

Google's Trust Ranking Patent Shows How User Behavior Is A Signal via @sejournal, @martinibuster

Google long ago filed a patent for ranking search results by trust. The groundbreaking idea behind the patent is that user behavior can be used as a starting point for developing a ranking signal.

The big idea behind the patent is that the Internet is full of websites all linking to and commenting about each other. But which sites are trustworthy? Google's solution is to utilize user behavior to indicate which sites are trusted and then use the linking and content on those sites to reveal more sites that are trustworthy for any given topic.

PageRank is basically the same thing only it begins and ends with one website linking to another website. The innovation of Google's trust ranking patent is to put the user at the start of that trust chain like this:

User trusts X Websites > X Websites trust Other Sites > This feeds into Google as a ranking signal

The trust originates from the user and flows to trust sites that themselves provide anchor text, lists of other sites and commentary about other sites.

That, in a nutshell, is what Google's trust-based ranking algorithm is about.

The deeper insight is that it reveals Google's groundbreaking

approach to letting users be a signal of what's trustworthy. You know how Google keeps saying to create websites for users? This is what the trust patent is all about, putting the user in the front seat of the ranking algorithm.

Google's Trust And Ranking Patent

The patent was coincidentally filed around the same period that Yahoo and Stanford University published a Trust Rank research paper which is focused on identifying spam pages.

Google's patent is not about finding spam. It's focused on doing the opposite, identifying trustworthy web pages that satisfy the user's intent for a search query.

How Trust Factors Are Used

The first part of any patent consists of an Abstract section that offers a very general description of the invention that that's what this patent does as well.

The patent abstract asserts:

- That trust factors are used to rank web pages.
- The trust factors are generated from "entities" (which are later described to be the users themselves, experts, expert web pages, and forum members) that link to or comment about other web pages).
- Those trust factors are then used to re-rank web pages.
- Re-ranking web pages kicks in after the normal ranking algorithm has done its thing with links, etc.

Here's what the Abstract says:

"A search engine system provides search results that are ranked according to a measure of the trust associated with entities that have provided labels for the documents in the search results.

A search engine receives a query and selects documents relevant to the query.

The search engine also determines labels associated with selected documents, and the trust ranks of the entities that provided the labels.

The trust ranks are used to determine trust factors for the respective documents. The trust factors are used to adjust information retrieval scores of the documents. The search results are then ranked based on the adjusted information retrieval scores.”

As you can see, the Abstract does not say who the “entities” are nor does it say what the labels are yet, but it will.

Field Of The Invention

The next part is called the Field Of The Invention. The purpose is to describe the technical domain of the invention (which is information retrieval) and the focus (trust relationships between users) for the purpose of ranking web pages.

Here’s what it says:

“The present invention relates to search engines, and more specifically to search engines that use information indicative of trust relationship between users to rank search results.”

Now we move on to the next section, the Background, which describes the problem this invention solves.

Background Of The Invention

This section describes why search engines fall short of

answering user queries (the problem) and why the invention solves the problem.

The main problems described are:

- Search engines are essentially guessing (inference) what the user's intent is when they only use the search query.
- Users rely on expert-labeled content from trusted sites (called vertical knowledge sites) to tell them which web pages are trustworthy
- Explains why the content labeled as relevant or trustworthy is important but ignored by search engines.
- It's important to remember that this patent came out before the BERT algorithm and other natural language approaches that are now used to better understand search queries.

This is how the patent explains it:

“An inherent problem in the design of search engines is that the relevance of search results to a particular user depends on factors that are highly dependent on the user's intent in conducting the search—that is why they are conducting the search—as well as the user's circumstances, the facts pertaining to the user's information need.

Thus, given the same query by two different users, a given set of search results can be relevant to one user and irrelevant to another, entirely because of the different intent and information needs.”

Next it goes on to explain that users trust certain websites that provide information about certain topics:

“...In part because of the inability of contemporary search engines to consistently find information that satisfies the user's information need, and not merely the user's query terms, users frequently turn to websites that offer

additional analysis or understanding of content available on the Internet.”

Websites Are The Entities

The rest of the Background section names forums, review sites, blogs, and news websites as places that users turn to for their information needs, calling them vertical knowledge sites. Vertical Knowledge sites, it’s explained later, can be any kind of website.

