@IPULLRANKAGENCY

iPULLRANK

Technical Content Optimization

Presented by: Michael King
Contact: mike@ipullrank.com
Twitter: @iPullRank

Disclaimer.

No one knows Google's algorithms.

Disclaimer.

No one knows Google's algorithms.

Patents are not an indication that something is currently in use.

Disclaimer.

No one knows Google's algorithms.

Patents are not an indication that something is currently in use.
I'm not aiming to offend anyone with this presentation.

Search Engine Expectations

How do Search Engines See Content?

My Hypothesis

Just my educated opinion...

MyCool King
@iPullRank

Thinking in terms of hacking, I wonder if Google's algos have always been more elegant than we thought and we've just been using links as the equivalent of brute force password cracking.

There have probably always been better "exploits" we've just always overindexed on links.

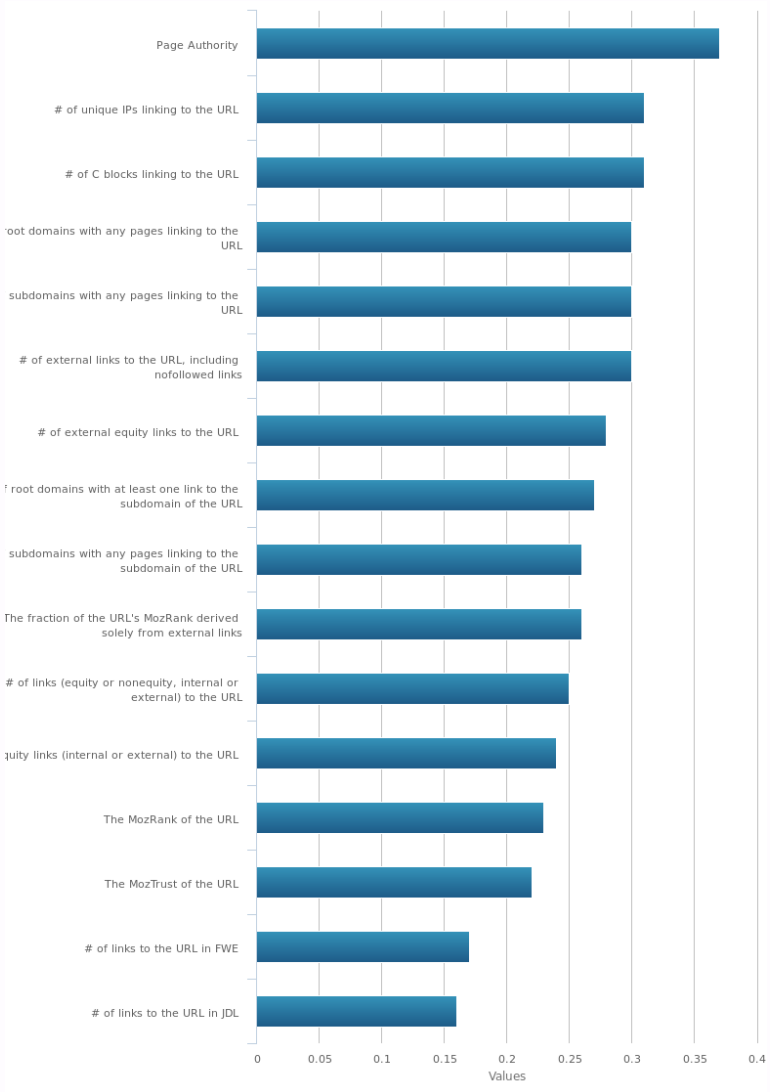
11:21 AM - 31 Jul 2018

6 Retweets **41** Likes

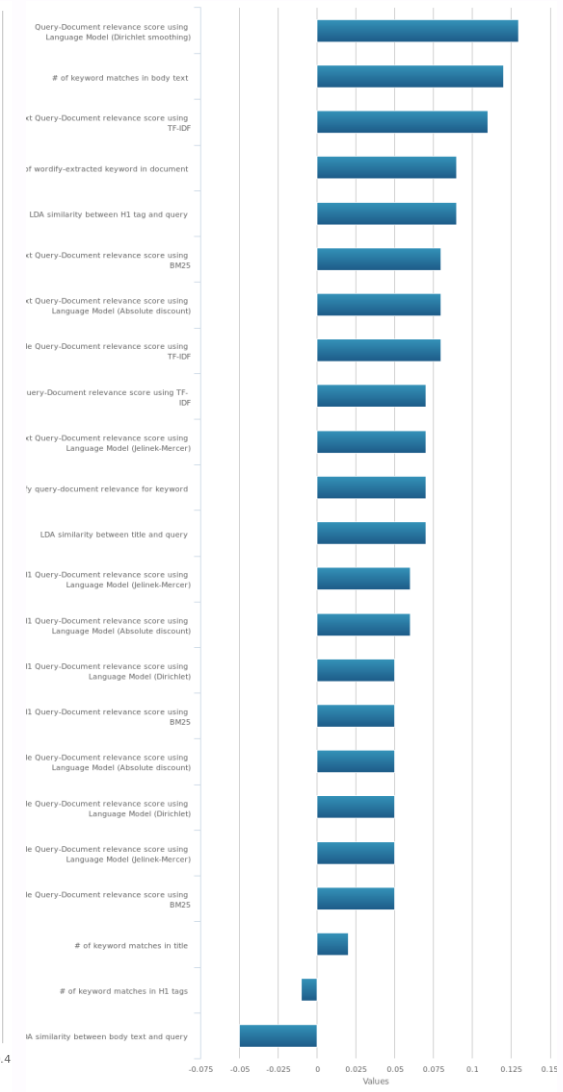
Our Correlation Studies Showcase Our Bias, Not Google's

I believe that because we are biased by the impact of links, we focus heavily on link acquisition. As a result our correlation studies reflect that.

Link Analysis Correlations

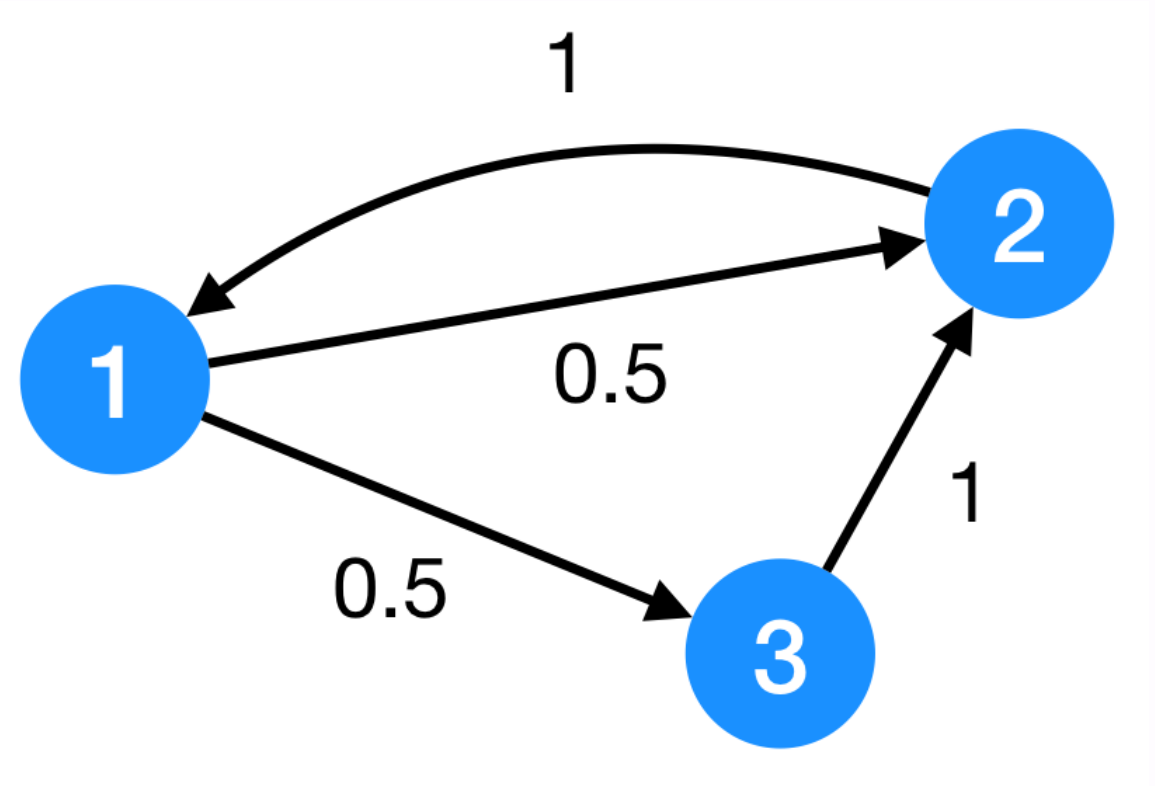


Text Analysis Correlations



PageRank

PageRank certainly changed the game. However, at its simplest, it is largely just about link volume.



