

INK Whitepaper: Content Rank Theory

Author: Alexander De Ridder, Chief Technical Officer of INK
For publication, August 1st, 2019

Introduction: An SEO Extinction Event

I began coding in 1994 on my 386 PC, but it was not until after my Computer Science studies at the University of Ghent when I began working for a marketing firm in the same city that I was first bitten by the internet marketing bug.

I was coding webpage backends using CGIs¹ and later in PHP 2.0. AJAX² had just become a *thing* because it was new on the local computer store's bookshelves.

36 hours later, I had built a multi-player board game in the browser using AJAX. Since I was now surely the foremost expert in AJAX in tiny Belgium, I landed my first marketing job that same week, book-in-hand to the interview and arguing how I could help innovate in marketing by creating dynamic campaigns, personalization, and tracking using the latest technologies.

By 2006, I had migrated to the United States, and after creating a patented call tracking solution for Google Adwords, I landed a job as lead r&d from 2009 to 2016, building computer vision solutions using classic machine learning techniques with the goal of bringing internet marketing technology and techniques to retail stores. Think of real-life cookies. Attribution, Personalization, Yum!

Neural Networks!

When Google first was able to detect cats³ in YouTube videos in 2012, and Facebook was subsequently able to recognize faces with 97.35% using DeepFace⁴ in 2014, I understood the world of internet marketing was about to change forever. Neural networks would change everything.

In 2016, I developed a Rank Candidate Theory and co-founded Edgy Labs with the purpose of validating it. In essence, I understood that global rank factors were about to be killed off as Google would begin applying their Neural Network advances to Search.

Ranking Factors

Content creators have loved tools based on universal Ranking Factors for many years. Because they worked. Yoast, for example, “runs on more than 9 million sites, and optimizes 11.4% of the top 1 million sites in the world.”⁵

Examples of Ranking Factors for search are “you should write more than 300 words” and “your keyword density should be below 3%.”

Those rules are going extinct, because of advances in AI. We have all seen it happen - Google is getting better at personalizing search for each individual search query.

Lexical Optimization

In response to this extinction event, we’ve seen a number of companies release more advanced tools for optimizing content. Their main innovation is what one would call “Custom Ranking Factors.”

The basic premise of a custom ranking factor tool is to analyze the top 10 or 20 of Google, and derive insights from this content based on the keyphrase you are targeting.

These tools universally have two things in common:

1. You have to pay up, big time. Prices range between \$100 and \$3,000 per month.
2. All the tools we’ve examined are built as a side-project of an SEO suite (me-too) or are built with marketers and SEOs as their primary audience. There is no solution that’s specifically for writers.

Furthermore, a subset of the new breed of Custom Ranking Factor tools for content is justifying their high cost with claims of “Artificial Intelligence”, “Machine Learning”, and/or “Data Science.”

Consumers are dazzled with terms like TF-IDF and are told that if they include this or that keyphrase to their content, they will improve their chances to rank for a given piece of content.

Tools like this are essentially performing Lexical optimization when Google has moved and evolved away from such old metrics on to Neural Network powered Semantic optimization.

Advances in search algorithm AI have led to a second SEO extinction event: lexical optimization. This is simply a more advanced form of keyword stuffing.

It’s not hard to prove that Google has moved beyond TF-IDF and other forms of Lexical analysis as a form of information retrieval. You’ve noticed it yourself - Google may rank your content for words that you don’t even include in your text - it’s gotten smart enough to know what you are talking about.

Left Behind

As with all progress, some will feel left behind by rapid and sweeping change.

All our market research and interviews with writers point to a single conclusion: optimizing content for SEO has become an increasingly difficult and frustrating process.

While everyone was concerned with how these changes affected business revenue and built tools to help marketers capitalize on that, the writers who actually made the internet what it is today have been left behind.

In 2016, we've felt the early foreshock tremors of this emerging event and began developing and experimentally validating our Rank Candidate Theory.

Subsequently, we've built a semantic optimization AI that we're extremely proud of and decided to put all that intelligence in the formula that made Yoast so popular: easy to use, and free.

Our final decision involved taking the user experience for writers to the next level. Could we create a user interface that made optimizing content for search so easy that we could remove the writer's frustration? Could we build a tool so popular that it would become to web content what Adobe Photoshop became to photographic content in years past?

That vision gave birth to our product, INK. And in this white paper, we'll go into the theory in more depth. While this whitepaper is not intended to be a scientific paper, and while we are writing it in a way to protect our intellectual property, I am confident that you will enjoy reading through the methods, findings, analysis, and conclusions.

