ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH

NHẬP MÔN ĐIỆN TOÁN - CO1005
BÁO CÁO BÀI TẬP LỚN

Đề tài 31: CÔNG CỤ TÌM KIẾM GOOGLE

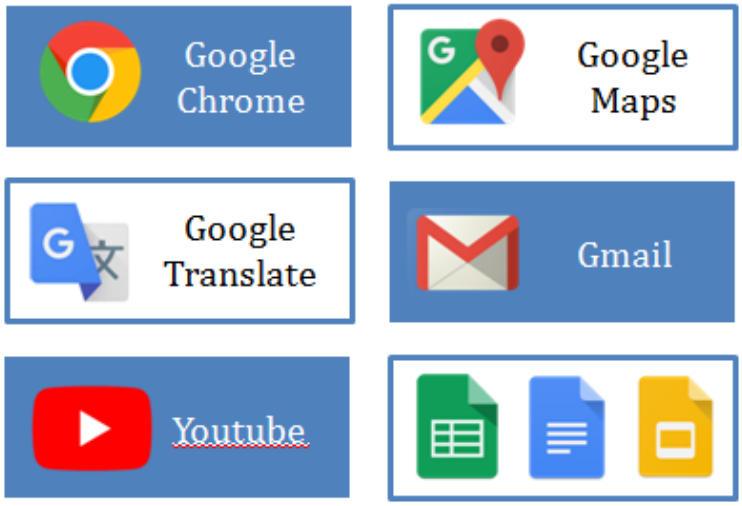
NHÓM THỰC HIỆN: TUT

Họ và tên:	*MSSV:*
Tô Duy Hưng	1810198
Lê Đức Huy	1810166
Nguyễn Bá Tiến	1810578
Châu Thanh Tần	1810501
Nguyễn Phạm Ngọc Quý	1810473

Mục lục

1 Introduction

Google là một công ty Internet tầm cỡ thế giới có trụ sở tại California – Hoa Kỳ, được thành lập vào năm 1998 bởi Larry Page và Sergey Brin. Ngày nay, Google đã cho ra mắt rất nhiều ứng dụng Internet hữu ích như: Chrome, Youtube, Gmail, Google Maps, Google Translate, bộ ứng dụng Google Office - Docs, Sheets, Slides,...



Hình 1: Các ứng dụng của Google

Tuy nhiên, bộ máy tìm kiếm Google Search mới là dịch vụ cung cấp chính và quan trọng nhất của công ty, chính điều này đã làm nên tên tuổi lớn của Google trên thị trường Internet ngày nay.

Dịch vụ này cho phép người truy cập tìm kiếm thông tin trên Internet bằng cách sử dụng công cụ tìm kiếm Google. Đây là công cụ tìm kiếm được sử dụng nhiều nhất trên World Wide Web ở tất cả các nền tảng.

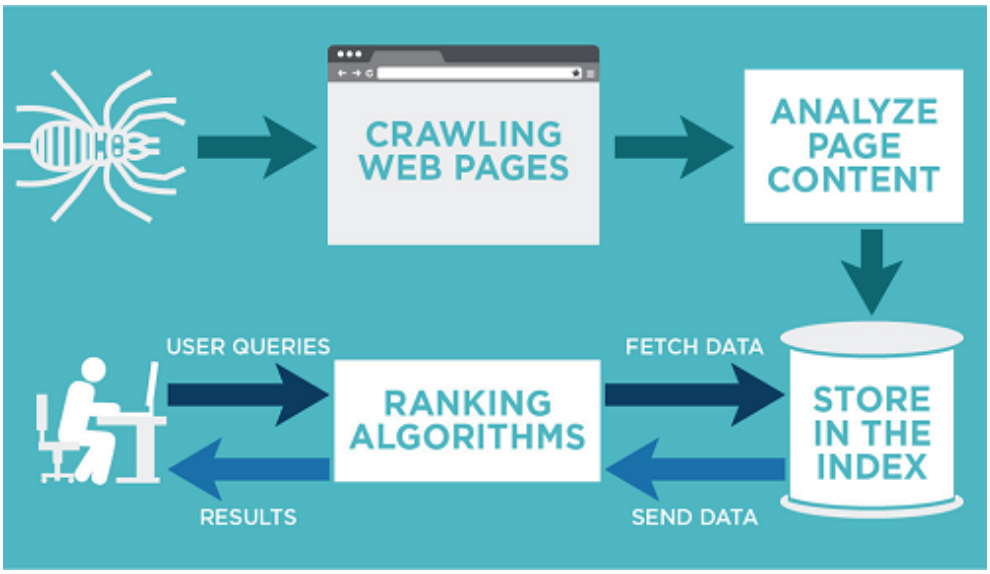
*** Sứ mệnh của Google là:** Sắp xếp thông tin của thế giới, giúp thông tin trở nên hữu ích và có thể truy cập được trên toàn cầu thông qua các định hướng sau:

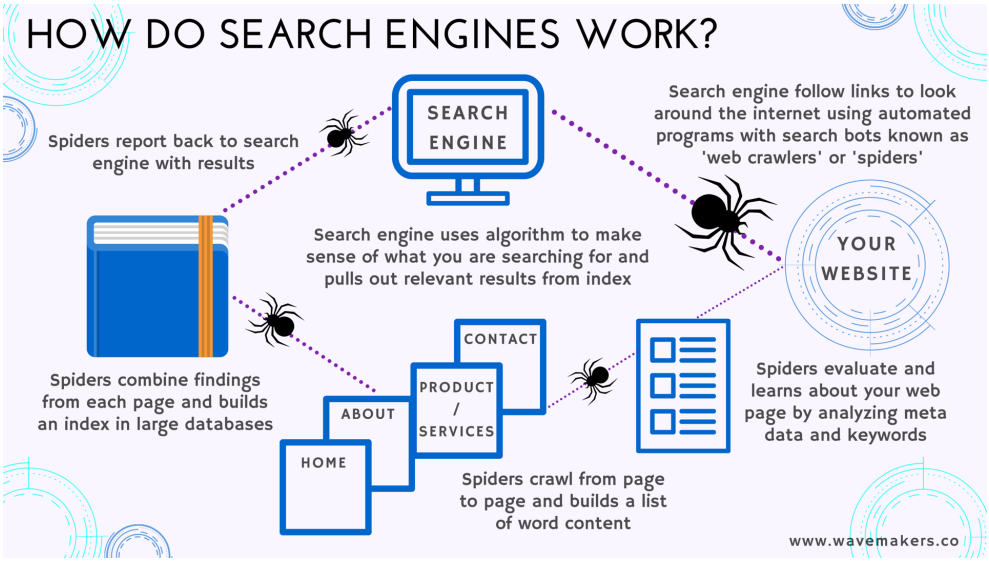
- Tập trung vào người dùng.
- Hỗ trợ chủ sở hữu trang web.
- Cung cấp quyền truy cập tối đa thông tin.
- Liên tục phát triển Google Search để cải thiện kết quả của người dùng.

2 How Search Works

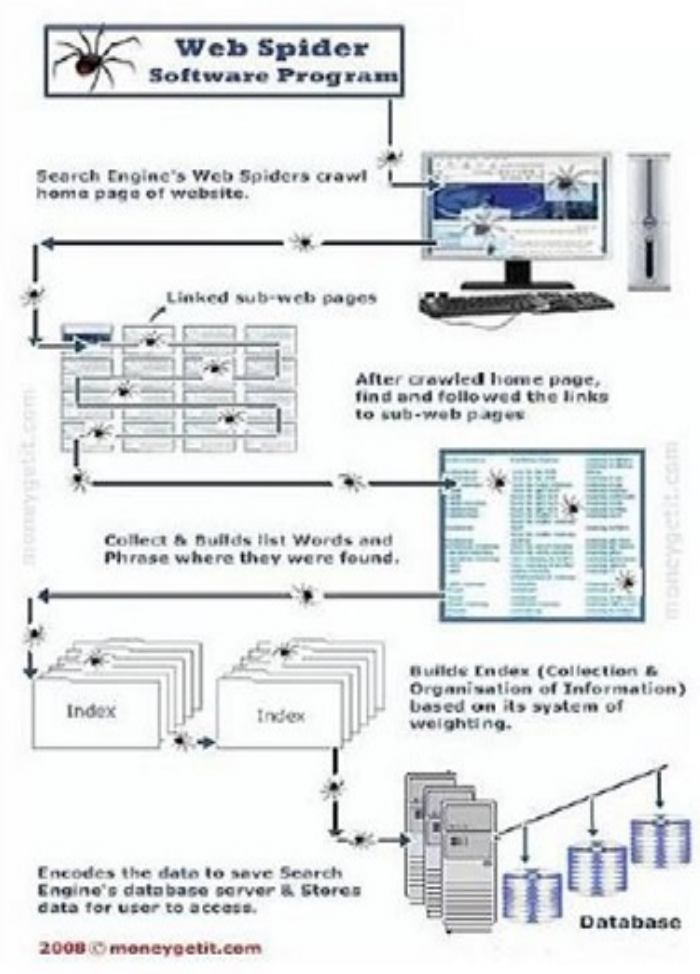
Google Search được thực thi thông qua 3 bước sau:

1. Thu thập dữ liệu và lập chỉ mục - Crawling and Indexing Data.
2. Xử lý truy vấn - Processing Queries.
3. Đưa ra kết quả - Return the results.