Google Cares About Content Relevance and Parity

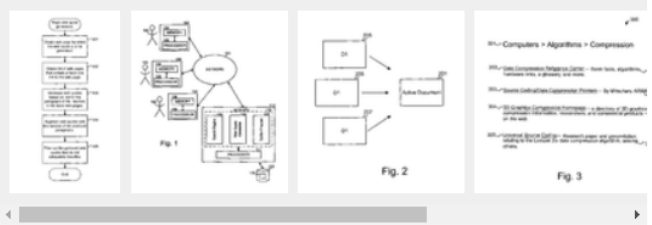
In addition to the volume of links, the source and target content relevance is examined.

Using text surrounding hypertext links when indexing and generating page summaries

Abstract

Web quotes are gathered from web pages that link to a web page of interest. The web quote may include text from the paragraphs that contain the hypertext links to the page of interest as well as text from other portions of the linked web page, such as text from a nearby header. The obtained web quotes may be ranked based on quality or relevance and may then be incorporated into a search engine's document index or into summary information returned to users in response to a search query.

Images (6)



Classifications

G06F17/30864 Retrieval from the Internet, e.g. browsers by querying, e.g. search engines or meta-search engines, crawling techniques, push systems

US8495483B1
US Grant

Download PDF | Find Prior Art | Similar

Inventor: Jeffrey A. Dean, Martin Farach-Colton, Sanjay Ghemawat, Benedict Gomes, Georges R. Hank

Current Assignee : Google LLC

Original Assignee: Google LLC

Priority date : 2002-03-13

Family: US (1)

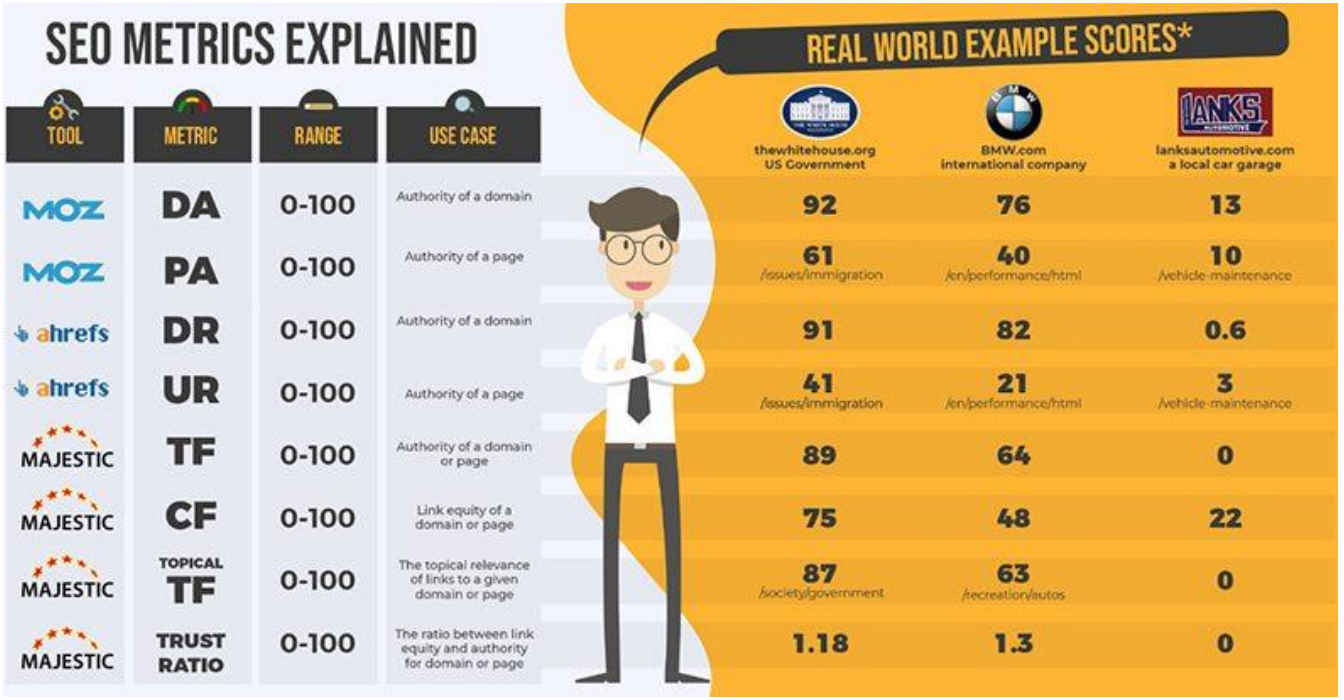
Date	App/Pub Number	Status
2003-03-12	US10386110	Active
2013-07-23	US8495483B1	Grant

