

# Ba điều cần biết khi sử dụng Google

Hoàng Xuân Phú

Trên Internet hay xuất hiện dạng câu: *"Nếu gõ ... vào Google, chỉ trong ... giây, sẽ có khoảng ... kết quả."* Câu này chứa ba thông tin, nhưng cả ba đều sai hoặc bị hiểu sai. Một cái sai do người sử dụng Google. Hai cái sai kia do Google tạo ra.

## 1. Thiếu dấu ngoặc kép

Ví dụ, hãy dùng Google để tra cứu cụm từ *"Làng có đám cưới, người ta phải dùng trâu để đón cô dâu"* trong bài *"Làng đang vỡ"* của Trần Đăng Khoa.

Có thể, bạn thu được thông tin *"About 215,000 results"* (khoảng 215.000 kết quả). Còn tôi, vào lúc 13 giờ chiều ngày 6 tháng 2 năm 2020, chỉ thu được thông tin *"About 6 results"* (khoảng 6 kết quả).

Vì sao thông tin nhận được lại chênh nhau đến như vậy? Vì bạn đã gõ thiếu hai dấu ngoặc kép. Hai dấu ngoặc kép ở hai đầu cụm từ cần tìm có vai trò cố định lại thứ tự xuất hiện của các chữ trong cụm từ. Thành thử, nếu thiếu hai dấu ngoặc kép đó, Google sẽ tìm kiếm tất cả những tài liệu có các chữ thuộc cụm từ cần tìm, trong đó các chữ ấy không nhất thiết đứng cạnh nhau và không nhất thiết phải xuất hiện đúng với thứ tự như trong cụm từ cần tìm.

Chẳng hạn, bài *"Độc đáo đám cưới rước dâu bằng xe trâu ở Thanh Hóa"* đăng trên Vietnamnet được liệt kê ở đầu danh sách các kết quả tìm kiếm của Google đối với cụm từ *"Làng có đám cưới, người ta phải dùng trâu để đón cô dâu"*. Chỉ vì bài ấy chứa các chữ *"đám", "cưới", "người", "dâu", "có", "dùng", "trâu", "để", "đón", "làng"* và *"phải"* trong trích đoạn sau đây (của Google):

*"Đám cưới 'thần tốc' của cô gái Hải Phòng và người đàn ông ngoại quốc ... tâm đặc biệt là bởi màn rước **dâu có** "1-0-2", **dùng** xe **trâu để** đi **đón** nàng. ... Không xa hoa nhưng độc đáo, có sự chứng kiến chung vui của bà con xóm **làng**. ... GHPGVN yêu cầu khi tổ chức, thực hành các nghi lễ cầu an **phải** đảm ..."*

Để thu được kết quả chính xác hơn và tránh được nhiều nghiệm ngoại lai, thì cần thực hiện hai quy tắc sau:

(1) Khi tìm các văn bản chứa một cụm từ nào đó, phải viết thêm hai dấu ngoặc kép ở hai đầu cụm từ, để cố định thứ tự xuất hiện của các chữ thuộc cụm từ, đồng thời không cho những chữ lạ xen vào giữa cụm từ ấy.

(2) Khi tìm các văn bản chứa mấy cụm từ khác nhau, thì ngoài việc dùng hai dấu ngoặc kép để giới hạn từng cụm từ, còn viết thêm dấu cộng giữa các cụm từ. Khi dùng dấu + để nối các cụm từ, thì trong văn bản các cụm từ ấy có thể xuất hiện cách xa nhau và không tuân theo thứ tự cố định nào cả.

Ví dụ:

- Nếu muốn tìm các văn bản có chứa cụm từ "có đám cưới" (hay "đón cô dâu"), thì gõ "có đám cưới" (hay "đón cô dâu"), với cả dấu ngoặc kép ở đầu và cuối cụm từ, để thu được "khoảng 383.000 kết quả" (hay "khoảng 161.000 kết quả").

- Nếu muốn tìm các văn bản có chứa cả hai cụm từ "có đám cưới" và "đón cô dâu", thì gõ "có đám cưới" + "đón cô dâu", tức là ngoài bốn dấu ngoặc kép (ở hai đầu của hai cụm từ) thì thêm dấu + giữa hai cụm từ, để thu được "khoảng 3.480 kết quả".

Hai quy tắc trên được áp dụng không chỉ cho Google, mà cho cả nhiều công cụ tìm kiếm khác.