The patent explains that trust is why users turn to those sites:

“This degree of trust is valuable to users as a way of evaluating the often bewildering array of information that is available on the Internet.”

To recap, the “Background” section explains that the trust relationships between users and entities like forums, review sites, and blogs can be used to influence the ranking of search results. As we go deeper into the patent we’ll see that the entities are not limited to the above kinds of sites, they can be any kind of site.

Patent Summary Section

This part of the patent is interesting because it brings together all of the concepts into one place, but in a general high-level manner, and throws in some legal paragraphs that explain that the patent can apply to a wider scope than is set out in the patent.

The Summary section appears to have four sections:

- The first section explains that a search engine ranks web pages that are trusted by entities (like forums,

news sites, blogs, etc.) and that the system maintains information about these labels about trusted web pages.

- The second section offers a general description of the work of the entities (like forums, news sites, blogs, etc.).
- The third offers a general description of how the system works, beginning with the query, the assorted hand waving that goes on at the search engine with regard to the entity labels, and then the search results.
- The fourth part is a legal explanation that the patent is not limited to the descriptions and that the invention applies to a wider scope. This is important. It enables Google to use a non-existent thing, even something as nutty as a “trust button” that a user selects to identify a site as being trustworthy as an example. This enables an example like a non-existent “trust button” to be a stand-in for something else, like navigational queries or Navboost or anything else that is a signal that a user trusts a website.

Here's a nutshell explanation of how the system works:

- The user visits sites that they trust and click a “trust button” that tells the search engine that this is a trusted site.
- The trusted site “labels” other sites as trusted for certain topics (the label could be a topic like “symptoms”).
- A user asks a question at a search engine (a query) and uses a label (like “symptoms”).
- The search engine ranks websites according to the usual manner then it looks for sites that users trust and sees if any of those sites have used labels about other sites.
- Google ranks those other sites that have had labels assigned to them by the trusted sites.

Here's an abbreviated version of the third part of the Summary

that gives an idea of the inner workings of the invention:

“A user provides a query to the system...The system retrieves a set of search results... The system determines which query labels are applicable to which of the search result documents. ... determines for each document an overall trust factor to apply... adjusts the ...retrieval score... and reranks the results.”

Here's that same section in its entirety:

- “A user provides a query to the system; the query contains at least one query term and optionally includes one or more labels of interest to the user.
- The system retrieves a set of search results comprising documents that are relevant to the query term(s).
- The system determines which query labels are applicable to which of the search result documents.
- The system determines for each document an overall trust factor to apply to the document based on the trust ranks of those entities that provided the labels that match the query labels.
- Applying the trust factor to the document adjusts the document's information retrieval score, to provide a trust adjusted information retrieval score.
- The system reranks the search result documents based at on the trust adjusted information retrieval scores.”

The above is a general description of the invention.

The next section, called Detailed Description, deep dives into the details. At this point it's becoming increasingly evident that the patent is highly nuanced and can not be reduced to simple advice similar to: “optimize your site like this to earn trust.”

A large part of the patent hinges on a trust button and an advanced search query: label:

Neither the trust button or the label advanced search query have ever existed. As you'll see, they are quite probably stand-ins for techniques that Google doesn't want to explicitly reveal.

Detailed Description In Four Parts

The details of this patent are located in four sections within the Detailed Description section of the patent. This patent is not as simple as 99% of SEOs say it is.

These are the four sections:

1. System Overview
2. Obtaining and Storing Trust Information
3. Obtaining and Storing Label Information
4. Generated Trust Ranked Search Results

The System Overview is where the patent deep dives into the specifics. The following is an overview to make it easy to understand.

System Overview

1. Explains how the invention (a search engine system) ranks search results based on trust relationships between users and the user-trusted entities who label web content.
2. The patent describes a "trust button" that a user can click that tells Google that a user trusts a website or trusts the website for a specific topic or topics.
3. The patent says a trust related score is assigned to a website when a user clicks a trust button on a website.
4. The trust button information is stored in a trust database that's referred to as #190.

Here's what it says about assigning a trust rank score based

on the trust button:

“The trust information provided by the users with respect to others is used to determine a trust rank for each user, which is measure of the overall degree of trust that users have in the particular entity.”