2.1 Crawling and Indexing Data - Thu thập dữ liệu và lập chỉ mục



Google thu thập dữ liệu thông qua Spider (hay còn gọi là Crawler) và sắp xếp dữ liệu bằng Indexing (lập chỉ mục). Spider có thể thu thập dữ liệu qua:

- Link trên các site đã index theo chỉ định của Meta Name.
- Add URL form.
- IP Server Reversed, DNS.
- Full Domain Search.

Cơ chế hoạt động của Spider:

- Đầu tiên, Spider sẽ thực hiện nhiệm vụ Crawling, nghĩa là lấy danh sách các máy chủ và trang Web phổ biến. Nó sẽ bắt đầu tìm kiếm với một Site nào đó, đánh dấu các từ chính và cốt yếu trên trang và lần theo các liên kết tìm thấy bên trong Site. Bằng cách này, hệ thống tìm kiếm của Google sẽ nhanh chóng thực hiện công việc trên toàn bộ các phần của trang Web.

- Khi Spider xem xét các trang Web, nó lưu ý các từ bên trong trang Web và nơi tìm thấy từ đó. Các từ xuất hiện trong các thẻ **Title, Meta Description,...** được Spider nhận định là các từ quan trọng có liên quan tới sự tìm kiếm của người dùng.

- Kế đó Google sẽ xây dựng các chỉ mục (index) để lưu thông tin mà Spider thu thập được. Bởi vì các quản trị viên luôn thay đổi, cập nhật thông tin lên các Website nên Spider sẽ luôn thực hiện Crawling và chỉ mục của Google luôn được cập nhật. Các thông tin lưu trong chỉ mục trên CSDL của Google đều được mã hoá. Việc xây dựng chỉ mục cho phép thông tin được tìm thấy một cách nhanh chóng.

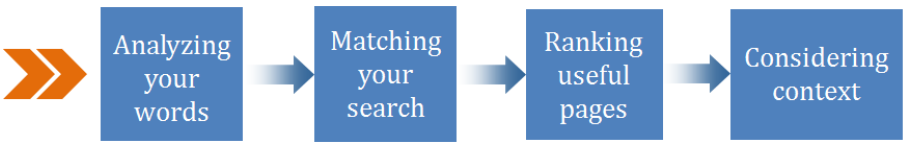
Ví dụ: Website của bạn làm về ngành giáo dục (học online, tài liệu học tập), Google sẽ lưu chỉ mục Website của bạn vào mục "giáo dục"; nếu bạn làm trang đăng tin thể thao hàng ngày, nó sẽ lưu vào phần "thể thao". Như vậy việc tìm kiếm sẽ không bị chồng chéo lên nhau và quy trình sẽ nhanh hơn.

2.2 Processing queries - Xử lý truy vấn

Sau khi đã tạo được chỉ mục, vấn đề nảy sinh là người dùng thực sự muốn câu trả lời, chứ không phải là hàng tỷ trang web, và vì thế, hệ thống xếp hạng của Google phải tìm qua hàng trăm tỷ trang web trong chỉ mục Tìm kiếm để cung cấp cho người dùng kết quả hữu ích và phù hợp chỉ trong một phần nhỏ của giây.

Các hệ thống xếp hạng này được tạo thành từ một loạt các *thuật toán* phân tích nội dung bạn đang tìm kiếm và thông tin gì cần trả về cho bạn. Trong quá trình phát triển Tìm kiếm, để làm cho tính năng này hữu ích hơn, Google đã tinh chỉnh các thuật toán của mình để đánh giá tìm kiếm và kết quả của bạn một cách chi tiết hơn nhằm làm cho dịch vụ hoạt động tốt hơn cho bạn.Đây chính là giai đoạn xử lý truy vấn.

Có tất cả 4 bước được thực hiện trong giai đoạn xử lý truy vấn này bao gồm:



Hình 2: 4 steps in Processing Queries

2.2.1 Analyzing your words

Việc hiểu được nghĩa của tìm kiếm của bạn là điều quan trọng để trả về câu trả lời thích hợp. Vì thế để tìm các trang có thông tin liên quan, bước đầu tiên của Google là phân tích các từ trong truy vấn của bạn có nghĩa gì. Google xây dựng các mô hình ngôn ngữ để cố giải mã những chuỗi từ nên tra trong chỉ mục. Hệ thống này mất 5 năm để phát triển và đã cải thiện đáng kể kết quả trong hơn 30% lượt tìm kiếm bằng các ngôn ngữ khác nhau.

Quy trình này gồm 3 bước:

- Phân tích lỗi chính tả
- Phân tích hệ thống từ đồng nghĩa. Ví dụ như: thay = thay thế, đổi = trao đổi, ...

Hình 3: Hình ảnh minh họa cho phân tích hệ thống từ đồng nghĩa

- Phân tích xem người dùng đang tìm kiếm loại thông tin nào: cụ thể hay truy vấn rộng, có phải là từ khoá thịnh hành để đưa ra kết quả xuất bản mới nhất.

2.2.2 Matching your search

Sau khi phân tích từ ngữ truy vấn, Google sẽ thực hiện các công việc sau:

- Finding Web: Google tìm các trang web có thông tin khớp với truy vấn của người dùng.
- Finding in Index: Khi bạn tìm kiếm, ở mức độ cơ bản nhất, các *thuật toán* sẽ tra cụm từ tìm kiếm trong chỉ mục để tìm các trang thích hợp dựa trên tần suất, vị trí các từ khoá đó xuất hiện trên một trang, liệu chúng xuất hiện trong tựa đề, tiêu đề hay nội dung.
- Matching: Hệ thống sẽ có các thuật toán đo mức độ phù hợp giữa kết quả tìm kiếm tiềm năng và nội dung đang tìm kiếm.
  Ví dụ khi bạn tìm từ "chó", có thể bạn không muốn một trang có từ "chó" xuất hiện hàng trăm lần. Google cố tìm hiểu xem liệu trang có cung cấp câu trả lời cho truy vấn của bạn không và không chỉ lặp lại truy vấn hay không. Vì thế các thuật toán Tìm kiếm phân tích liệu trang có nội dung thích hợp hay không — chẳng hạn như ảnh, video về chó hay thậm chí danh sách các giống chó.
- Prioritizing: Ưu tiên các trang có cùng ngôn ngữ với truy vấn hoặc ngôn ngữ tùy chọn của bạn.

2.2.3 Ranking useful pages

Đối với một truy vấn thông thường, có hàng nghìn, thậm chí là hàng triệu trang web có thể cung cấp thông tin liên quan. Vì thế, để giúp xếp hạng các trang tốt nhất đầu tiên, Google đã viết các thuật toán để

đánh giá mức độ hữu ích của các trang web này, đặc biệt là thuật toán **Pagerank**.

Các thuật toán này phân tích hàng trăm yếu tố khác nhau để cố hiển thị thông tin tốt nhất có sẵn trên web, từ số lượt visits, độ mới mẻ của nội dung cho đến số lần xuất hiện của cụm từ tìm kiếm của bạn và liệu trang có cung cấp trải nghiệm người dùng tốt hay không. Để đánh giá độ đáng tin cậy và nguồn có căn cứ về chủ đề, Google tìm các trang web có vẻ được nhiều người dùng đánh giá cao đối với cùng truy vấn. Nếu trang web có nhiều backlinks [1] từ các trang web nổi bật khác về chủ đề này, đó là một dấu hiệu tốt cho thấy thông tin có chất lượng cao.

Đối với các trang cố vươn lên đầu các kết quả tìm kiếm bằng cách lặp lại từ khoá, mua liên kết, vi phạm nguyên tắc quản trị web [2], Google có các thuật toán để xác định spam và xóa các trang web vi phạm khỏi kết quả.
****Trang nào đáp ứng nhiều tiêu chí hơn sẽ có thứ hạng cao hơn.****

Hình 4: Một số nội dung cơ bản của Webmaster Guildlines

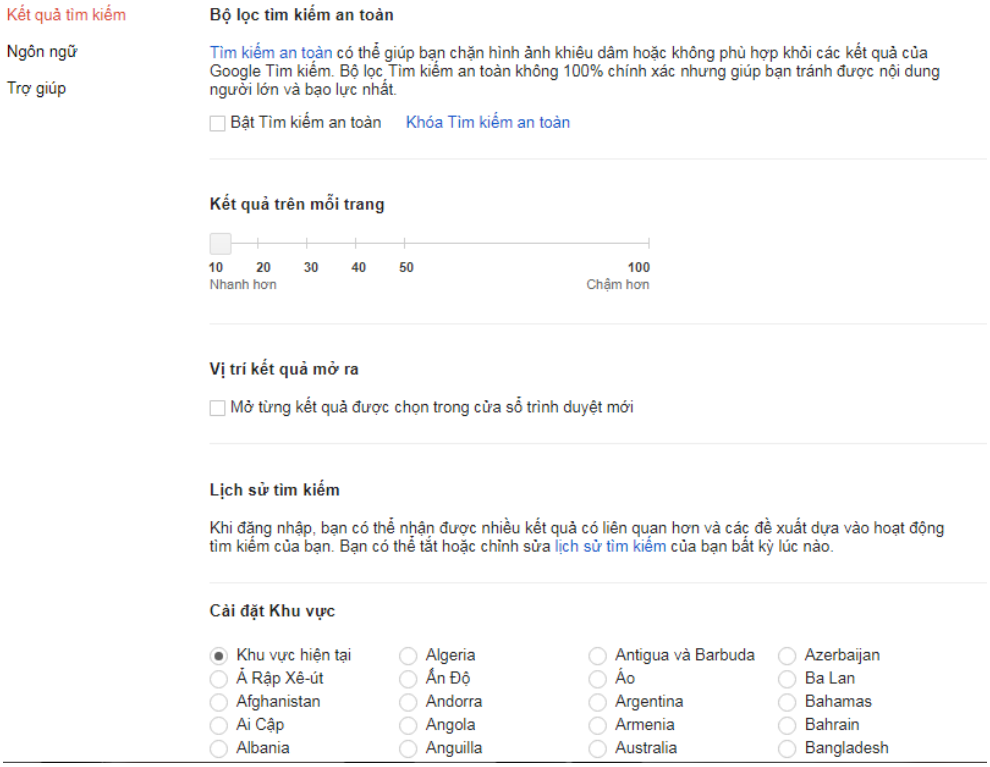
[1] backlinks là những liên kết hướng tới website hoặc webpage từ những web khác nhau
[2] Có thể tham khảo thêm về Webmaster Guildlines tại: https://support.google.com/webmasters/answer/35769?hl=en