The Rank Candidate Theory

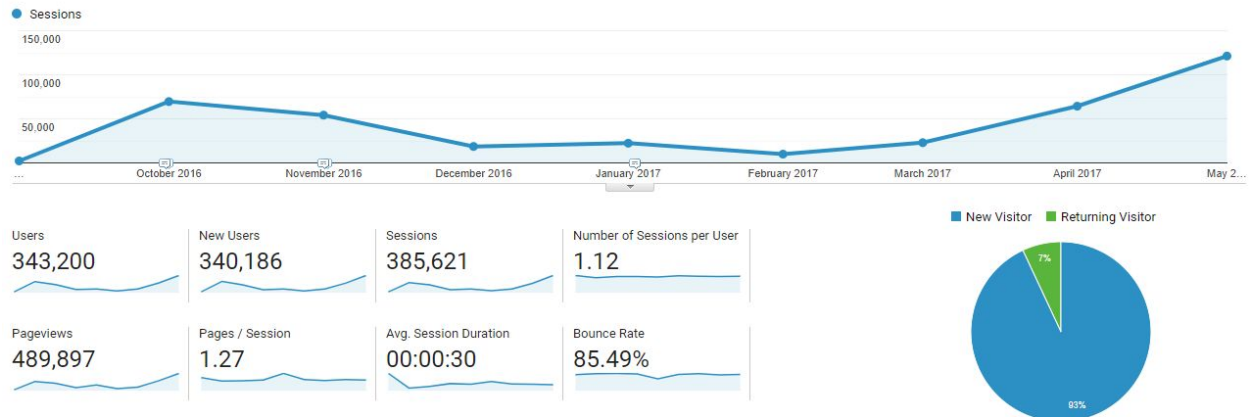
The Theory, 2016 Edition

Google crawls your page, scrapes your content, and determines from a deep understanding of the content if you meet the user's intent. If you do, they consider your page a *rank candidate*. They will send some traffic to your page to test user satisfaction. If users like your page, your rankings will go up. Content Relevance and User Experience will, therefore, become increasingly important ranking factors.

Testing Method

We intended to validate our Rank Candidate theory by buying a brand new domain (edgylabs.com) - meaning zero domain authority - in September of 2016, optimizing it for technical SEO, and publishing content.

The key to the strategy was to not influence our key experimental content with link building strategies. By doing this, we were able to test how Google responded to different content optimization strategies, using content relevance as a new "page authority" signal, capable of outperforming the link-based rank approaches of the past.



By the power of content optimization - organically growing a brand new domain to 100,000+ monthly sessions from a cold start.

Observations & Interpretation

1. We noticed that content - even if it lacks backlinks AND even if it comes from a site with little domain authority AND and provided it has a decent technical optimization - will be tested by Google with some traffic ONLY if it meets searcher intent because it has achieved Semantic / Topical Completeness. In practical terms, many pieces of content fail at becoming a Rank Candidate.
2. Typical behavior for the Google Search Engine would be to send a trickle of traffic to your page, followed by either no subsequent traffic or a lot of traffic. This period of time is where we believe that Google is testing your page for user satisfaction. We believe that the more engaging the content, the better the final outcome. This is what we understand as RankBrain activity. ⁶ In other words, AI makes the decision if you are worthy of a chance, but RankBrain uses actual users to contribute the data which helps further the decision of where you will end up ranking.

Conclusions

You should first worry about how to become a rank candidate, and afterward, you need to ensure that your content achieves good engagement metrics with your audience.

(In case you are wondering what we've done with our site... the site is still running strong, but the experimental focus has shifted. After all, we couldn't maintain our initial experimental conditions because of the site's growing authority. Today we are testing against other ranking factors. We shut down edgylabs.com and made a new PWA, edgy.app, where we are testing how Google responds to content migrations, testing how technical SEO and speed optimization.)