Info: Patent citations (18), Non-patent citations (11), Cited by (1), Legal events, Similar documents, Priority and Related Applications

Our Metrics Mostly Measure Volume

First, our link metrics are, at best, just approximations. They are not necessarily indicative of anything that Google is actually doing or measuring.

Second, our link metrics are mostly some measure of link volume.



Citation Trust Flow

Except for Majestic.

Topical PageRank

It is not realistic to expect that Google has been using pure PageRank. Rather, it makes way more sense to expect they have defaulted to some form of Topical PageRank for several years.

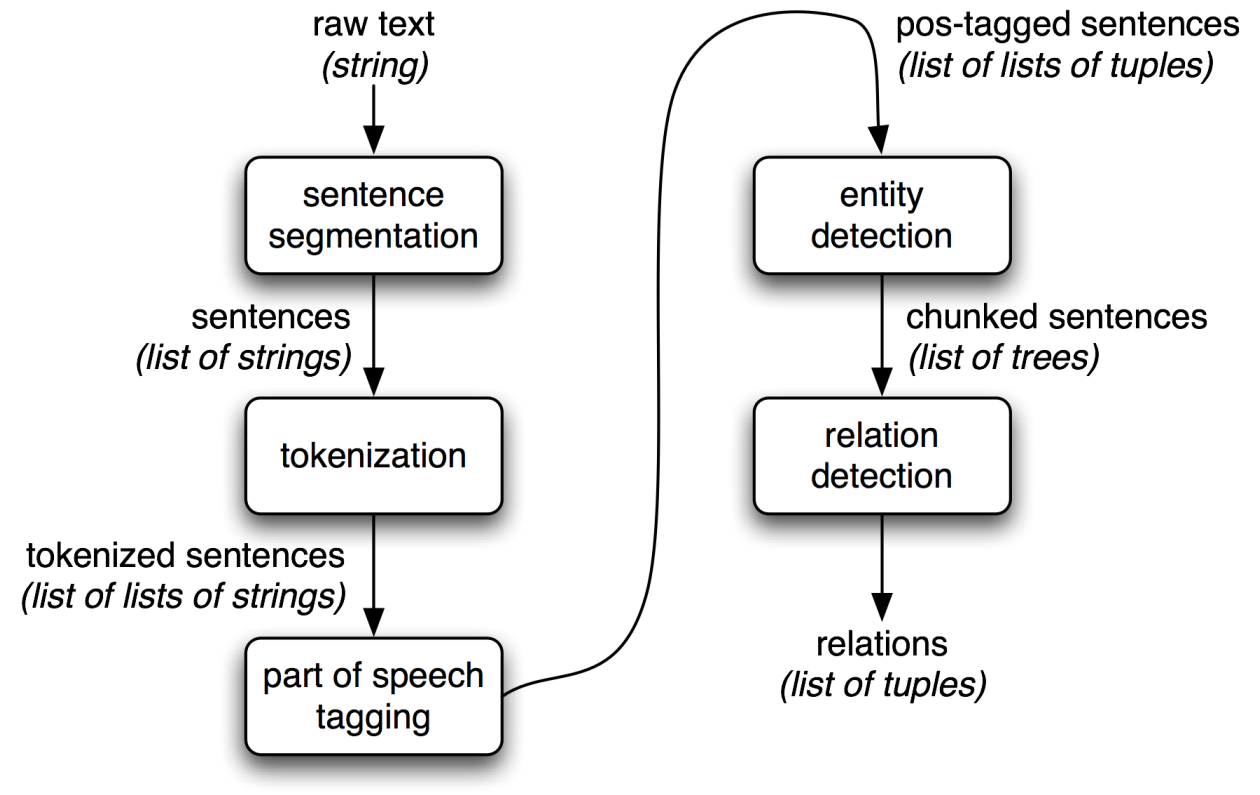
h/t @CyrusShepard

Disclaimer.