2.2.4 Considering context

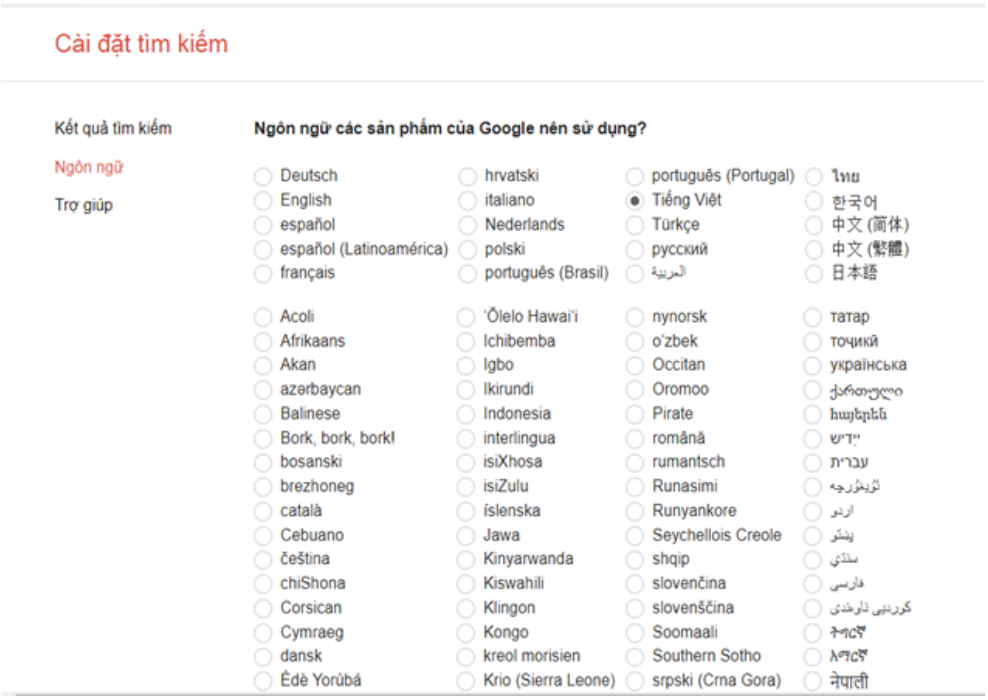
Bên cạnh đó, các thông tin chẳng hạn như vị trí của bạn, lịch sử tìm kiếm và cài đặt Tìm kiếm đều giúp Google tùy chỉnh kết quả cho phù hợp và hữu ích với bạn nhất trong khoảnh khắc đó.

Google sử dụng quốc gia và vị trí của người dùng để cung cấp nội dung thích hợp với khu vực. Ví dụ: nếu bạn ở Chicago và tìm "football", Google có thể sẽ hiển thị cho bạn kết quả về môn bóng bầu dục và câu lạc bộ Chicago Bears trước tiên. Ngược lại, nếu bạn tìm "football" ở Luân Đôn, Google sẽ xếp hạng các kết quả về bóng đá và giải Premier League cao hơn.

Ngoài ra, Cài đặt Tìm kiếm cũng là một chỉ báo quan trọng về việc bạn có khả năng thấy kết quả nào hữu ích, chẳng hạn như liệu bạn có thiết lập một ngôn ngữ ưu tiên hay chọn tham gia Tìm kiếm an toàn (một công cụ giúp lọc các kết quả không phù hợp) hay không.



Hình 5: Bảng cài đặt tìm kiếm theo khu vực, lịch sử tìm kiếm



Hình 6: Bảng cài đặt tìm kiếm theo ngôn ngữ

2.3 Return the best results - Trả kết quả cho người dùng

Dựa vào các tiêu chí sắp xếp ở trên, Google sẽ hiện thị kết quả tìm kiếm phù hợp nhất cho người dùng trong thời gian rất nhanh.

Không chỉ vậy, trước khi cung cấp cho bạn kết quả, Google đánh giá xem tất cả các thông tin liên quan đến nhau như thế nào: liệu chỉ có một chủ đề duy nhất trong kết quả tìm kiếm hay có nhiều chủ đề? Có phải có quá nhiều trang tập trung vào một cách diễn giải hời hợt, từ đó đảm bảo cung cấp thông tin đa dạng bằng nhiều định dạng chi tiết và hữu ích nhất cho loại tìm kiếm của bạn. Và khi web phát triển, Google cũng từng ngày phát triển hệ thống xếp hạng của mình để cung cấp kết quả tốt hơn cho nhiều truy vấn hơn.

3 Search Algorithms - Các thuật toán tìm kiếm

3.1 PageRank

3.1.1 Giới thiệu chung:

PageRank là một thuật toán phân tích liên kết (link) được phát triển bởi Larry Page và sau đó được Sergey Brin tiếp tục nghiên cứu với giả thuyết: "Sự lớn mạnh của một trang web có thể được đánh giá bởi

số hyperlink [3] được trỏ đến trang web". Dự án này là tiền đề cho sự ra đời của Google vào năm 1988. Giải thuật PageRank được Google sử dụng để làm thước đo đánh giá mức độ phổ biến của trang và xếp hạng các trang web đó trong kết quả công cụ tìm kiếm của họ. PageRank được hiển thị trên GoogleToolbar là một số nguyên từ 0 đến 10, thường thì trang có PageRank càng cao thì vị trí của nó trên trang tìm kiếm càng được ưu tiên.

Trên thực tế, Google sẽ xem một liên kết từ trang A đến trang B như một bình chọn. Tuy nhiên, Google xem xét nhiều hơn là dựa trên khối lượng bình chọn hoặc số liên kết mà trang nhận được, nó còn phân tích về trang tác động đến bình chọn, bình chọn được thực hiện bởi các trang càng "quan trọng" thì càng làm cho các trang khác trở nên "quan trọng".

3.1.2 Mô tả thuật toán:

PageRank là phân bố xác suất, thể hiện khả năng khi một người click chuột ngẫu nhiên vào đường link và sẽ tới được trang web cụ thể. Ví dụ giá trị pagerank = 0.5 có nghĩa là 50% cơ hội một người nào đó click vào một link ngẫu nhiên để được chuyển đến văn bản đó.

PageRank có thể được tính cho các tập văn bản với tài liệu có độ dài bất kỳ. Khi bắt đầu tính toán thì sự phân bố đó được chia đều cho tất cả những văn bản trong tập văn bản.

Các tính toán PageRank cần một số lần "lặp đi lặp lại" qua các văn bản trong tập để có thể đạt được giá trị thực tế một cách thiết thực hơn.

Công thức chung, giá trị PageRank đối với một trang bất kỳ **u** có thể tính như sau:

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

v là các trang có chứa trong trong tập hợp B_u - tập chứa các trang có link đến trang u

Tuy nhiên,lý thuyết PageRank cho rằng, ngay cả một người dùng giả thiết click ngẫu nhiên vào các trang web thì cuối cùng cũng sẽ phải dừng lại. vì vậy thuật toán PageRank phải tính đến yếu tố damping [4] sử dụng mô hình khi người dùng bất kì sẽ cảm thấy chán sau một vài lần click và chuyển đến vài trang web khác một cách ngẫu nhiên. Như vậy:

$$PR(A) = \frac{1-d}{N} + d\left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \cdots\right)$$

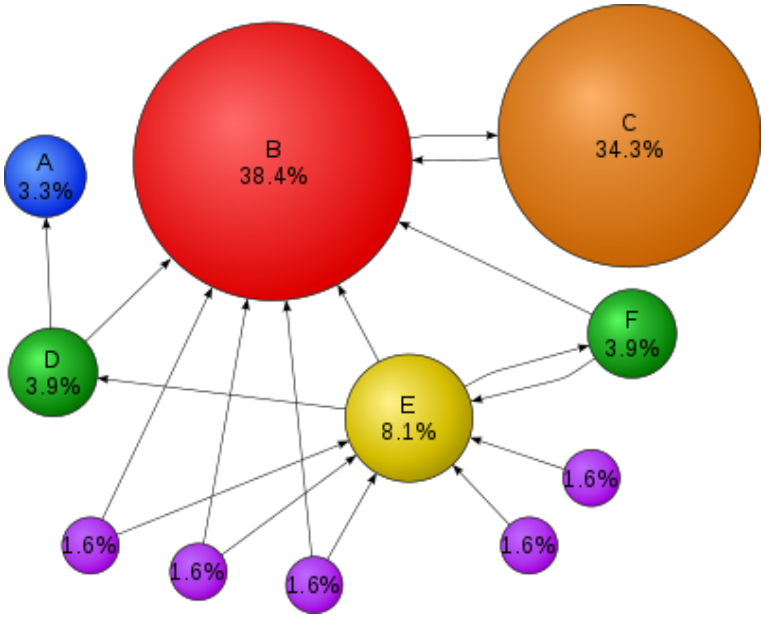
Nếu như trang web không có đường link đến các trang khác, nó sẽ thành ngõ cụt và việc truy cập ngẫu nhiên sẽ dừng lại. Nhưng nếu người dùng đến trang không có các link khác, thì người dùng sẽ chọn ngẫu nhiên một trang khác để tiếp tục truy cập. Khi tính Pagerank, những trang không có link trỏ đi các trang khác sẽ được giả định có link trỏ đến tất cả các trang trong tập văn bản. Và như vậy giá trị Pagerank sẽ được chia đều cho các trang khác. Nói một cách khác, để công bằng với những trang web có outbound link, thì các truy cập ngẫu nhiên sẽ được thêm vào tất cả những trang trong Web, với xác suất d=0.85, được