On Ranking Factor-Based Approaches

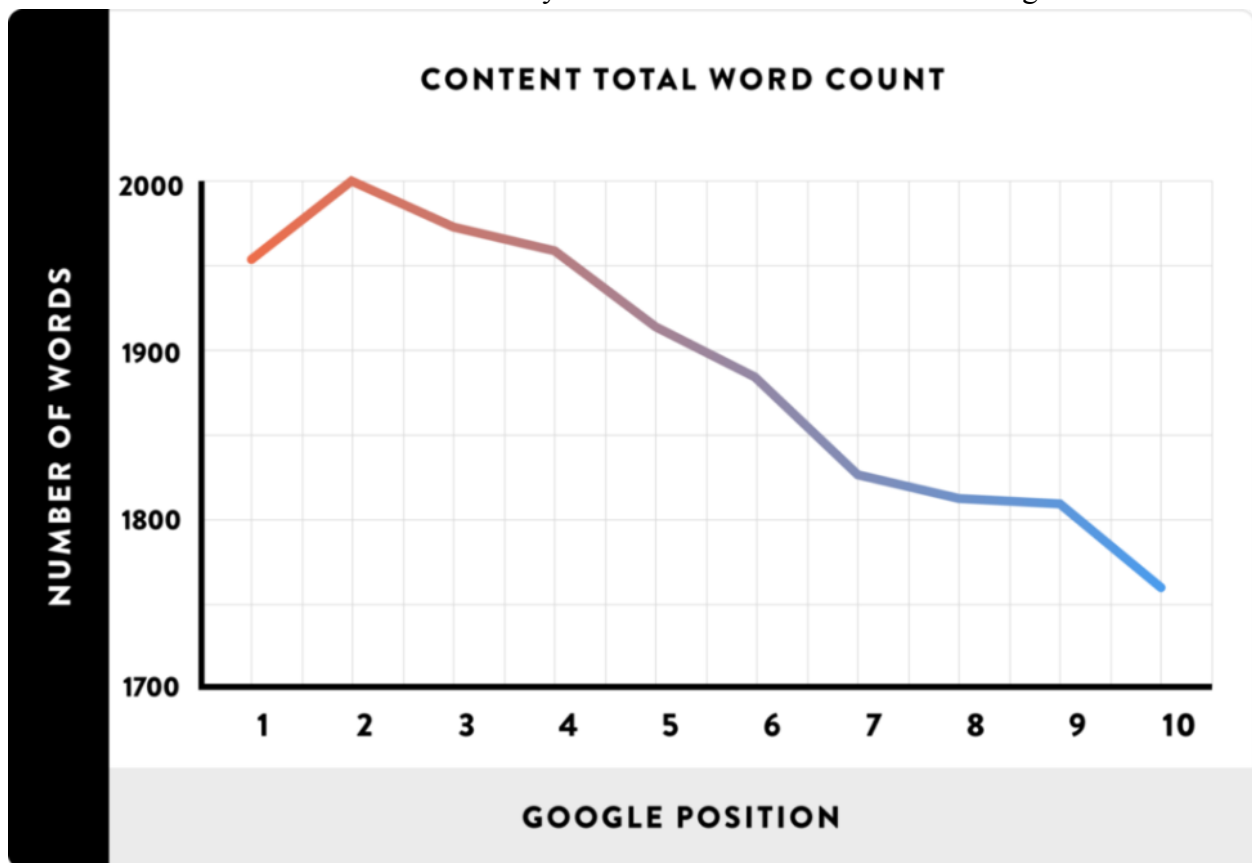
There's no doubt that technology has attempted to cut the learning curve for updates in SEO, but it relies only on limited information and rules. For example, tools such as Yoast ⁷, Hemingway ⁸, Readable.io ⁹, and more are using rule-based recommendations.

In other words, they don't take into consideration that every search query indicates a unique searcher intent. One size doesn't fit all.

In 2019, every search query is unique, and you need a solution that keeps up with Google and allows for more personalization of content. You can't rely on your site's topical authority, nor on general SEO best practice rules.

One industry leader who analyzes and explains these details further is Brian Dean of Backlinko. Each year, in collaboration with BuzzSumo ¹⁰, he publishes an amazing and large scale blog content data study.¹¹

Looking at one graph from his study below, a content writer might conclude that to be safe, you should write about 2000 words in each of your articles. That would be the wrong conclusion.



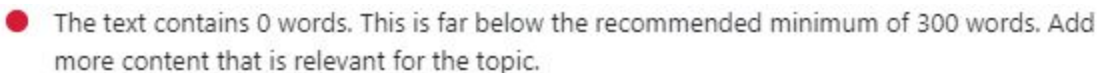
Correlation between word count and top 20 page positions in Google ¹²

The above graph shows that for an average high-ranking article, Google prefers longer content.

But what about your particular topic? You only need to outperform your competitors. If your competitors score the top positions in Google with just 500 words, then you don't need 2,000 words to beat them.

A Real-Life Example

In the example below, the Yoast tool recommends you write at least 300 words. This is an example of a Ranking Factor based approach. They say 'minimum', so they are not incorrect. But how is this really helpful to a content writer trying to figure out how much to write?



● The text contains 0 words. This is far below the recommended minimum of 300 words. Add more content that is relevant for the topic.

User Screenshot of a Yoast recommendation.

