ANALYSIS OF RANKBRAIN ALGORITHM USING MACHINE LEARNING

Shantam 1, Harish Kumar 2 and Dr. Rajesh Kumar Tiwary 3

1Department of Mechanical Engineering, RVSCET Jamshedpur, India  
2Department of Computer Science and Engineering, RVSCET Jamshedpur, India  
3Head of the Department of Computer Science and Engineering, RVSCET Jamshedpur, India

ABSTRACT:

RankBrain is Google’s name for a machine-learning artificial intelligence system that’s used to help process its search results. Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data. The methods Google already uses to refine queries generally all flow back to some human being somewhere doing work, either having created stemming lists or synonym lists or making database connections between things. Sure, there’s some automation involved. But largely, it depends on human work. True artificial intelligence, or AI for short, is where a computer can be as smart as a human being, at least in the sense of acquiring knowledge both from being taught and from building on what it knows and making new connections. In terms of RankBrain, it seems to us ‘RankBrain and AI’ are fairly synonymous. You may hear them both used interchangeably, or you may hear machine learning used to describe the type of artificial intelligence approach being employed. RankBrain is nothing, but a part of Google’s overall search “algorithm,” a computer program that’s used to sort through the billions of pages it knows about and find the ones deemed most relevant for particular queries. RankBrain is one of the “hundreds” of signals that go into an algorithm that determines what results appear on a Google search page and where they are ranked. In the few months it has been deployed, RankBrain has become the third-most important signal contributing to the result of a search query. Thus, it is helping to increase the efficiency of Hummingbird algorithm using machine learning.

Keywords: Google, RankBrain, Machine Learning, Artificial Intelligence(AI), Search, Algorithm, Page, Rank, Signal, Hummingbird Algorithm, ranking factor.

[1] INTRODUCTION TO RANKBRAIN

RankBrain is a machine-learning artificial intelligence system that helps Google process some of its search results, in particular rare or one-of-a-kind queries. It was launched in early 2015 and is used globally by Google. RankBrain is primarily focused on helping refine queries that Google processes, but the company says that the system is also used to help rank web pages. In fact, Google says RankBrain is its third most important ranking factor.
[2] RANKBRAIN & MACHINE LEARNING USING AI

The Artificial Intelligence That’s Now Processing Google Search Results. A machine-learning AI is handling a large fraction of Google's search queries. Machine learning is where a computer teaches itself how to do something, rather than being taught by humans or following detailed programming. Gary Illyes said that Google uses it mostly for “coming up with new signals and signal aggregations.” So they may look at two or more different existing non-machine-learning signals and see if adding machine learning to the aggregation of them can help improve search rankings and quality.

[3] ARTIFICIAL INTELLIGENCE

- True artificial intelligence, or AI for short, is where a computer can be as smart as a human being, at least in the sense of acquiring knowledge both from being taught and from building on what it knows and making new connections.
- True AI exists only in science fiction novels, of course. In practice, AI is used to refer to computer systems that are designed to learn and make connections.
- How’s AI different from machine learning? In terms of RankBrain, it seems to us they’re fairly synonymous.

So RankBrain is the new way Google ranks search results?
No, RankBrain is part of Google’s overall search “algorithm,” a computer program that’s used to sort through the billions of pages it knows about and find the ones deemed most relevant for particular queries.

[4] GOOGLE’S SEARCH ALGORITHM

- It’s called Hummingbird, as we reported in the past. For years, the overall algorithm didn’t have a formal name. But in the middle of 2013, Google overhauled that algorithm and gave it a name, Hummingbird.
- It’s not really a complete replacement of the Google search algorithm, also known as Hummingbird, however. The algorithm is the system that processes what people search for and combs through billions of pages to rank the ones believed to be best first.

  Hummingbird is the overall search algorithm, just like a car has an overall engine in it. The engine itself may be made up of various parts, such as an oil filter, a fuel pump, a radiator and so on. In the same way, Hummingbird encompasses various parts, with RankBrain being one of the newest.

- Hummingbird also contains other parts with names familiar to those in the SEO space:
  1. Panda, Penguin and Payday designed to fight spam
  2. Pigeon designed to improve local results
  3. Top Heavy designed to demote ad-heavy pages
  4. Mobile Friendly designed to reward mobile-friendly pages
  5. Pirate designed to fight copyright infringement

PageRank is part of the overall Hummingbird algorithm that covers a specific way of giving pages credit based on the links from other pages pointing at them. PageRank is special because it’s the first name that Google ever gave to one of the parts of its ranking algorithm, way back at the time the search engine began, in 1998.
[5] SIGNALS GOOGLE USES FOR RANKING

Signals are things Google uses to help determine how to rank webpages: -
For example, it will read the words on a webpage, so words are a signal. If some words are in bold, that might be another signal that’s noted. The calculations used as part of PageRank give a page a PageRank score that’s used as a signal. If a page is noted as being mobile-friendly, that’s another signal that’s registered. (very important) – depends on websites’ quality and every website signal calculation vary which gives them rank.