[3] siêu liên kết
[4] xác suất người dùng tiếp tục click trong bất cứ bước nào, ước lượng bằng 0.85

ước tính từ tần số trung bình mà người dùng sử dụng khi đánh dấu một tính năng bằng trình duyệt.

$$PR(A) = 1 - d + d\left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \cdots\right)$$

Xét ví dụ sau:



Hình 7: Thuật toán PageRank đối với hệ thống liên kết đơn giản

Qua thuật toán này, ta có những nhận xét rằng:
- Những trang không có backlinks thì cùng giá trị PR và $PR = \frac{1-d}{N} \approx 1.6\%$
- Các trang có cùng (hoặc đối xứng) backlinks thì có cùng giá trị PR, như D và F
- 1 backlink từ 1 trang PR cao sẽ tốt hơn rất nhiều so với nhiều backlinks từ các trang PR thấp.

Gọi G là một đồ thị các trang Web. Đặt G(V,E) với V=(1,2,...,n) là tập n đỉnh của đồ thị (mỗi đỉnh là một trang Web cần tính hạng trang) còn E là tập các cạnh, $E = \{(i,j)$ nếu có siêu liên kết từ trang i trang j $\}$. Chúng ta giả thiết rằng đồ thị trang Web là liên thông, nghĩa là từ một trang bất kì có thể có đường liên kết tới một trang Web khác trong đồ thị đó. Với mỗi trang Web i, kí hiệu N(i) là số liên kết đi ra từ trang Web thứ i và B(i) là số các trang Web có liên kết đến trang i.

Khi đó hạng trang r(i) của trang Web i được định nghĩa như sau:

$$r(i) = \sum_{j \in B(i)} \binom{r(j)}{n(j)}$$

Việc ta chia cho N(i) cho thấy rằng những trang có liên kết tới trang i sẽ phân phối hạng của chúng cho các trang Web mà chúng liên kết tới.

Các phương trình này được viết lại dưới dạng ma trận r=rP trong đó:

$r = [r1, r2, ..., rn]$ là vector PageRank, với ri là hạng của trang Web i trong đồ thị trang Web.

P là ma trận chuyển n x n với giá trị các phần tử được xác định:

$$\begin{cases} 1 & \text{nếu liên kết từ i đến j} \\ 0 & \text{ngược lại} \end{cases} \tag{1}$$

Từ đó công thức PageRank được viết lại: r=rP

3.1.3 Nhận xét và kết luận

- Ưu điểm:
  - Là một phương pháp tính hạng khá tốt và có quá trình tính toán độc lập với người dùng nên có thể thực hiện độc lập và không ảnh hưởng đến tốc độ tìm kiếm.
  - Đã từng đem lại rất nhiều kết quả khả quan.

- Nhược điểm:
  - Thuật toán chỉ quan tâm đến các liên kết giữa các trang Web mà không quan tâm đến nội dung trang Web nên có thể dễ bị đánh lừa bởi công nghệ spam.

- **Tuy nhiên,** vào năm 2016, Google đã gỡ bỏ PageRankToolbar và xóa điểm số PageRank công khai.

  Lý do chính là bởi nhược điểm đã trình bày bên trên của PageRank: liên kết spam.

  Thật công bằng khi nói rằng SEO từ lâu đã bị ám ảnh bởi PageRank như một yếu tố xếp hạng, có lẽ bởi vì cái gọi là "thanh công cụ PageRank" cung cấp một thước đo có thể nhìn thấy, theo đúng nghĩa đen, về mức độ xứng đáng của một trang web.

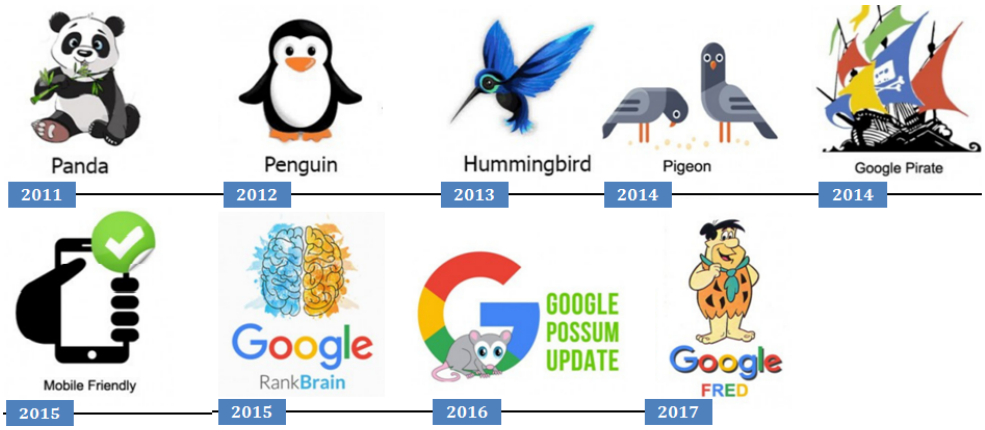
  Tại thời điểm đó không có thước đo trực quan nào tồn tại đối với bất kỳ yếu tố xếp hạng nào khác, điều đó khiến cho PageRank là yếu tố duy nhất quan trọng. Kết quả là, mọi người sớm bắt đầu mua và bán các liên kết "PR cao". Nó đã trở thành một ngành công nghiệp khổng lồ.

  Có rất nhiều cách để người bán liên kết xây dựng các liên kết "PR cao". Vào giữa những năm 2000, một trong những chiến thuật mua lại chính là để lại nhận xét trên blog.

  Đối với Google, đây là một vấn đề lớn. Liên kết ban đầu là một thẩm phán tốt về chất lượng bởi vì chúng được đưa ra tự nhiên cho các trang xứng đáng. Các liên kết không tự nhiên làm cho thuật toán của họ kém hiệu quả hơn trong việc làm sáng tỏ các trang chất lượng cao từ những trang có chất lượng thấp.

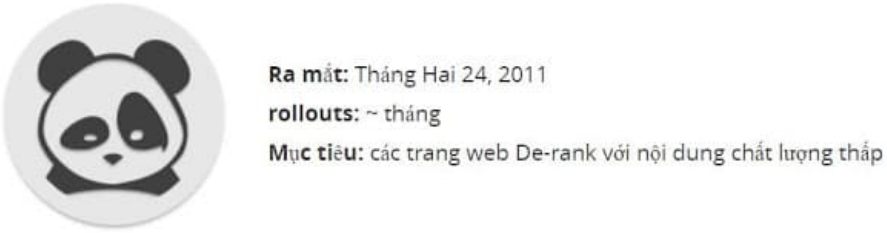
  Vậy nên Google đã gỡ bỏ PageRankToolbar và xóa điểm số PageRank công khai và nói rằng: *"Khi Internet và sự hiểu biết của chúng ta về Internet đã phát triển phức tạp, điểm số PageRank của thanh công cụ đã trở nên ít hữu ích hơn đối với người dùng dưới dạng chỉ số đơn lẻ. Gỡ bỏ hiển thị PageRank khỏi thanh công cụ giúp tránh nhầm lẫn giữa người dùng và quản trị viên web về tầm quan trọng của chỉ số"*

- **Nhưng,** chắc chắn một điều rằng Google PagRank **vẫn chưa chết** và vẫn đóng một vai trò cực kì quan trọng đối với các thuật toán tìm kiếm khác của Google.



Hình 8: Một số thuật toán tìm kiếm khác của Google

3.2 Thuật toán Panda (gấu trúc):



3.2.1 Giới thiệu

Google Panda là thuật toán chuyên xử lý các website có chất lượng nội dung thấp, nghèo nàn, spam, copy,.... Một khi bị Panda áp đặt hình phạt, website của bạn sẽ không thể có được thứ hạng cao trên công cụ tìm kiếm.

Ban đầu, Panda có vai trò như một bộ lọc, chuyên lọc các website có nội dung kém chất lượng ra khỏi cuộc chơi đua hạng trên Google Search. Tuy nhiên, đến đầu năm 2016, thuật toán Panda đã chính thức được Google đưa vào danh sách là một phần trong những thuật toán cốt lõi xếp hạng tìm kiếm của họ. Nghĩa là đã được hơn 1 năm rồi, Panda luôn trong trạng thái Real – time (luôn luôn hoạt động, chứ không hoạt động cục bộ, theo chu kỳ như xưa nữa) do đó website của bạn có thể bị phạt cũng có thể được gỡ phạt nhanh hơn trước rất nhiều.