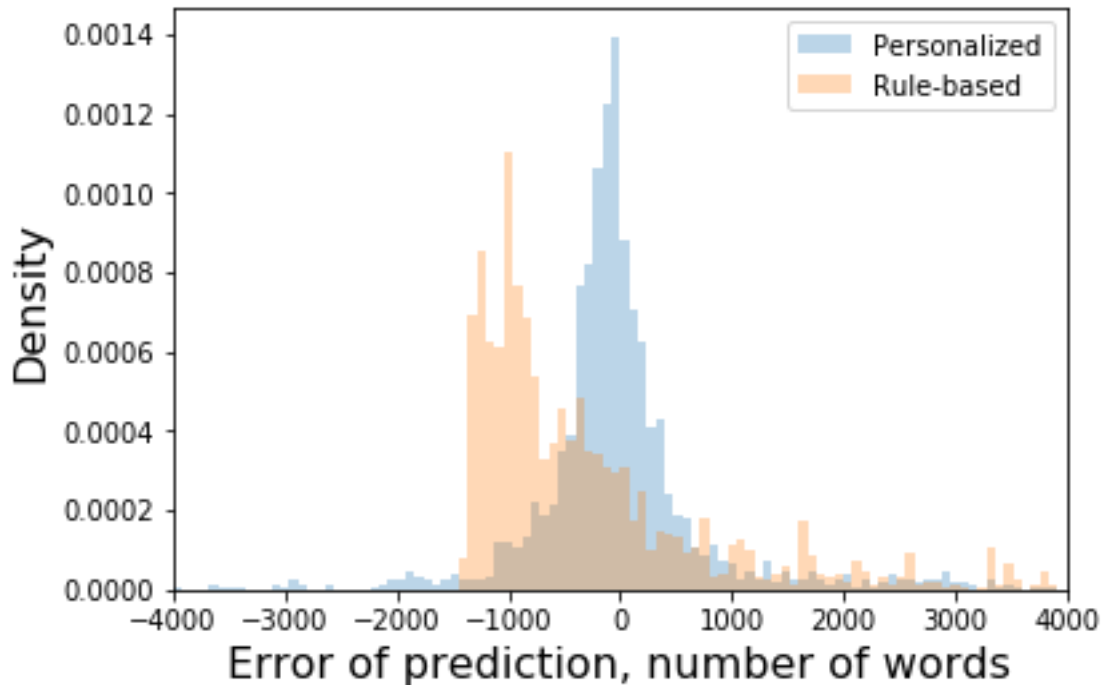
As pointed out in the introduction, Global Ranking factors are experiencing an extinction-level event. But to what extent have these Global Ranking Factors become less relevant? We've done a study on how Global Ranking Factors compare with Customized Ranking Factors.

Study: Customized Search Heuristics Outperform Global Factors

We achieved a better prediction of search performance with a simple heuristic which suggests the weighted average of word counts of pages with positions from 2 to 10. For a fair comparison, we can then take an average of these suggestions and use it as a rule-based recommendation (to make the average of these suggestions coincide). We can then compare both recommendations with a ranked 1st-page word count.

In the graph below, you can see that the heuristic recommendation is much better centered around 0 (which means no error.) In fact, on average, it's over two times better in both directions (261 extra words against 568 and 269 missing words against 576 than in 1st page).

This way you can be more confident that your article is not penalized for being too short or too long. In practical terms, you can stop worrying about hitting 2,000 words and artificially inflating your content with perhaps less relevant information. Instead, focus on providing the right amount of text based on the searcher's intent, which makes the search engines and the search audience happy.



Comparison of personalized and rule-based recommendations for content length

We repeated this study for many other ranking factors and came to the same conclusion over and over. You could easily replicate this heuristic and publish your own findings on those. (If we've inspired you, feel free to leave a shout-out in your article!)

On Custom Ranking Factors based on Lexical Optimization

As mentioned in the introduction, a new generation of tools has sprung up that recognize how changes in Google make rule-based tools outdated. They offer Custom Ranking Factors, and this is a very attractive proposition.

We have found that many of these tools consider Lexical Optimization as AI, and market their tools as "AI-powered". Examples of Lexical Optimization include Natural Language Parsing and Word-Frequency tables.

Lexical Optimization Approach #1

It works like this, for example: take the top 10 search results in Google. Extract the content. Parse the text and sort words that have high frequencies, because those must be important. Then check the user's content. If they didn't use those words, suggest they include those words with the same frequency.

There are a few fundamental flaws with this method. First of all, Lexical Optimization based tools cannot account for synonyms and different word forms, which extends to more complex scenarios such as words with similar meanings.

That is mostly why such approaches like TF-IDF ¹³ and its variations were abandoned as primary tools for use in Natural Language Understanding ¹⁴ over ten years ago in favor of more comprehensive techniques such as word embeddings. ¹⁵

Google's John Mueller had this to say about TF-IDF: ¹⁶

- “This is a fairly old metric and things have evolved quite a bit over the years”
- “My general recommendation here is not to focus on these kinds of artificial metrics”

More importantly for SEO, such approaches cannot understand the depth and breadth of the topical coverage, but that is exactly what Google is looking for.

Lexical Optimization Approach #2

A related Lexical optimization approach will suggest that you include long-tail keyphrases in your text that content that is ranking highly in Google for your main keyword is ‘also ranking for’.

The obvious problem is one of the writer's original tone and voice. If you can't contribute anything unique or of value, and you will just copy other writer's ideas, then why would Google find your content interesting? The reasons those other articles rank for other keyphrases is because of how they have developed their content from different angles.