Trust Rank Button

The patent refers to the “trust rank” of the user-trusted websites. That trust rank is based on a trust button that a user clicks to indicate that they trust a given website, assigning a trust rank score.

The patent says:

“...the user can click on a “trust button” on a web page belonging to the entity, which causes a corresponding record for a trust relationship to be recorded in the trust database 190.

In general any type of input from the user indicating that such as trust relationship exists can be used.”

The trust button has never existed and the patent quietly acknowledges this by stating that any type of input can be used to indicate the trust relationship.

So what is it? I believe that the “trust button” is a stand-in for user behavior metrics in general, and site visitor data in particular. The patent Claims section does not mention trust buttons at all but does mention user visitor data as an indicator of trust.

Here are several passages that mention site visits as a way to understand if a user trusts a website:

“The system can also examine web visitation patterns of the

user and can infer from the web visitation patterns which entities the user trusts. For example, the system can infer that a particular user trust a particular entity when the user visits the entity's web page with a certain frequency."

The same thing is stated in the Claims section of the patent, it's the very first claim they make for the invention:

*"A method performed by data processing apparatus, the method comprising:
determining, based on web visitation patterns of a user, one or more trust relationships indicating that the user trusts one or more entities;"*

It may very well be that site visitation patterns and other user behaviors are what is meant by the "trust button" references.

Labels Generated By Trusted Sites

The patent defines trusted entities as news sites, blogs, forums, and review sites, but not limited to those kinds of sites, it could be any other kind of website.

Trusted websites create references to other sites and in that reference they label those other sites as being relevant to a particular topic. That label could be an anchor text. But it could be something else.

The patent explicitly mentions anchor text only once:

"In some cases, an entity may simply create a link from its site to a particular item of web content (e.g., a document) and provide a label 107 as the anchor text of the link."

Although it only explicitly mentions anchor text once, there are other passages where it anchor text is strongly implied,

for example, the patent offers a general description of labels as describing or categorizing the content found on another site:

“...labels are words, phrases, markers or other indicia that have been associated with certain web content (pages, sites, documents, media, etc.) by others as descriptive or categorical identifiers.”

Labels And Annotations

Trusted sites link out to web pages with labels and links. The combination of a label and a link is called an annotation.

This is how it's described:

“An annotation 106 includes a label 107 and a URL pattern associated with the label; the URL pattern can be specific to an individual web page or to any portion of a web site or pages therein.”

Labels Used In Search Queries

Users can also search with “labels” in their queries by using a non-existent “label:” advanced search query. Those kinds of queries are then used to match the labels that a website page is associated with.

This is how it's explained:

“For example, a query “cancer label:symptoms” includes the query term “cancel” and a query label “symptoms”, and thus is a request for documents relevant to cancer, and that have been labeled as relating to “symptoms.”

Labels such as these can be associated with documents from any entity, whether the entity created the document, or is a

third party. The entity that has labeled a document has some degree of trust, as further described below.”

What is that label in the search query? It could simply be certain descriptive keywords, but there aren’t any clues to speculate further than that.

The patent puts it all together like this:

“Using the annotation information and trust information from the trust database 190, the search engine 180 determines a trust factor for each document.”

Takeaway:

A user’s trust is in a website. That user-trusted website is not necessarily the one that’s ranked, it’s the website that’s linking/trusting another relevant web page. The web page that is ranked can be the one that the trusted site has labeled as relevant for a specific topic and it could be a web page in the trusted site itself. The purpose of the user signals is to provide a starting point, so to speak, from which to identify trustworthy sites.

Experts Are Trusted

Vertical Knowledge Sites, sites that users trust, can host the commentary of experts. The expert could be the publisher of the trusted site as well. Experts are important because links from expert sites are used as part of the ranking process.

Experts are defined as publishing a deep level of content on the topic:

“These and other vertical knowledge sites may also host the analysis and comments of experts or others with knowledge, expertise, or a point of view in particular fields, who again

can comment on content found on the Internet.

For example, a website operated by a digital camera expert and devoted to digital cameras typically includes product reviews, guidance on how to purchase a digital camera, as well as links to camera manufacturer's sites, new products announcements, technical articles, additional reviews, or other sources of content.

To assist the user, the expert may include comments on the linked content, such as labeling a particular technical article as "expert level," or a particular review as "negative professional review," or a new product announcement as ;new 10MP digital SLR'."