3.2.2 Mục tiêu:

Thuật toán Panda thường quét vào những lỗi sau:

- Nội dung trùng lặp quá nhiều.

- Đi copy nội dung từ website khác.
- Nội dung quá thô sơ, nghèo nàn.
- Spam do người dùng.
- Nhồi nhét từ khóa.
- Nội dung website không thu hút được người dùng.

3.2.3 Làm sao để tránh thuật toán Panda:

- **Bước 1: Kiểm tra nội dung trùng lặp trên website của bạn:**

  Trùng lặp nội dung trên website là lỗi phổ biến nhất đưa bạn đến con đường làm miếng mồi ngon cho Panda. Google khuyến cáo người quản trị website rằng hãy kiểm tra thường xuyên hơn sự trùng lặp nội dung trên trang web. Đặc biệt lỗi trùng lặp tiêu đề và mô tả bị Google phạt rất nặng.

  Có rất nhiều công cụ hỗ trợ bạn kiểm tra trùng lặp nội dung trên website như SEO PowerSuite 's website Auditor.

  Nếu các công cụ không thể xem được nội dung trùng lặp trên website của bạn thì hãy kiểm tra lại file Robots.txt hoặc xem lại bạn đã cho Google index nội dung hay chưa.

- **Bước 2: Kiểm tra đạo văn, nội dung copy:**

  Nội dung website trùng lặp với một website khác đi trước bạn sẽ là con đường kích hoạt hình phạt từ Panda rất nhanh. Nếu bạn nghi ngờ rằng có nội dung trên trang của bạn hãy sử dụng ngay công cụ CopyScape. CopyScape cung cấp dịch vụ hoàn toàn miễn phí và không mất công đăng ký rườm rà như một số công cụ khác nhưng những gì nó đem lại cho chúng ta lại rất hữu dụng. Tốc độ cập nhật nội dung của công cụ này cũng cực kỳ nhanh chóng, đặc biệt nó đưa ra ngay website có nội dung tương tự được Google index trước, điều này mang lại kiểm chứng chính xác nhất.

  Rất nhiều chủ đề, ngành công nghiệp có thể không bao giờ đảm bảo được rằng website có 100

- **Bước 3: Xác định nội dung kém chất lượng**

  Nội dung kém chất lượng là một khái niệm khá mơ hồ, các bạn sẽ đặt ngay câu hỏi thế nào là kém chất lượng? Các website cung cấp nội dung với số lượng từ ngữ văn bản quá ít, đặt nhiều quảng cáo, có nhiều liên kết ra ngoài đến website khác, nội dung mang lại rất ít giá trị cho người dùng,... được cho là một trang kém chất lượng.

  Nếu nội dung bài viết dưới 250 từ (Nếu các bạn dùng mã nguồn WordPress có công cụ hỗ trợ SEO Yoast sẽ thấy nó khuyến khích bạn viết nội dung tối thiểu là 250 từ) và trang có hơn 100 link out (liên kết đến website ngoài) thì một khả năng rất cao bạn đã bị đánh giá là website có nội dung kém chất lượng và khả năng cao bạn đã được Google Panda "chăm sóc".

  Để kiểm tra chất lượng nội dung website các bạn có thể sử dụng công cụ hỗ trợ SEO WebSite Auditor.

- **Bước 4:**

  Nhồi nhét từ khóa là thuật ngữ để chỉ những bạn đang ở tình trạng SEO quá đà, quá tối ưu từ khóa cho một page.

  Hãy đảm bảo rằng nội dung được xây dựng một cách tự nhiên nhất, đừng cố gắng spam các từ khóa, Nếu các bạn đủ sức hãy sử dụng một từ khóa rộng bao hàm nhiều từ khóa. Ví dụ, các bạn đang làm một diễn đàn SEO, mà các bạn muốn nó lên top cả các từ khóa như diễn đàn seo web, diễn đàn seo uy tín ở Việt Nam chẳng hạn, thay vì việc nhồi nhét tất cả các từ khóa này lên website, bạn hãy tối ưu website với một từ khóa duy nhất là Diễn đàn SEO web uy tín ở Việt Nam và bắt đầu quá trình

Off page, điều hướng link nội bộ với tất cả các từ khóa bạn đang muốn lên top. Bạn sẽ được đánh giá cao hơn nhiều và tránh được Google Panda nhòm ngó.

Để kiểm tra xem website có nhồi nhét từ khóa hay không bạn hãy sử dụng công cụ SEO PowerSuite 's WebSite Auditor.

- **Bước 5: Khắc phục các lỗi có thể đưa bạn thành miếng mồi ngon cho Panda:**

  Một khi phát hiện ra các lỗi mà Panda nhắm vào bạn hãy khẩn trương sửa chữa ngay chúng càng sớm càng tốt bởi có thể bạn chưa bị phạt bởi phiên bản cũ nhưng sau khi thuật toán cập nhật phiên bản mới mạnh hơn, tinh vi hơn bạn có thể thành miếng mồi ngon nhanh chóng của Gấu Trúc!

3.3 Thuật toán Penguin (chim cánh cụt):

3.3.1 Giới thiệu:

Mục tiêu của thuật toán Penguin (Chim cánh cụt) chính là các website có hồ sơ liên kết kém chất lượng, liên kết spam và không tự nhiên. Chim cánh cụt là nỗi ám ảnh lớn nhất đối với không chỉ SEO mũ đen mà ngay cả SEO mũ trắng nếu không cẩn thận đi đúng con đường cũng sẽ dễ dàng trở thành con mồi của nó.

Từ cuối năm 2016, Penguin được xếp vào là một trong những thuật toán cốt lõi của Google và luôn hoạt động Real time nghĩa là luôn luôn hoạt động trong thời gian thực. Điều đó có nghĩa là bạn có thể bị phạt nhanh hơn và cũng có thể được giải thoát bởi hình phạt này nhanh hơn. Đối với những webmaster mà có tên miền bị phạt bởi Penguin thì đây có lẽ là tin mừng họ mong chờ nhất từ Google, Bởi trước đây, một khi bị phạt bởi Penguin, bạn chỉ còn cách chờ đợi đợt cập nhật nâng cấp tiếp theo của thuật toán thì mới có cơ hội được xem xét tha hay không. Có những người phải chờ đợi mấy năm trời, như vậy chẳng thà bạn đi mua luôn một tên miền mới còn hơn ngồi chờ đợi ngày ra tù với thời gian khủng khiếp ấy.

3.3.2 Mục tiêu:

- Hồ sơ liên kết của bạn là các website kém chất lượng, các website spam.
- Liên kết đến từ các website có mục đích chỉ để xây dựng liên kết. Private Blog Networks (PBNs) là xu hướng xây dựng liên kết trong năm 2017. Đây là hình thức xây dựng rất tốn kém nhưng lại là xu hướng bởi những lợi ích to lớn mà nó mang lại nếu bạn làm đúng cách, và chỉ cần chờ bạn sai lầm, Penguin sẽ có một miếng mồi ngon ngay lập tức, hậu quả có thể là cả hệ thống của bạn ra đi.
- Backlink từ các website không liên quan gì đến chủ đề của bạn.
- Mua bán liên kết. Google luôn nỗ lực để giảm thiểu việc sử dụng text link bởi các SEOer. Nhân đây, tôi cũng muốn khẳng định một điều với một số bạn Newbie rằng text link không bao giờ mất giá trị của nó. Bởi backlink luôn là một phần của xếp hạng tìm kiếm. Textlink nó giống như kiểu bạn sản

xuất công nghiệp vậy, được toàn trang liên kết đến website đích, thay vì một vài liên kết trong bài viết. Có bạn nhắn tin hỏi tôi rằng có phải text link đã mất giá trị không, sao mình cứ đặt là tụt hạng vậy. Đó là bởi vì website của bạn chưa đủ sức để tiếp nhận text link đó. Bạn cứ tưởng tượng sơn hào hải vị đúng là tốt cho sức khỏe thật, cho bạn ăn một con cua gạch thì ăn được và tốt cho bạn chứ bắt bạn ăn 10kg một lúc bạn sẽ bội thực và chết ngay!

- Bạn nhồi nhét quá nhiều liên kết vào một từ khóa tập trung.

3.3.3 Làm sao để tránh thuật toán Penguin:

- **Bước 1: Giữ cho sự tăng trưởng liên kết một cách tự nhiên nhất:**

  Google thông báo rằng họ chỉ cho các website có thứ hạng cao khi họ cho rằng hồ sơ liên kết của bạn là tự nhiên. Google cũng cho biết thêm rằng liên kết tự nhiên là liên kết do người dùng đặt và chia sẻ lên các website khác.

  Nói như vậy có nghĩa là Google không thể phân biệt được đâu là do người dùng đặt backlink đâu là do bạn xây dựng.

  Vậy vì sao họ lại phạt website của bạn? Họ dựa vào đâu để áp đặt hình phạt lên bạn? Một yếu tố đầu tiên dẫn đến bạn trở thành con mồi ngon của chim cánh cụt đó chính là tốc độ tăng trưởng liên kết của bạn.

  Nếu website tự dưng có một lượng liên kết lớn đổ về, tăng vọt so với sự phát triển trước đây thì chắc chắn là bạn sẽ bị Google phạt rồi.

  Có rất nhiều công cụ hỗ trợ việc thống kê liên kết cho bạn như Ahrefs, Moz Pro, SEO SpyGlass,... Dưới đây là hình ảnh thống kê sự tăng trưởng liên kết của một website bởi công cụ phân tích backlink Ahrefs.