Suggesting to add such keyphrases to your text while your content is not semantically optimized for it can backfire on your SEO, come across as artificial to your users, and is a losing tactic long term.

It's as John Mueller said in the same Webmaster Central:

“Instead, I would strongly recommend focusing on your website and its users and making sure that what you're providing is something that Google will in the long term still recognize and continue to use as something valuable.”

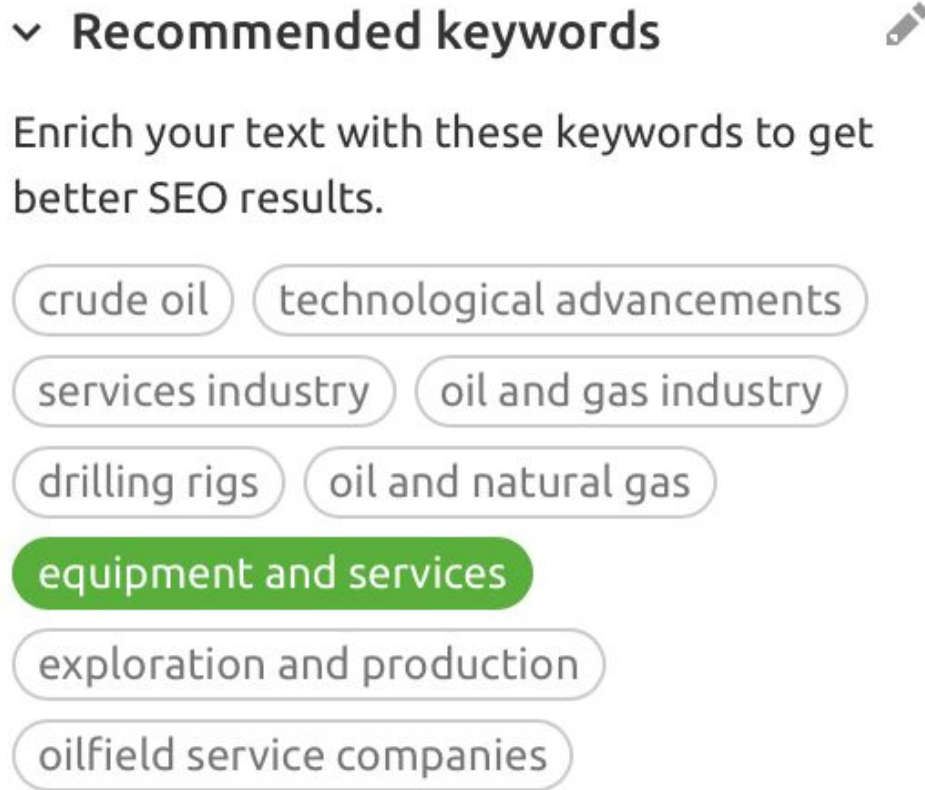
If you think about it, the promise of this approach is not to help you rank for your primary focus keyphrase. Instead, it promises that you might instead pick up some extra traffic from related keywords.

We believe this idea has some merit, but requires a semantic approach to recommend words which are truly semantically relevant. Just because these phrases appear in other top 10 Google Results does not mean that they are relevant for your article as well.

A Real-Life Example

SEMrush makes a content optimization tool available as part of their SEO Suite. We love SEMrush, particularly their Site Audit feature, and have been happy customers for years.

Recently, we had a beta tester of INK send us the screenshot below. They asked us why INK is not recommending keywords to include in the copy. Let's take a closer look:



User Screenshot of SEMrush' Content Tool

Let's say that you've got an article about "drilling rigs." What might the user expect to find for such a search? What would be a top 1 article look like for that keyphrase?

Now do the same exercise for "services industry." What does such top content look like?

Finally, mentally compare the two articles. Do they look a lot different in your mind? We think they are radically different pieces of content.

Including a keyword and hoping you will magically fool Google into ranking for that additional keyphrase? If only it was that easy.

Notice how Google defines Keyword Stuffing:

"Keyword stuffing" refers to the practice of loading a webpage with keywords or numbers in an attempt to manipulate a site's ranking in Google search results. ¹⁷

The new methods are definitely more sophisticated, but it's just a modern version of keyword stuffing. Keyword stuffing worked for a while until the ban hammer came down on this practice. Google is becoming more sophisticated.

E-A-T and the Cost of Being Ordinary

The marketing slogans are amazing. "We'll give you an outline that will show you exactly how to rank for a given keyphrase." Who wouldn't want such a wizard at your fingertips?

Some tools promise the users "outlines", which could work something like this: Take the top 10 results in Google for your target keyphrase. Extract the content. Extract all paragraphs which contain your target keyphrase. Compile an outline of competitors' paragraphs.