Links From Expert Sites

Links and annotations from user-trusted expert sites are described as sources of trust information:

"For example, Expert may create an annotation 106 including the label 107 "Professional review" for a review 114 of Canon digital SLR camera on a web site "www.digitalcameraworld.com", a label 107 of "Jazz music" for a CD 115 on the site "www.jazzworld.com", a label 107 of "Classic Drama" for the movie 116 "North by Northwest" listed on website "www.movierental.com", and a label 107 of "Symptoms" for a group of pages describing the symptoms of colon cancer on a website 117 "www.yourhealth.com".

Note that labels 107 can also include numerical values (not shown), indicating a rating or degree of significance that the entity attaches to the labeled document.

Expert's web site 105 can also include trust information. More specifically, Expert's web site 105 can include a trust list 109 of entities whom Expert trusts. This list may be in the form of a list of entity names, the URLs of such

entities' web pages, or by other identifying information. Expert's web site 105 may also include a vanity list 111 listing entities who trust Expert; again this may be in the form of a list of entity names, URLs, or other identifying information."

Inferred Trust

The patent describes additional signals that can be used to signal (infer) trust. These are more traditional type signals like links, a list of trusted web pages (maybe a resources page?) and a list of sites that trust the website.

These are the inferred trust signals:

*"(1) links from the user's web page to web pages belonging to trusted entities;
(2) a trust list that identifies entities that the user trusts; or
(3) a vanity list which identifies users who trust the owner of the vanity page."*

Another kind of trust signal that can be inferred is from identifying sites that a user tends to visit.

The patent explains:

"The system can also examine web visitation patterns of the user and can infer from the web visitation patterns which entities the user trusts. For example, the system can infer that a particular user trusts a particular entity when the user visits the entity's web page with a certain frequency."

Takeaway:

That's a pretty big signal and I believe that it suggests that promotional activities that encourage potential site visitors

to discover a site and then become loyal site visitors can be helpful. For example, that kind of signal can be tracked with branded search queries. It could be that Google is only looking at site visit information but I think that branded queries are an equally trustworthy signal, especially when those queries are accompanied by labels... ding, ding, ding!

The patent also lists some kind of out there examples of inferred trust like contact/chat list data. It doesn't say social media, just contact/chat lists.

Trust Can Decay or Increase

Another interesting feature of trust rank is that it can decay or increase over time.

The patent is straightforward about this part:

"Note that trust relationships can change. For example, the system can increase (or decrease) the strength of a trust relationship for a trusted entity. The search engine system 100 can also cause the strength of a trust relationship to decay over time if the trust relationship is not affirmed by the user, for example by visiting the entity's web site and activating the trust button 112."

Trust Relationship Editor User Interface

Directly after the above paragraph is a section about enabling users to edit their trust relationships through a user interface. There has never been such a thing, just like the non-existent trust button.

This is possibly a stand-in for something else. Could this trusted sites dashboard be Chrome browser bookmarks or sites that are followed in Discover? This is a matter for

speculation.

Here's what the patent says:

"The search engine system 100 may also expose a user interface to the trust database 190 by which the user can edit the user trust relationships, including adding or removing trust relationships with selected entities.

The trust information in the trust database 190 is also periodically updated by crawling of web sites, including sites of entities with trust information (e.g., trust lists, vanity lists); trust ranks are recomputed based on the updated trust information."

What Google's Trust Patent Is About

Google's Search Result Ranking Based On Trust patent describes a way of leveraging user-behavior signals to understand which sites are trustworthy. The system then identifies sites that are trusted by the user-trusted sites and uses that information as a ranking signal. There is no actual trust rank metric, but there are ranking signals related to what users trust. Those signals can decay or increase based on factors like whether a user still visits those sites.

The larger takeaway is that this patent is an example of how Google is focused on user signals as a ranking source, so that they can feed that back into ranking sites that meet their needs. This means that instead of doing things because "this is what Google likes," it's better to go even deeper and do things because users like it. That will feed back to Google through these kinds of algorithms that measure user behavior patterns, something we all know Google uses.

Featured Image by Shutterstock/samsulalam

<https://www.searchenginejournal.com/googles-trust-ranking-pate>

nt-shows-how-user-behavior-is-a-signal/550203/