• All these signals get processed by various parts within the Hummingbird algorithm to figure out which pages Google shows in response to various searches.
• The moment You understand Google’s algorithm, google changes it.

And RankBrain is the third-most important signal. From out of nowhere, this new system has become what Google says is the third-most important factor for ranking webpages.

The first- and second-most important signals: -
1. My personal guess is that links remain the most important signal, the way that Google counts up those links in the form of votes.
2. As for the second-most important signal, I’d guess that would be “words,” where words would encompass everything from the words on the page to how Google’s interpreting the words people enter into the search box outside of RankBrain analysis.

[6] RANKBRAIN HELPS TO REFINE QUERIES

• Shoe and Shoes
• Google became smarter to do stemming: - Running is a variation of run.
• Similarly typing sneakers – Google would get – running shoes.

Figure: 1. Query search – understanding the intent.
ANALYSIS OF RANKBRAIN ALGORITHM USING MACHINE LEARNING

Figure: 2. Query search – understanding the intent & generating the content.

Figure: 3. Query search – understanding the intent & generating the content.
Therefore, similar intent content was generated when a particular intent was defined by the user while doing a particular search. That’s why Google brought RankBrain, already 3% of the shares it has lost in the US market, and for saving Google’s future – search algorithm is getting revolutionized.

[7] SEARCH UPDATE IMPACT ON SEO & CONTENT STRATEGIES

Panda – (Low Quality Content or Spammy Content)

- The Panda update was first launched in February 2011, though it has been updated several times since then. This update is designed to target sites with low-quality content and prevent them from ranking well in search engine results pages.
- Sites that have pages of spammy content, too many ads or excessive duplicate content, for example, often experience Panda penalties.
- It was recently announced that Panda was added to Google’s core ranking algorithm, which has caused considerable buzz in the industry.

Penguin – (Targeting spammy links – Do follow or Quality backlinks or No follow links)

- The Penguin update was first released about a year after the Panda update, in April 2012. The two are often grouped together when discussing Google’s big push to raise the quality of content that appears in search engine results.
• This update focused largely on targeting spammy links. Google looks at backlinks as a 
signal of a website’s authority and reputation, taking a site or page’s backlink profile into 
consideration when determining rankings.
• Back when its core algorithm was less sophisticated, people figured out that they 
could **effectively game search engine** rankings simply by obtaining significant numbers 
of (often spammy and irrelevant) backlinks.
• **Penguin combatted this manipulative technique** by targeting pages that depended upon 
poor-quality links, such as link farms, to artificially raise their rankings. Websites with 
spammy backlink profiles have been forced to remove or disavow bad links in order to 
avoid ranking penalties.
• Quality links still have something of value to offer websites, although Google emphasizes 
that sites should focus on developing a **quality backlink profile** organically. This means 
creating informative pieces that people will want to source with a backlink.
• To attract attention to your piece, you can leverage the search, social and content trifecta. 
By creating high-quality pieces and then distributing them on social media, you start to 
attract attention to your work.

**Hummingbird – (LSI or semantic search capabilities)**
• The **Hummingbird update** followed in the summer of 2013. This update was designed to 
improve Google’s semantic search capabilities. It was becoming increasingly common 
for people to use Google in a conversational way, to type their queries as though they 
were asking a friend.
• This update was designed to help Google respond by understanding **intent and context**.
• With this update, the development of content had to shift slightly again. With the 
emphasis on intent, Google was not simply playing a matching game where they connect 
the keywords in the query with the keywords in the content.
• Content needed now to go beyond just the keyword. It needed to **demonstrate an 
understanding of what users are interested in and what they would like to learn**.
• While keywords still are an important part of communicating with the search engine 
about the **topic of the content**, the way they were used shifted. **Long-tail keywords 
became more important**, and intent became crucial.
• Content developers needed to direct their focus toward understanding why customers 
might be typing particular words into the search engine and producing content that 
addressed their needs.

**Mobile Update**
• The year 2015 saw several major updates that impacted content development. The 
first, Google’s **mobile-friendly update**, occurred in April. This update was unique because 
Google actually **warned** website users in advance that it was coming.
• With this update, Google recognized that mobile was beginning to dominate much of 
search and online customer behaviour — in fact, just a couple months after the mobile-
friendly update was announced, Google noted that mobile searches had **officially 
surpassed desktop**. The mobile-friendly update forced sites to become mobile-friendly or 
risk losing visibility to sites that were.
• With this update, Google wanted sites to take into account what mobile users wanted to do online and how these needs were being addressed.
• This meant that SEOs and content marketers had to start considering design factors such as:
  1. Responsive design or a mobile page.
  2. Having site navigation front and centre and easy for customers to use with their fingers.
  3. Avoiding frustrations caused by issues such as buttons too close together.
  4. Having all forms as efficient and as easy as possible to fill out on a smartphone screen.
• This mobile update also brought to the forefront the importance of brands optimizing for mobile, even going beyond what was required by Google to avoid a penalty.
• For example, customers on mobile are often very action-oriented. They want to be able to call you or find your address. They want to view the information on your screen easily, without excessive scrolling. While long-form content is commonly read on mobile devices, making it easier for people to get back to the top is very beneficial.
• Mobile users also tend to be very local-oriented. Content developed for mobile devices should take local SEO into account to maximize the mobile opportunities that present themselves.