A very important issue for such tools is idea plagiarism and originality of your content - the most they can do is to suggest you make your content similar to already existing in the top on a very primitive word level.

This won't make your content original enough to be ranked high because search engines try not only to rank and sort content by relevance, but also provide a user with as diverse content as possible (otherwise you would see 10 copies of the same articles on the 1st page).

Therefore, by following recommendations that encourage you to duplicate other highly-ranking pages, you might even decrease your own page's value for Google.

In 2018, we learned from the Search Quality Evaluation Guidelines that Google values Expertise, Authoritativeness, Trustworthiness. ¹⁸

- Are you authoritative if you imitate your competitor's content ideas?
- Are you trustworthy if your content is at increased risk of plagiarism?
- Are you showing your expertise if you stick to an outline which makes you ordinary?

While competitive research tools are useful, the average writer is not the average SEO expert. Such tools are often geared toward the SEO expert, who then hands over the outline to the content writer.

The content writer follows the instructions without the proper context, and risks their, or their domains' reputation as a source with a low E-A-T score, and just like lead poisoning, progressively lowering their ability to rank in the future.

INK and the Rank Candidate Theory

Now that we've completed a survey of the status quo of tools landscape, it's time to explain how we've approached the problem of content optimization at INK.

True to our Rank Candidate Theory, INK is built to achieve these two goals:

- Optimize chances to qualify content as a rank candidate with Google
- Optimize language to improve user satisfaction with the content
 - By optimizing language to resonate with the precisely-targeted audience

We'll take you through each of those two objectives, but a few definitions are in order:

- Search Intent: What the searcher is expecting to find in response to a search query.
- Relevance: Content is deemed relevant if it meets user expectation from an informational or topical point of view.
- Semantic: Related to the meaning or information in language.
- Semantic / Topical Breadth: The broad spectrum of ideas or topics contained in the content.
- Semantically / Topical Complete: The content is relevant for the entire semantic space, meaning that the content achieved minimum topical completeness.
- Semantic/Topical Depth: Do you cover each information in the desired amount of depth?

How INK Helps You Qualify as a Rank Candidate

Search engines get more users if they serve more relevant results, and are therefore financially incentivized to reward more relevant content.

You can achieve content relevance by optimizing for topical completeness with sufficient depth.

How INK Measures Topical Completeness

Achieving this was no small feat. Our patent-pending approach is the result of nearly two years of r&d. Essentially, you first have to be able to map text to meaning. That is typically achieved

via a technique called word embeddings. Our technique is unique, but the end result is a concept map.

We can generate such maps for any piece of content. Those maps are the coolest thing ever because they are so powerful: we can compare maps with each other, calculate the similarity between them, find the differences, and so forth.

One of the biggest development challenges here is “discrimination.” Concepts are everywhere, but the true meaning is derived from the relationship between concepts. Here’s an example:

- Apple is a fruit.
- Apple is a technology company.

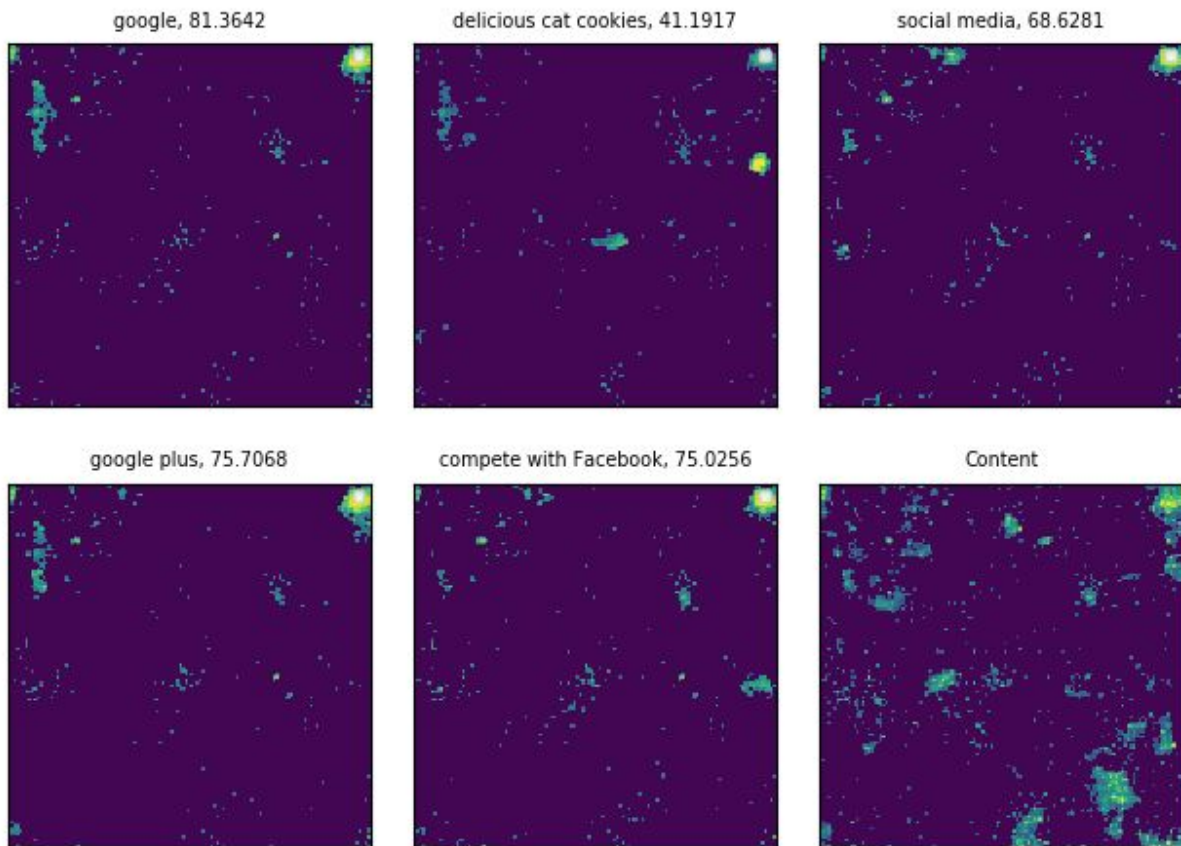
So if we played Pictionary, I might draw a fruit, and something like a computer, and you would be able to understand what I’m talking about.