## 2. Kết quả hoang đường

Khi hiển thị kết quả, Google cũng chỉ thông báo "About ... results", tức "Khoảng ... kết quả". Chữ "Khoảng" cho thấy đó chỉ là ước lượng. Tuy nhiên, điều mà nhiều người không biết là Google chỉ thu được rất ít kết quả tìm kiếm, cách rất xa, thậm chí cách quá xa so với con số ước lượng được công bố.

Để thấy được điều đó, hãy quan sát kỹ thí nghiệm tìm kiếm hai cụm từ "có đám cưới" và "đón cô dâu", bằng cách gõ "có đám cưới" + "đón cô dâu" (bao gồm cả bốn ngoặc kép " và dấu +) vào ô điền dữ liệu của Google.

Trang kết quả đầu tiên mà ta nhận được bắt đầu với thông tin "About 3.820 results (0,38 seconds)" (tức "Khoảng 3.820 kết quả (0,38 giây)") và kết thúc với dòng sau:

1 2 3 4 5 6 7 8 9 10 Next

Có một số (ban đầu là số 1) màu đen, báo hiệu số trang mà ta đang đọc. Các số khác (ban đầu là từ số 2 đến số 10) màu xanh, khi nhấn chuột vào một trong các số đó thì sẽ được dẫn đến trang có số tương ứng. Và khi nhấn chuột vào Next, thì sẽ được dẫn đến trang tiếp theo (sau trang đang đọc).

Bằng cách đó, ta có thể nhảy từ trang này sang trang khác. Khi đọc đến trang 8, dòng kết thúc của nó là

Previous 3 4 5 6 7 8 9 10 11 12 Next

Theo quy luật, khi đọc sang trang 9, các số trang hiện ra trên dòng kết thúc phải từ 4 đến 13. Thế nhưng, dòng kết thúc lại là

[Previous](#) 2 3 4 5 6 7 8 9 10 11 [Next](#)

Và khi đọc đến trang 10, dòng kết thúc lại là

[Previous](#) 1 2 3 4 5 6 7 8 9 10

Không đi tiếp được nữa, vì 10 đã là số trang lớn nhất và chữ [Next](#) đã biến mất. Thay vào đó, xuất hiện chú thích sau:

*"In order to show you the most relevant results, we have omitted some entries very similar to the 97 already displayed.*

*If you like, you can [repeat the search with the omitted results included](#)."*

Nghĩa là: "Để có thể cung cấp cho bạn những kết quả phù hợp nhất, chúng tôi đã bỏ qua một số mục rất giống 97 mục đã được hiển thị. Nếu muốn, bạn có thể tìm lại để nhận được cả các kết quả đã bị bỏ qua."

Hợp lý đến mức chẳng chê vào đâu được, đúng không? Thế nhưng, khi nhấn chuột vào "[repeat the search with the omitted results included](#)" để tìm lại, ta cũng chỉ đến được trang 33, với dòng cuối cùng như sau:

[Previous](#) 24 25 26 27 28 29 30 31 32 33

Vâng, 33 là số trang lớn nhất và chữ [Next](#) lại biến mất. Thành thử, dẫu muốn có được nhiều kết quả hơn, ta cũng buộc phải dừng lại tại đây.

Từ trang 1 đến trang 32, mỗi trang hiển thị 10 kết quả. Trang 33 hiển thị 4 kết quả. Gộp lại, 324 kết quả là tất cả những gì mà Google có thể cung cấp (tại thời điểm này), chẳng còn gì hơn. Vậy mà, Google lại thông báo là tìm được "*Khoảng 3.820 kết quả*".

Nếu quan niệm chữ "*khoảng*" biểu thị số kết quả làm tròn, thì số kết quả chính xác phải nằm trong khoảng từ 3.815 đến 3.825, chữ không thể là 324. Nếu quan niệm chữ "*khoảng*" chỉ là ước lượng thì cũng không ổn, vì từ "*kết quả*" có nghĩa là số "*kết quả*" thu được sau khi tìm kiếm.

Tiếc rằng, đối với những người thực sự cần có nhiều kết quả tìm kiếm, thì số kết quả mà Google thông báo quá hoang đường.