Quality Update
• Not long after the mobile update went live, people began reporting evidence of another Google update, which has since been nicknamed the Quality Update.

[8] ANALYSIS OF RANKBRAIN USING MACHINE LEARNING
• RankBrain was announced in the fall of 2015, and it was also a unique change to the Google algorithm. With this update, the search engine ventured into the world of AI (artificial intelligence) and machine learning.
• This system was designed to learn and predict user behaviours, which helps Google interpret and respond to the hundreds of millions of completely unique, never-before-seen queries that it encounters each day.
• It is also assumed that RankBrain helps Google to interpret content and intent in some way. Although Google has given little information about how their new AI works, they have said that it has become the third most important ranking signal. For site owners, this has placed an even greater emphasis on creating content that matches the user intent.
• Since RankBrain has gone live, some marketers have spoken about the importance of making sure that the technical side of SEO, such as schema markup, is all up to date. It is likely that as search engines become more dependent upon AI, these little details will become significant.
[9] RANKBRAIN CHANGES ENTITY SEARCH

- RankBrain is one of the “hundreds” of signals that go into an algorithm that determines what results appear on a Google search page and where they are ranked, but this is the third most important signal.
- RankBrain is more likely a “query processor” than a true “ranking factor.” It is currently unclear how exactly RankBrain functions as a ranking signal, since those are typically tied to content in some way.
- The launch of the Hummingbird algorithm was a radical change.
- Semantic search seeks to improve search accuracy by understanding searcher intent and the contextual meaning of terms as they appear in the searchable dataspace, whether on the Web or within a closed system, to generate more relevant results. Semantic search systems consider various points including context of search, location, intent, variation of words, synonyms, generalized and specialized queries, concept matching and natural language queries to provide relevant search results. Major web search engines like Google and Bing incorporate some elements of semantic search.
- The Move from “Strings” to “Things”

[10] RANKBRAIN FUNCTIONING PRINCIPLES

RankBrain actually changes the rankings of search results rather for “a lot” of them. That fits in with what we’ve understood about RankBrain: it seems largely used as a query refinement tool. Imagine that RankBrain sees a search for “best flower shop in Los Angeles.” It might understand that this is similar to another search that’s perhaps more popular, such as “best LA flower shops.” If so, it might then simply translate the first search behind the scenes into the second one. It would do that because for a more popular search, Google has much more user data that helps it feel more confident about the quality of the results. In the end, RankBrain did change the ranking of those results. But it did that simply because it triggered a different search, not because it used some special ranking factor to influence which exact listing appeared in what order. For the SEO and search marketers worried about what they should do now that RankBrain has ramped up, the answer remains the same: nothing, but focus on great content. So focus on that, which has always been the case with SEO, and you’re on the right track. Google uses RankBrain for every search, impacts rankings of “lots” of them. Last year, RankBrain was used for less than 15% of queries. Now Google's confidence has increased enough that it's used all the time.

Here are Google’s top 3 search ranking factors, and Google's Andrey Lipattsev reveals:

1) Links
2) Content
3) **and RankBrain – 3rd most important signal**
   These are the top three ranking signals in Google's search algorithm.

4) **[11] CONCLUSION - THE FOUR PILLARS OF THE FUTURE OF SEO**

SEO has come a long way from being all about on-page optimization, building backlinks and creating “relevant” content. When I read popular search engine blogs, I notice a definite trend: SEO is moving toward a more inclusive strategy that goes beyond **new ways of link building or content marketing**.

1. **RankBrain**
   According to Gary Illyes and Andrey Lipattsev of Google, RankBrain does not act on your backlink profile, content quality or click-through rate. It only helps the algorithm interpret queries better and match them with relevant page content.

2. **The Knowledge Graph & rich answers**
   The Knowledge Graph is a knowledge base used by Google to enhance its search engine’s search results with semantic-search information gathered from a wide variety of sources.

![Rich Answers Continuing to Grow](image)

**Figure: 5. Graph representing for the use of Knowledge Paragraphs**

3. **Accelerated Mobile Pages (AMP)**

Shantam, Harish Kumar and Dr. Rajesh Kumar Tiwary
In February 2016, Google integrated results from its Accelerated Mobile Pages project into its search results in the form of a “Top Stories” carousel in mobile results. Six months later, Google started displaying links to AMP pages in the main organic search results.

Today, Google has 150 million indexed AMP documents in its index, and, encouraged by mainstream adoption outside the publishing industry (including eBay and Bing), has just announced that users searching from mobile devices will be directed to the relevant AMP pages even if an equivalent app page exists.

4. Real-time, integrated penalty filters

Penguin is now a real-time signal processed within Google’s search algorithm — data on your pages is refreshed every time Google re-crawls and re-indexes them.

Figure: 6. Periodic Table for the SEO success
REFERENCES