At the same time, there are so many possible concepts in any given piece of text. The challenge is to select the most important concepts without distorting the combination in a way that meaning is not lost.

In the end, we’ve achieved the ability to retain the most relevant concepts out of a quarter of a billion possible concept combinations.

What’s super cool about these concept maps, is that they are kind of like a brain scan. You will see patterns emerge according to different thoughts (keyphrases) in relation to a piece of content.

Below you can see such concept maps generated for an article [Why Did Google Plus Fail?](#)¹⁹ If you look closely, you’ll see that there are more common clusters of features between the content and relevant keyphrases (such as "Google" and "compete with Facebook") than with non-relevant ("delicious cat cookies") or too common keyphrases (such as "social media").



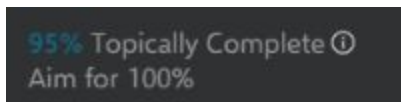
You're looking at the first pictures of what you could call a Google Brain Scan.

Instead of using obsolete word-frequency based approaches - which even fail to recognize synonyms and different word forms - we developed a text analyzer that builds a high dimensional concept map. When you visually map-related concepts closer together, our semantic concept models look a lot like brain activity.

We get to look inside a search engine's brains. This opens up endless possibilities that we're just starting to tap into, all of which will be built into the capability of INK to help you write more effectively.

That's because these concept maps are trained on what Google deems relevant. This is our big achievement: INK has managed to create a concept map of how Google deems content as relevant.

We can now score your content by comparing it to what Google is expecting from a relevance point of view. That's how our Topical Completeness Score works:



Screenshot of INK showing your Topical Completeness

If you have been let down by tools that offer silly word recommendations, you will be delighted at how intelligent INK really is. That's because INK has developed - from scratch - a true search-engine-grade semantic Artificial Intelligence (AI) that understands the meaning and concepts behind the content.²⁰

It's the closest you can get to understanding Google Rankbrain from a semantic relevance point of view without actually being at Google. Our tool is agnostic to word forms or exact synonyms because it understands the meaning behind words in a context.

We believe that words change, but ideas are consistent!

Our beta users loved the scoring. Some have asked when we will have additional tools available to help achieve a higher score. As of today, we've already got seven months of r&d into a solution (hint: it won't be anything like what exists on the market today), but we still have to finish fine-tuning our new AI's, evaluate the effectiveness, and test it rigorously.

How We've Evaluated our Content Mapping Technology

We evaluated the quality of this approach by analyzing hundreds of thousands of search engine results, and to no surprise, we found that high ranking content has important ideas (thus similar concept maps) in common, while low ranking or irrelevant content is missing these concepts. Namely, by tuning parameters, we achieved over 50% average difference between relevant and non-relevant pages for a random keyphrase.

We ran thousands of pipelines on distributed spark clusters over the course of the last year to discover the best ways to fine-tune our INK's AI-engine. We want to acknowledge the authors of open-source software we used. This includes but is not limited to Spotify's²¹ Luigi²², Apache Software Foundation²³, UC Berkeley²⁴ AMP Lab²⁵ and DataBricks²⁶ for Apache Spark²⁷, Google Brain²⁸ for TensorFlow²⁹ and many others for open-source python libraries.

Search-Engine Grade

For this type of solution, calculations can be so intense, that code architecture and approach are critical. From the very beginning, one of the guiding r&d pipeline objectives was to achieve search-engine-grade performance.