- **Bước 2: Kiểm tra khả năng bạn bị Google Penguin bị phạt**

  SEO SpyGlass có thể đưa ra được các dự đoán giúp bạn đánh giá các yếu tố như tình hình hồ sơ liên kết thế nào, tốc độ tăng trưởng có ổn không, chất lượng liên kết thế nào,...

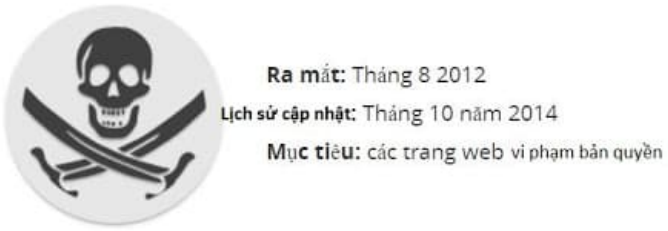
  Ở phiên bản miễn phí, SEO SpyGlass sẽ giúp bạn phân tích 1000 liên kết, để có thể phân tích nhiều hơn nữa bạn cần sử dụng phiên bản pro.

- **Bước 3: Loại bỏ các liên kết xấu**

  Cách tốt nhất đó là bạn hãy liên hệ với người quản trị website có backlink về trang web của bạn và yêu cầu họ hỗ trợ bằng cách gỡ các liên kết đó ra. Tuy nhiên, điều này cũng không hề dễ dàng gì nếu bạn chẳng quen biết gì với người quản trị kia, chưa kể trường hợp xấu đó là đối thủ chơi đểu bơm backlink bẩn cho bạn thì việc yêu cầu trợ giúp gỡ liên kết rõ ràng là vô nghĩa!

  Rất may, Google cũng hỗ trợ cho chúng ta vấn đề này triệt để. Họ đã cung cấp công cụ chặn liên kết xấu Google Discavow có trong Webmaster Tool mà Google phát hành miễn phí. Để chặn liên kết xấu các bạn cần lọc ra các tên miền và URL muốn chặn. Bằng cách này, Google sẽ loại bỏ các liên kết xấu ra khỏi hồ sơ liên kết cho bạn.

3.4 Thuật toán Pirate:



3.4.1 Giới thiệu:

Thuật toán Pirate xuất hiện nhằm trừng phạt các website vi phạm bản quyền nội dung, bị nhiều báo cáo vi phạm gửi đến Google.

Để yêu cầu Google xử lý bản quyền các bạn truy cập vào liên kết sau: https://support.google.com/legal/troubleshooter/ Trang này sẽ giúp bạn truy cập đúng nơi để báo cáo nội dung bạn muốn yêu cầu xóa khỏi các dịch vụ của Google theo luật hiện hành. Việc cung cấp cho chúng tôi thông tin đầy đủ sẽ giúp chúng tôi điều tra yêu cầu của bạn.

Google cũng bị kiện rất nhiều liên quan đến vấn đề để cho các website ăn cắp nội dung được hiển thị cao trên công cụ tìm kiếm của họ, Vì thế không có gì khó hiểu khi họ rất mạnh tay xử phạt các website chuyên ăn cắp nội dung mà không được sự đồng ý của tác như phim ảnh, tài liệu khoa học,...

3.4.2 Mục tiêu:

- Website vi phạm bản quyền nội dung.
- Website bị báo cáo quá nhiều lần vi phạm bản quyền.

3.4.3 Làm sao để tránh thuật toán Pirate:

- Hãy đảm bảo rằng nội dung bạn xuất bản lại của người khác là không có đăng ký bản quyền hoặc được sự cho phép từ tác giả.

3.5 Thuật toán Hummingbird (chim ruồi):

3.5.1 Giới thiệu:

Thuật toán Google Hummingbird là một bước tiến lớn vượt bậc của công cụ tìm kiếm lớn nhất thế giới với mục đích hiểu được truy vấn tìm kiếm của người dùng, hiểu được nội dung website nói về vấn đề gì để từ đó đưa ra kết quả chính xác nhất.

Trước thời Hummingbird, kết quả tìm kiếm dựa rất nhiều vào từ khóa tìm kiếm của người dùng và sự xuất hiện của từ khóa trong văn bản nội dung website cung cấp.

Các từ khóa truy vấn của người dùng vẫn đóng một vai trò rất quan trọng trong việc đưa ra kết quả tuy nhiên Hummingbird giúp cho Google đưa ra các kết quả mà thậm chí nội dung của page đó không có từ khóa mà người dùng tìm kiếm nhưng nội dung của nó lại nói đến vấn đề mà người dùng đang cần.

Với thuật toán chim ruồi Hummingbird, Google đủ thông minh để hiểu được các từ đồng nghĩa, các từ viết tắt phổ biến.

3.5.2 Mục tiêu:

- Nhồi nhét từ khóa.
- Nhắm mục tiêu từ khóa không chính xác.

3.5.3 Làm sao để tránh thuật toán Hummingbird (chim ruồi):

- **Bước 1: Mở rộng nghiên cứu từ khóa của bạn**

  Với Google HummingBird, bạn có cơ hội thể hiện sự sáng tạo từ khóa cao nhất thay vì chỉ sử dụng những từ khóa ngắn ngủn được gợi ý từ các công cụ phân tích từ khóa như Google Keyword Planner trước đây.

  Bạn cũng không cần phải cố gắng viết nhiều, cố gắng đưa nhiều từ khóa vào nội dung bài viết làm gì, thay vào đó hãy cung cấp nội dung một cách tự nhiên nhất bởi Google đủ thông minh để biết bạn đang nói về vấn đề gì, các từ đồng nghĩa nó cũng hiểu hết. Bài viết về tạo nick Facebook ảo của tôi đó là một ví dụ, Google đủ thông minh đến mức hiểu được rằng nó cần cung cấp nội dung website của tôi cho người dùng dù tôi không hề viết một từ "fb" nào trong bài.

- **Bước 2: Khám phá suy nghĩ của người dùng**

  Bạn cần tìm hiểu người dùng, đặt bản thân vào là người đang đi tìm kiếm dịch vụ xem các từ khóa chúng ta có thể tìm kiếm là gì. Hãy sử dụng thêm các công cụ phân tích tín hiệu mạng xã hội như Awario để xem người dùng tìm kiếm bạn thế nào. Awario sẽ giúp bạn biết làm sao để khách hàng tìm kiếm thương hiệu của bạn, khách hàng nói gì về bạn trên mạng xã hội cũng như đối thủ.

- **Bước 3: Viết nội dung chất lượng, dừng sử dụng các công cụ spin lại:**

  Trước đây, bạn có thể sử dụng các công cụ spin nội dung để xây dựng hệ thống website tuy nhiên điều này ngày càng trở nên nguy hiểm với chúng ta. Tôi không phủ nhận có nhiều người vẫn rất thành công với cách này, tuy nhiên hãy chắc chắn rằng ít nhất bạn phải có data spin chuyên ngành riêng, chỉnh sửa lại tương đối sau khi spin thì mới giảm khả năng bị phát hiện bởi HummingBird được.Nếu các bạn không có nhiều kỹ thuật, tôi khuyên các bạn hãy tự xây dựng nội dung một cách trung thực, đào sâu ý nghĩa để cung cấp cho người dùng sẽ an toàn hơn nhiều.

3.6 Thuật toán Pigeon (chim bồ câu):

Ra mắt: Tháng Bảy 24, 2014 (US)
rollouts: 22 tháng 12 năm 2014 (Anh, Canada, Úc)
Mục tiêu: Cung cấp chất lượng cao, kết quả tìm kiếm địa phương có liên quan

3.6.1 Giới thiệu:

Thuật toán Google Pigeon hiện mới chỉ áp dụng tại các quốc gia sử dụng tiếng Anh là chính như nước Anh, Canada và Úc. Pigeon đã tác động tương đối lớn đến kết quả tìm kiếm của người dùng, một trong những yếu tố quyết định đến kết quả người dùng do thuật toán này gây ảnh hưởng đó chính là vị trí tìm kiếm.

Theo Google, Pigeon tạo ra sự thân thiện, gần gũi hơn giữa các thuật toán địa phương và các thuật toán cốt lõi của họ. Bản cập nhật này sử dụng vị trí là một yếu tố quan trọng để xếp hạng tìm kiếm.

Pigeon làm cho công cụ tìm kiếm trả về kết quả có ít nhất là 50% các website địa phương. Đây là một lợi thế lớn đối với các doanh nghiệp có Google My Business.

Hiện tại, ở Việt Nam chúng ta chưa có thuật toán này. Tuy nhiên, qua những bước đi chuyển hướng công cụ tìm kiếm về các tên miền địa phương của Google tôi tin, khả năng cao không lâu nữa thuật toán này cũng sẽ được phát hành toàn cầu.

3.6.2 Mục tiêu:

- Trang web tối ưu hóa kém.
- Thiết lập Google My Business không đúng cách.
- Mâu thuẫn các thông tin liên hệ của doanh nghiệp.
- Thiếu trích dẫn trong thư mục địa phương (nếu có liên quan).

3.6.3 Làm sao để tránh thuật toán Pigeon:

- **Bước 1: Tối ưu hóa trang web của bạn đúng cách:**

  Pigeon đưa vào các tiêu chuẩn SEO tương tự cho doanh nghiệp địa phương như đối với tất cả các kết quả tìm kiếm khác của Google. Điều đó có nghĩa các doanh nghiệp địa phương bây giờ cần phải đầu tư rất nhiều công sức vào tối ưu hóa trên trang. Hãy khởi đầu tốt đẹp với SEO PowerSuite 's WebSite Auditor khi bạn thực hiện công việc tối ưu hóa này.

