ đó có thể dùng các sự kiện trong nguồn thông tin để dự đoán những sự kiện khác. Điều này giúp cho bộ máy tìm kiếm thích ứng tốt hơn với những yêu cầu của người dùng trong tương lai.

* 1. **Tổng kết**

Chúng ta đã trình bày một vài hướng đi mới trong việc khai phá query log. Những vấn đề được nêu ra đều chưa được giải quyết một cách đầy đủ. Cụ thể, có rất ít bài báo đề cập đến vấn đề nâng cao tính hiệu quả của thuật toán đặt vị trí cho quảng cáo. Chúng tôi tin rằng một trong những hướng đi tốt nhất để giải quyết bài toán này là sử dụng thông tin trong query log.