### 3. Thời gian ảo

Ngay sau thông tin "*About ... results*", tức "*Khoảng ... kết quả*", Google thông báo: "*(... seconds)*", tức "*(... giây)*". Đó là thời gian gì? Ai cũng nghĩ đó là thời gian tìm kiếm, nên rất phục Google, tìm kiếm trên mạng toàn cầu mà nhanh đến thế. Bởi vậy mới viết rằng "*chỉ trong ... giây, sẽ có khoảng ... kết quả*".

Thực ra, đó không phải là thời gian tìm kiếm trên mạng toàn cầu. Để nhận ra điều này, bạn có thể tiến hành thí nghiệm như sau: Hãy vào một trang mạng có tiếng (để đảm bảo rằng nó nằm trong diện quan tâm

của Google), rồi tìm một bài thật mới và tìm trong đó một câu đủ dài và ít có khả năng giống với câu đã đăng ở nơi khác. Sau đó copy câu ấy vào ô ghi dữ liệu tìm kiếm của Google (và đừng quên gõ dấu ngoặc kép " ở đầu và cuối câu, như đã đề cập ở phần 1). Sau khi ra lệnh tìm kiếm, Google sẽ cho câu trả lời: *No results found for "..."*, tức không tìm kết quả nào cho câu ấy. Chứng tỏ, khi bạn yêu cầu tìm, thì Google không hề tìm trên mạng toàn cầu.

Ngày hôm sau, nếu lặp lại thí nghiệm với đúng câu ấy, thì có thể Google sẽ cho bạn một số kết quả. Tại sao? Hoặc là Google chủ động tìm kiếm thông tin theo định kỳ. Hoặc là Google sẽ tiến hành tìm kiếm một số cụm từ nào đó, sau khi đã thấy có người cần tìm kiếm nó. Các kết quả đó được lưu trữ trên các máy chủ của Google. Và khi mọi người tìm kiếm, Google chỉ "*lục trong túi*", tức là chỉ tìm kiếm trong hệ thống máy chủ của Google. Vì vậy mới trình ra kết quả trong thời gian chưa đến một giây.

Rõ ràng, khách hàng của Google chỉ quan tâm đến thời gian tìm kiếm trên mạng toàn cầu, chứ hoàn toàn không quan tâm tới thời gian truy xuất trên máy chủ của Google. Vì vậy, Google cũng không phải thông báo thời gian truy xuất trên máy chủ. Song Google vẫn thông báo thời gian, mà không viết rõ đó là thời gian gì, khiến mọi người hiểu lầm, đó là thời gian tìm kiếm trên mạng toàn cầu. Đã phục, lại càng phục Google. Có lẽ, đó là một kiểu quảng cáo, với hiệu quả rất cao. Nhưng có nên đạt hiệu quả quảng cáo cao bằng cách gây hiểu lầm hay không?

**Tóm lại**, khi viết câu "*Nếu gõ ... vào Google, chỉ trong ... giây, sẽ có khoảng ... kết quả*", hẳn người viết muốn nhấn mạnh cụm từ đang xét xuất hiện trên mạng toàn cầu nhiều đến như vậy. Và với thời gian tính theo giây, người viết muốn nhấn mạnh thêm, chỉ dưới một giây hay mấy giây mà Google đã tìm được nhiều như vậy, thì còn tìm được nhiều hơn hẳn nếu tăng thêm thời gian tìm kiếm. Tiếc rằng, số lượng kết quả được do Google thông báo có thể sẽ khổng lồ, nếu không viết hai dấu ngoặc kép " ở hai đầu cụm từ, để cố định thứ tự các chữ và không để chữ lạ lọt vào cụm từ cần tìm. Kể cả khi đã sử dụng ngoặc kép, thì số lượng kết quả được do Google thông báo vẫn lớn gấp nhiều lần số kết quả mà Google thực sự tìm được. Và thời gian hiển thị không phải là thời gian mà Google đã sử dụng để tìm trên mạng toàn cầu, sau khi nhận được yêu cầu tìm kiếm.

**Ghi chú:** Nếu bạn tiến hành cuộc thí nghiệm với các cụm từ đã nêu trong bài viết, thì thường nhận được các con số hơi khác. Vì kết quả tìm kiếm trên Google rất khác nhau tại các thời điểm khác nhau. Thậm chí, tại cùng một thời điểm thì kết quả tìm kiếm cũng khác nhau, nếu sử dụng các máy tính khác nhau. Nhưng về cơ bản thì có thể chứng kiến hiện tượng đã được trình bày trong bài viết này.

*Hà Nội, ngày 7 tháng 2 năm 2020*