  Các công cụ của phân tích nội dung trên bảng điều khiển sẽ cung cấp cho bạn những ý tưởng tốt về các khía cạnh tối ưu hóa trên trang web của bạn, phát hiện ra các lỗi, cảnh báo trên website cho bạn.

- **Bước 2: Thiết lập trang Google My Business:**

  Đầu tiên, bạn cần tạo một trang Google My Business. Tiếp theo bạn cần xác minh được địa điểm để được Google đăng công khai thông tin doanh nghiệp lên trên Google Map.

  Bạn cần chắc chắn rằng mọi thông tin về doanh nghiệp của bạn đưa lên là chính xác, đặc biệt mã vùng, số điện thoại, địa chỉ,...

  Chú ý rằng đánh giá tích cực của người dùng là một điểm cộng lớn giúp bạn có thứ hạng cao trên Google Search. Do đó, hãy tìm cách khuyến khích người dùng đánh giá cho bạn qua các chương trình như khuyến mãi chẳng hạn.

- **Bước 3: Đảm bảo thông tin doanh nghiệp của bạn là nhất quán trên mọi kênh:**

  Google sẽ xem xét thông tin doanh nghiệp của bạn không chỉ trên website doanh nghiệp mà còn cả trên các trang web khác nói về bạn.

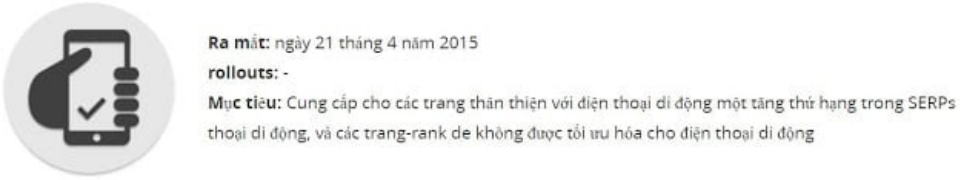
  Bạn cần đảm bảo rằng mọi thông tin đăng tải về doanh nghiệp của bạn đều đúng về địa chỉ, số điện thoại, tên doanh nghiệp, mã vùng,...

  Nếu có sự sai khác, bạn có thể bị Google đánh tụt thứ hạng hoặc thậm chí là không được hiển thị trên Google Search nữa nếu doanh nghiệp của bạn gian dối.

- **Bước 4: Nhận giới thiệu từ các website địa phương:**

  Các website công bố doanh nghiệp địa phương như Yelp, TripAdvisor có tác động rất tốt đến xếp hạng doanh nghiệp của bạn tại địa phương. Vì vậy, bạn hãy tìm cách liên hệ với họ để được đăng tải thông tin doanh nghiệp.

3.7 Thuật toán Mobile Friendly:



3.7.1 Giới thiệu:

Thuật toán Mobile Friendly (thân thiện với di động) được Google ra mắt ngày 21 tháng 4 năm 2015. Kể từ đây, chúng ta sẽ thấy có sự khác biệt tìm kiếm trên điện thoại di động và laptop nếu một website không thân thiện với thiết bị di động. Điều này càng thể hiện rõ hơn sau khi Google cập nhật một loạt thuật toán vào tháng 9 năm 2016. Và cuối cùng đó là sự khác biệt hẳn giữa kết quả tìm kiếm trên máy tính và điện thoại nếu website không thân thiện với di động khi Google bắt đầu lập chỉ mục tìm kiếm đầu tiên cho điện thoại di động. Có rất nhiều chuyên gia dự đoán rằng về lâu dài có khả năng rất cao Google sẽ sử dụng yếu tố trên di động để xếp hạng luôn website khi người dùng sử dụng laptop.

3.7.2 Mục tiêu:

- Website không có phiên bản dành cho thiết bị di động điều này có nghĩa là thiết bị hoàn toàn không thân thiện với di động rồi.

- Chữ quá nhỏ, hoặc các dòng quá gần nhau dẫn đến người dùng khó đọc.
- Sử dụng nhiều Plugin.
- Cấu hình khung hình chế độ xem không tốt

3.7.3 Làm sao để tránh thuật toán Mobile Friendly:

- Ngay từ khi thiết kế website bạn hãy đảm bảo rằng website của bạn thân thiện với di động. Để kiểm tra độ thân thiện với di động Google cung cấp tool miễn phí cho chúng ta sử dụng đó là công cụ Kiểm tra tính thân thiện của di động.

3.8 Thuật toán RankBrain:

Ra mắt: ngày 26 tháng 10 năm 2015 (có thể sớm hơn)
rollouts: -
Mục tiêu: Cung cấp kết quả tìm kiếm tốt hơn dựa trên sự liên quan & máy học tập

3.8.1 Giới thiệu:

RankBrain là hệ thống máy móc giúp Google phân tích, giải mã các ý nghĩa những truy vấn tìm kiếm khác nhau của người dùng để từ đó đưa ra kết quả chính xác nhất trên công cụ tìm kiếm.

Chỉ có một thành phần xử lý truy vấn trong RankBrain nhưng đó cũng chính là yếu tố xếp hạng của thuật toán này. Lần đầu tiên khi công bố thuật toán RankBrain, Google khẳng định rằng đây là một trong 3 yếu tố xếp hạng quan trọng nhất của họ. Bằng cách nào đó, RankBrain có thể đánh giá tóm tắt được nội dung của một website để rồi từ đó đưa ra kết quả tốt nhất liên quan đến tìm kiếm của người dùng.

Về mặt cơ bản, RankBrain cũng xếp hạng website dựa vào các yếu tố truyền thống như backlink, tối ưu on page,... Ngoài ra, RankBrain còn xem xét đánh giá các truy vấn cụ thể khác để từ đó đưa ra kết quả liên quan nhất, chất lượng nhất cho người dùng.

3.8.2 Mục tiêu:

- Các website thiếu sự liên quan đến truy vấn cụ thể.
- Website cung cấp trải nghiệm người dùng kém.

3.8.3 Làm sao để tránh thuật toán RankBrain:

- **Bước 1: Tối ưu hóa trải nghiệm người dùng:**

Dĩ nhiên, RankBrain không phải là lý do để bạn đem lại trải nghiệm cho người dùng tốt hơn. Nhưng Google sinh ra là để làm hài lòng người dùng, chúng ta muốn có thứ hạng cao hơn trên công cụ tìm kiếm tại sao lại không cố gắng làm hài lòng người dùng, bởi nó sẽ ảnh hưởng trực tiếp đến thứ hạng của bạn.

Google Analytics là một công cụ tuyệt vời để phân tích dữ liệu người dùng từ đó đánh giá mức độ phát triển của một website đến đâu. Hai chỉ số mà tôi đặc biệt quan tâm khi phân tích hành vi người dùng đó là tỷ lệ thoát và thời gian xem trang web của khách hàng. Nếu 2 chỉ số này quá xấu, điều chắc chắn là thứ hạng của bạn sẽ không bao giờ cao được.

- **Bước 2: Nghiên cứu đối thủ cạnh tranh:**

  Một trong những cách xếp hạng quan trọng của RankBrain đó là phân tích các truy vấn của người dùng vào từng website rồi sử dụng nó như một tín hiệu để đưa ra kết quả tìm kiếm chính xác nhất. Như vậy, tính năng này của RankBrain sẽ giúp Google hiểu được bất kỳ hành động nào của người dùng trên website của bạn. Vì vậy bạn cần có những đánh giá quan sát trực tiếp không chỉ bản thân mà còn cả đối thủ xem traffic họ đến từ đâu chẳng hạn,. . . và còn rất nhiều điều nữa.

3.9 Thuật toán Possum:

Ra mắt: ngày 01 Tháng Chín năm 2016
rollouts: -
Mục tiêu: Cung cấp, kết quả đa dạng hơn tốt hơn dựa trên vị trí của người tìm kiếm và địa chỉ doanh nghiệp

3.9.1 Giới thiệu:

Bản cập nhật gây xôn xao, khiến cộng đồng SEO chao đảo vào đầu tháng 9 năm ngoái của Google khiến kết quả tìm kiếm địa phương thay đổi chóng mặt do đó nó được đặt tên là thuật toán Possum.Sau khi Possum được cập nhật, Google Search đưa ra kết quả tìm kiếm phong phú hơn nhiều dựa vào vị trí tìm kiếm của người dùng.

Nếu bạn ở gần doanh nghiệp cung cấp dịch vụ mà bạn đang quan tâm thì doanh nghiệp đó có khả năng xuất hiện trên công cụ tìm kiếm rất cao. Hơi nghịch lý một chút là chính Possum cũng khuyến khích các doanh nghiệp ngoài địa phương tăng cường sự hiện diện ở nơi khác.

Trước đây, nếu doanh nghiệp của bạn nằm ngoài vùng địa lý sẽ không được ưu tiên hiển thị và đưa vào danh sách địa phương tuy nhiên bây giờ điều này đã thay đổi, bạn vẫn có thể tăng cường hiện diện ở địa phương khác. Ngoài ra các doanh nghiệp chia sẽ một doang nghiệp khác cung cấp dịch vụ liên quan truy vấn của người dùng cũng có cơ hội được xuất hiện trên tìm kiếm địa phương.

3.9.2 Mục tiêu:

- Xử lý các doanh nghiệp trùng lặp địa chỉ, cung cấp dịch vụ tương tự nhau.
- Đối thủ cạnh tranh có địa chỉ gần người tìm kiếm hơn.