Our approach was to think about how search engines like Google might determine the semantic relevance of content at scale.

We believe that INK has achieved this. In practical terms, our AI can map more than a quarter of a billion different concept combinations in milliseconds for any type of content.

With INK, you literally have billion-dollar search-engine-like technology guiding your content optimization recommendations.

How INK Helps You Achieve User Satisfaction

The second step to ranking high in Google revolves around user delight with your content.

Improving user engagement is a core objective of content optimization.

We believe the three pillars of content optimization are:

- Help content be found
- Once content is found, ensure it is engaging for the user
- Once the user is engaged, ensure it converts (achieves its objective)

It's early days, but already INK is boosting user engagement. For example, it examines what writing style search users for your target keyphrase are best responding to:

- How to use Passive Voice
- How to use Adverbs
- How to write to the ideal reading grade level

For example, in one study we found that users frequently stop reading and start 'skimming', when they encounter a sentence which is 2 reading grade levels above what INK recommends.

INK will help you by highlighting those sentences as Very Hard To Read, so you can keep your readers' attention for longer.

How INK Revolutionizes Content Performance Optimization for Writers

There's a moment during every project where you sit down and try to determine the best business model. You talk to many advisors and you look at your audience.

We know there is a lot of money in the SEO market, and we know there is a demand there. On the other hand, we know that content writers are not as problem aware, and often don't get the same compensation packages as their marketing colleagues.

Some may see this as a disadvantage, but we saw it as an incredible opportunity. As I hope this whitepaper has established beyond a doubt, a solution like this is much-needed. The demand and need for frustration-free content performance optimization in the writing community is enormous.

One vague tweet about our solution flooded our PR agent with over 400 DMs. Our beta users are ecstatic. A few months ago, we broke something on the backend and INK didn't work for half an hour. We received messages like: "How am I supposed to write content now? Writing without INK is like flying an airplane blindfolded." Once you start writing your content in INK, it's extremely sticky. You can't imagine going without it.

Therefore, we've decided that the best business model for us is to first let INK fulfill its destiny and become widely beloved. The easiest way to get there is to remove any obstacles - and make this version of INK available for free.

In the introduction, I mentioned that existing tools universally have two things in common:

1. You have to pay up, big time. Prices range between \$100 and \$3,000 per month. The most advanced solutions don't publish their prices on their site.
2. All the tools we've examined are built as a side-project of an SEO suite (me-too), or are built with marketers and SEOs as their primary audience. There is no solution that's specifically for writers.

In contrast, INK was built from the ground up with our main audience in mind. Our goal is to take your feedback and make INK better and better. In addition, we've got a very exciting roadmap of AI magic ahead of us. But that story is to be continued for another day. Welcome to the Content Performance Optimization era!

Grab your copy of INK for free at inkforall.com.

References

[1] https://en.wikipedia.org/wiki/Common_Gateway_Interface

[2] [https://en.wikipedia.org/wiki/Ajax_\(programming\)](https://en.wikipedia.org/wiki/Ajax_(programming))

[3] <https://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html>

- [4] <https://en.wikipedia.org/wiki/DeepFace>
- [5] <https://en.wikipedia.org/wiki/Yoast>
- [6] <https://en.wikipedia.org/wiki/RankBrain>
- [7] <https://yoast.com/>
- [8] <http://www.hemingwayapp.com/>
- [9] <https://readable.com/>
- [10] <https://buzzsumo.com/>
- [11] <https://backlinko.com/content-study>
- [12] <https://backlinko.com/search-engine-ranking>
- [13] <https://en.wikipedia.org/wiki/T%E2%80%93idf>
- [14] https://en.wikipedia.org/wiki/Natural-language_understanding
- [15] https://en.wikipedia.org/wiki/Word_embedding
- [16] <https://www.youtube.com/watch?v=J47Wk5-ayQw&feature=youtu.be&t=1845>
- [17] <https://support.google.com/webmasters/answer/66358?hl=en>
- [18] <https://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf>
- [19] <https://edgy.app/why-did-google-plus-fail-a-google-autopsy>
- [20] https://en.wikipedia.org/wiki/Artificial_intelligence
- [21] <https://www.spotify.com/us/>
- [22] <https://github.com/spotify/luigi>
- [23] https://en.wikipedia.org/wiki/The_Apache_Software_Foundation
- [24] https://en.wikipedia.org/wiki/UC_Berkeley
- [25] <https://en.wikipedia.org/wiki/AMPLab>
- [26] <https://en.wikipedia.org/wiki/Databricks>
- [27] https://en.wikipedia.org/wiki/Apache_Spark
- [28] https://en.wikipedia.org/wiki/Google_Brain
- [29] <https://en.wikipedia.org/wiki/TensorFlow>