A Complete Guide to the Google RankBrain Algorithm by @schachin

Editor's note: *This post is part of an ongoing series looking back at the history of Google algorithm updates. Enjoy!*

The Google RankBrain algorithm went online in April 2015 but wasn't introduced to the world until October 2015 via a Bloomberg news story. Here's how RankBrain was described at the time in the article:

> *"RankBrain uses artificial intelligence to embed vast amounts of written language into mathematical entities – called vectors – that the computer can understand. If RankBrain sees a word or phrase it isn't familiar with, the machine can make a guess as to what words or phrases might have a similar meaning and filter the result accordingly, making it more effective at handling never-before-seen search queries."*
>
> *ADVERTISEMENT*

RankBrain is the only live Artificial Intelligence (AI) that Google uses in its search results. While Google uses machine learning to teach the algorithms, AI isn't being used in the wild – and for good reason. If search broke, Google's engineers would have no clue how to fix it.

RankBrain, however, is used to sort live search results to help give users a best fit to their search query.

RankBrain as a Ranking Signal

RankBrain has been called Google's third most important ranking signal (behind content and links).

But is RankBrain *really* a "ranking signal"?

Not really. At least not in the way we think of traditional ranking signals.

RankBrain is a method of processing search queries in a way that infers a "best fit" for queries that are unknown to Google.

About 15 percent of the queries Google processes every day are new – in other words, nobody has ever searched using these exact terms before.

How can there be so many unknown queries? It's a hard concept to wrap your brain around.

But if you think about all the different ways we talk about a person, place, or thing you can quickly see how there could be millions of way to ask even one simple question. This will likely even expand exponentially as we move more to voice search as smartphones get better at Voice to Text and devices move into the home that take only voice.

So, in the simplest terms, RankBrain is a processing algorithm that uses machine learning to bring back the best match to your query when it isn't sure what that query "means."

At first, RankBrain was only present in a small number of Google queries (about 15 percent). However, over time, it has expanded and is involved in almost all queries entered into Google.

That being said, if Google is sure of the query meaning RankBrain has very little influence. RankBrain is only there

to help when Google is unsure of the queries meaning.

What Does it Mean for Google to ‘Know’ a Query Set?

When Google rolled out Hummingbird and moved from “strings to things”, it moved from inferring a match to your search query by using on-page and off-page factors to understanding relationships of people, places, and things to each other by seeding the algorithm with known relationships.

This was in part determined from a database called Freebase at first. Then Google used WikiData. Now they use data fed machine learning for the most part.

How does this work?

This means that instead of determining your article about “red apples” was about red apples from optimization signals such as inbound link anchor text and H1 tags, it already knew that a red apple was a round edible fruit that came in the color known as red.

The database told Google that this string was actually a thing called “red apple”. Then Google can pull back all the best match results for the term “red apple.”

However, maybe you meant “red apple” as in a “red apple computer.” If Google isn’t sure you meant “apple the fruit” or “apple the computer”, it might throw a few alternate results into your query set.

So instead of 10 fruit related results, you might get 8 fruit related and 2 computer related, or vice versa.

This is how Google RankBrain works in the most basic of ways.

When Is RankBrain Heavily Influencing a Query Result?

RankBrain impacts queries in all languages and all countries.

When RankBrain is most in “play” is when the query is unique and unknown.

For instance, before RankBrain was announced, I had written an article about something I was observing in my own Google searches.

It started when I was searching for information on water rights in Nevada during the California drought. (We share a river with them). When I looked up Clark County or Las Vegas water rights, there was a lot of information on Google related to the topic. However, when I searched Mesquite NV water rights (a town 90 miles north) I got back the water authority and nothing related to water rights. Instead, I got pages on mesquite trees, mesquite wood, mesquite barbecue chips, etc.

At the time I didn’t know what it was called, just that this existed. However, this is what we now know as a result where RankBrain was in full effect.

Why? Because Google did not know what the relationship between the “thing or place” mesquite and the “thing” water rights was, it sent back a “kitchen sink” of results.

The idea of the “best guess” kitchen sink is that over time Google will learn what would be the best match to that query.

If you’ve been in search long enough you might remember when you would do a search and Google would show you what words it actually used in that search (despite what you typed in). This was the precursor to RankBrain.

What Google RankBrain Is Not

Up until now, we've discussed RankBrain in general, non-specific, layman's terms.

So what really happens behind the scenes?

RankBrain is not a Natural Language Processor, otherwise known as NLP.

NLP is the holy grail of search, where a computer can break down full sentences and understand the intent of the user from the users sentence structure and linguistics.

Words infer meaning on other words and the NLP can understand language similar to the way a human does, albeit through a different process.

While RankBrain is a step closer toward that ultimate goal, RankBrain can't infer meaning from your searches based on language alone.

RankBrain requires a database of relationships, and vectors of known relationships between similar queries, to pull back a best guess. Inference occurs when the queries are not understood, but the results returned are still based on that data.

So How Does RankBrain Actually Work?

RankBrain used a series of databases based on people, places, and things (otherwise known as entities) to seed the algorithm and its machine learning processes.

These words (queries) are then broken down into word vectors using a mathematical formula to give those words an "address". Similar words share similar "addresses."

When Google processes an unknown query, it uses these mathematically mapped out relationships to assume a best fit to the query and returns multiple related results.

Over time Google refines the results based on user interaction and machine learning to improve the match between users search intent and the search results that Google returns.

It's important to note that words that search engines used to throw away like "and" or "the" are not in RankBrain's analysis. RankBrain is also meant to help better understand queries to deliver the best search results, particularly for negative-oriented queries, such as queries using words like "without" or "not."

Also, as explained on The Next Web:

> *"RankBrain converts the textual contents of search queries into 'word vectors,' also known as 'distributed representations,' each of which has a unique coordinate address in mathematical space. Vectors close to each other in this space correspond to linguistic similarity."*

The process at a mathematical level is much more involved, but at a process summary level, it's not overly complex.

Words go in. Words get assigned a mathematical address. Words are retrieved based on your query and the words it locates in the "best fit" vector.

These word "interpretations" are used to return results.

Behind the scenes, data is continually fed into the machine learning process, so as to make those results more relevant the next time.

Simple on the surface, but incredibly complicated and difficult at the microlevel.

Can You Optimize for RankBrain?

Google's Gary Illyes tells us we can optimize for RankBrain by just writing naturally:

> *"Optimizing for RankBrain is actually super easy, and it is something we've probably been saying for 15 years now, is – and the recommendation is – to write in natural language. Try to write content that sounds human. If you try to write like a machine then RankBrain will just get confused and probably just pushes you back.*
>
> *But if you have a content site, try to read out some of your articles or whatever you wrote, and ask people whether it sounds natural. If it sounds conversational, if it sounds like natural language that we would use in your day to day life, then sure, you are optimized for RankBrain. If it doesn't, then you are 'un-optimized.'"*

However, if you were always writing good content, you are probably asking what else can you do though? What can give you that "edge"? How can you optimize for this "ranking signal".

This answer to this question isn't an answer, but another question:

Why would you want to try?

RankBrain might be beneficial to some unique use cases. However, for most sites, the time and energy you would use to try to rank for a query that is unknown to Google (meaning no one is using it) would be far better spent on other things.

Because not only are you trying to optimize for a query that few people are using, but it is constantly changing.

RankBrain results are designed to change and bring back better results. So optimizing for it would be like trying to hit a

moving target – all the time.

The best advice? Follow Illyes’s advice.

Write good content.

Make sure it sounds natural.

Move on.

http://tracking.feedpress.it/link/13962/7721284