3.9.3 Làm sao để tránh thuật toán Possum:

- **Bước 1: Theo dõi xếp hạng địa phương cụ thể:**

  Sau khi thuật toán Possum được cập nhật, vị trí tìm kiếm của bạn đóng một vai trò to lớn có sức ảnh hưởng mạnh đến công cụ tìm kiếm.

  Hãy sử dụng các công cụ để theo dõi xếp hạng tìm kiếm địa phương của bạn bằng cách sử dụng các công cụ như SEO PowerSuite 's Rank Tracker.

- **Bước 2: Phát triển danh sách từ khóa tìm kiếm địa phương:**

  Possum làm xuất hiện nhiều kết quả tìm kiếm với các từ khóa tương tự nhau. Vì thế bạn hãy cố gắng phát triển được từ khóa mở rộng đến các suy nghĩ của người dùng. Chú ý là dùng từ khóa mở rộng bao hàm nhiều từ khóa chứ không phải cố gắng nhồi nhét, spam từ khóa nhé.

3.10 Thuật toán Fred:

Ra mắt: ngày 08 tháng 3 năm 2017
rollouts: -
Mục tiêu: Lọc ra kết quả tìm kiếm chất lượng thấp có mục đích duy nhất là tạo ra doanh thu quảng cáo và liên kết

3.10.1 Giới thiệu:

Thuật toán Fred là một cập nhật của Google vào ngày mùng 8 tháng 3 năm 2017 làm cộng đồng SEO trên toàn cầu chao đảo, có rất nhiều website giảm traffic đến mức chỉ còn bằng 1 phần 10 trước đây. Cái tên Fred là do Gary Illyes của Google nói đùa trên Twitter của ông. Google đã xác nhận sự có mặt của thuật toán này, tuy nhiên họ từ chối đưa ra thảo luận về cập nhật này với lý do là để đảm bảo chất lượng tìm kiếm cho người dùng.

Kết quả nghiên cứu từ các công cụ theo dõi Google tự động cho thấy thuật toán Fred hướng đến các website chỉ có mục đích quảng cáo là chính, spam liên kết, chất lượng nội dung kém mà không hướng đến người dùng.

3.10.2 Mục tiêu:

- Website có nội dung kém, quảng cáo quá nhiều.
- Hồ sơ liên kết kém chất lượng

3.10.3 Làm sao để tránh thuật toán Fred:

- **Bước 1: Thực hiện đúng nội quy quản trị website của Google:**

  Google không đưa ra bình luận về thuật toán này tuy nhiên họ có nhắc nhở webmaster toàn cầu rằng hãy thực hiện công việc quản trị đúng với Nguyên tắc quản trị trang web của họ.

- **Bước 2: Xây dựng nội dung website chất lượng:**

  Các website bị đánh giá là chất lượng nội dung kém đều trở thành miếng mồi béo của Fred. Hãy cố gắng xây dựng nội dung chất lượng lên website của bạn để dữ chân khách hàng và nâng cao uy tín của bạn với Google.

  Đừng tham lam đặt quảng cáo quá nhiều, Đây là một điều Google rất ghét và ra sức ngăn chặn từ khi Google Adsense ra đời.

  Bạn cần phải hiểu rõ SEO là gì? Và các yếu tố chất lượng gây ảnh hưởng lớn trong suốt quá trình SEO, nếu đáp ứng tốt thì việc có một thứ hạng tốt trên Google, và ngược chắc bạn cũng sẽ hiểu website của bạn đang ở đây.

  Lời khuyên tốt nhất của chúng tôi là bạn nên tham gia các khóa đào tạo SEO chuyên sâu, hoặc các buổi tư vấn SEO từ các chuyên gia trong lĩnh vực này, hãy luôn cập nhật kiến thức thì bạn không còn phải lo lắng tới các thuật toán của Google có thể xảy ra đối với bạn.

4 Conclusion

4.1 Đánh giá tổng quát

Ngày nay, với sự phát triển ngày càng mạnh mẽ, Internet đã đóng một vai trò quan trọng trong đời sống của con người. Internet mang lại nhiều tiện ích hữu dụng cho người sử dụng như hệ thống thư điện tử, trò chuyện trực tuyến, thương mại điện tử, các dịch vụ về y tế, giáo dục ... Với một kho dữ liệu khổng lồ từ các máy chủ và liên mạng máy tính toàn cầu, để khai thác được tối ưu thông tin cần tìm kiếm, các công ty phần mềm đã viết ra một phần mềm gọi là: Công cụ tìm kiếm, đặc biệt phải kể đến Công cụ tìm kiếm **Google Search** chiếm gần như tuyệt đối thị phần tìm kiếm trên Internet (xấp xỉ 78%).

4.2 Vậy điều gì đã làm nền sự vượt trội này

- Google được nhiều người đánh giá là công cụ tìm kiếm hữu ích và mạnh mẽ nhất trên Internet nhờ có lượng máy chủ khổng lồ cùng với công nghệ tốt.
- Có thể tìm kiếm thông tin dưới nhiều hình thức khác nhau, kết quả tìm được bao gồm nhiều định dạng, chẳng hạn như văn bản, ảnh, video clip ...
- Người dùng có thể chọn lọc được đúng đối tượng mình cần tìm trong vô vàn nguồn thông tin chỉ trong thời gian cực kì ngắn. Điều này giúp cho quá trình thu thập thông tin được nhanh hơn, nhiều lần hơn, trong mỗi lần tìm kiếm sẽ thấy được vị trí thứ tự ưu tiên của thông tin.
- Người dùng có thêm thông tin liên quan đến đối tượng đang tìm kiếm để từ đó gợi mở và tạo ý tưởng cho một hướng làm việc mới hay hơn.
- Thuật toán, công cụ tìm kiếm luôn không ngừng nâng cao, phát triển từng ngày để phù hợp với sự phát triển của Internet và nhu cầu sử dụng của người dùng.

4.3 Nhưng bên cạnh đó vẫn tồn tại một vài nhược điểm:

- Không thu thập được những thông tin mang tính chất nhất thời, địa phương, chẳng hạn như tình hình các gian hàng, giá cả trong khu chợ địa phương hay đường đi ở các vùng rừng núi ít người ở.
- Thời gian cập nhật và thay đổi những thông tin mới còn dài, cần có những người sử dụng Internet (đặc

biệt là Google) trực tiếp thay đổi cài đặt hay cập nhật thông tin, dẫn đến thiếu chính xác ở một vài thời điểm. Ví dụ như:

- Việc thay đổi, xây mới hoặc loại bỏ các tuyến đường nếu không được người dùng cập nhật lên Internet sẽ cần thời gian dài để tự thay đổi ở Google Map.
- Khi người dùng chuyển địa điểm đang ở sang một nơi khác, nếu không thay đổi địa điểm trên cài đặt tìm kiếm của Google sẽ vẫn phải nhận hầu hết các thông tin, quảng cáo liên quan đến nơi ở trước mà không phải địa điểm đang ở hiện tại.
- Việc cài đặt địa điểm, ngôn ngữ tìm kiếm giúp người dùng khi tìm kiếm các thông tin sẽ được Google ưu tiên đưa ra các trang phù hợp vs địa điểm và ngôn ngữ đó. Tuy nhiên, nếu người dùng muốn tìm kiếm thông tin ở một nơi khác, nước khác,... sẽ gặp phải khó khăn vì các thông tin này không hoàn toàn được đưa ra khi tìm kiếm vì nó không phải là địa điểm hay ngôn ngữ đã cài đăt.
- Mọi người đều biết Google thu thập thông tin khi người dùng tìm kiếm trên Search để cập nhật vào việc xếp hạng trang hữu ích và đưa ra quảng cáo phù hợp. Tuy nhiên việc đưa ra thông tin dựa trên yếu tố như vậy không hoàn toàn đúng vì không thể xác định được người đó đang có nhu cầu, đã có nhu cầu hay chưa có nhu cầu. Việc đưa ra các trang web mà người dùng đang không có nhu cầu tìm hiểu dẫn đến việc người dùng xem các trang này là trang Web spam dù nội dung của nó hoàn toàn hữu ích và chất lượng.

- Chưa kiểm soát được chặt chẽ các trang Web, outbound links mang tính chất spam và chứa nhiều virus,...

*Cuối cùng, để kết thúc nội dung trình bày cho bài tập lớn lần này, Nhóm TUT xin kết thúc bằng một câu thơ lục bát, vừa để nêu ra xu hướng, vừa khẳng định vai trò của công cụ tìm kiếm Google trong đời sống hiện nay: ....

Tư liệu tham khảo:

1. Spider Crawling: https://atpsoftware.vn/cach-hoat-dong-cua-bo-may-tim-kiem-google.html

2. How Search Works: https://www.google.com/search/howsearchworks/

3. Thuật toán PageRank:
https://www.slideshare.net/manman89/thut-ton-pagerank
https://vi.wikipedia.org/wiki/PageRank

4. Các thuật toán tìm kiếm của Google:
https://www.seonamnguyen.com/9-thuat-toan-anh-huong-toi-seo-noi-tieng-cua-google.html

Tài liệu liên quan để người đọc tham khảo thêm:

1. Cơ chế hoạt động của Google:
https://support.google.com/webmasters/answer/70897?hl=vi
http://dautuseo.com/google-search-hoat-dong-nhu-the-nao/

2. Crawling và Indexing: https://dieuhau.com/google-crawling-va-indexing/