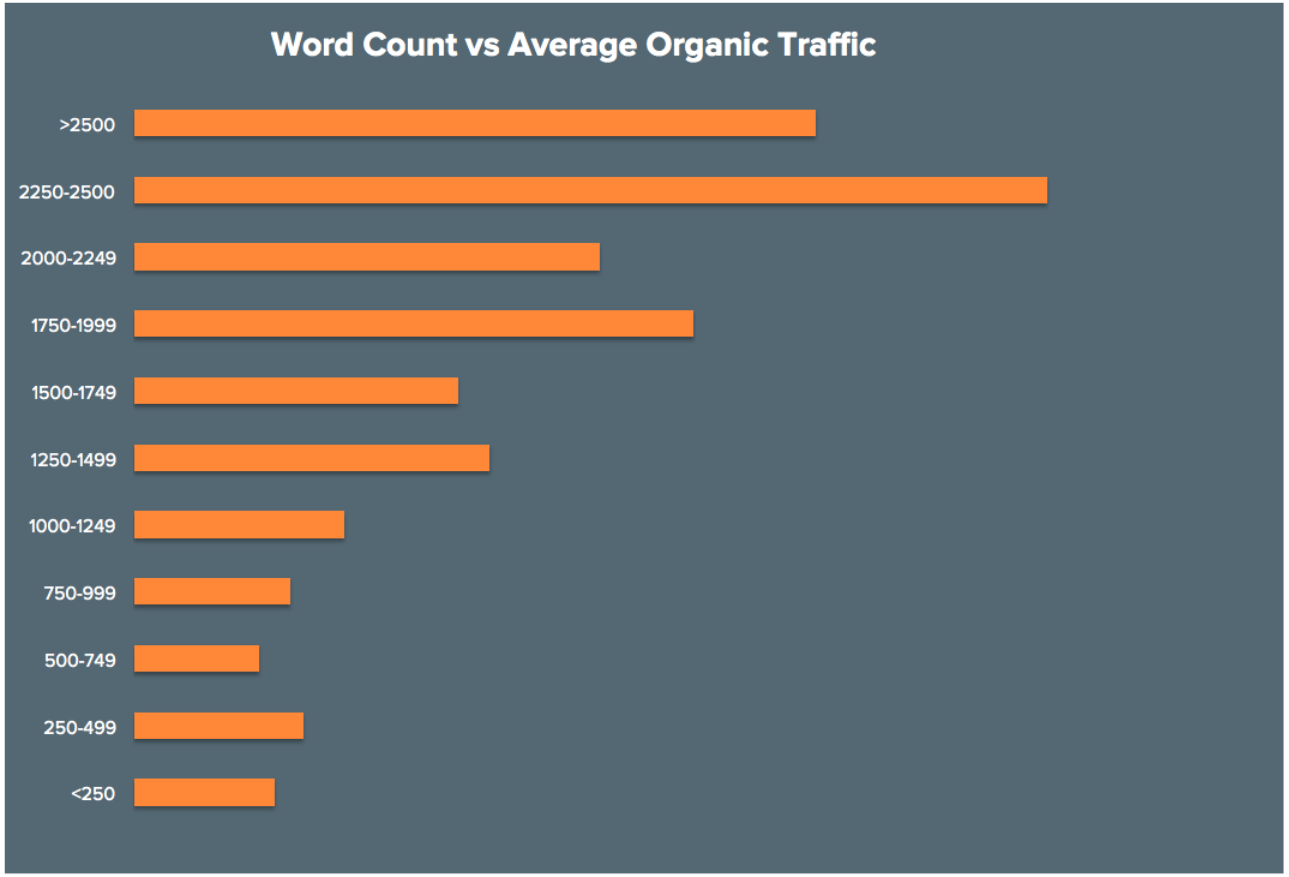
But we just want the link, no matter what?

How Content is Understood in Information Retrieval

Content is tokenized and then relationships between terms are determined statistically.

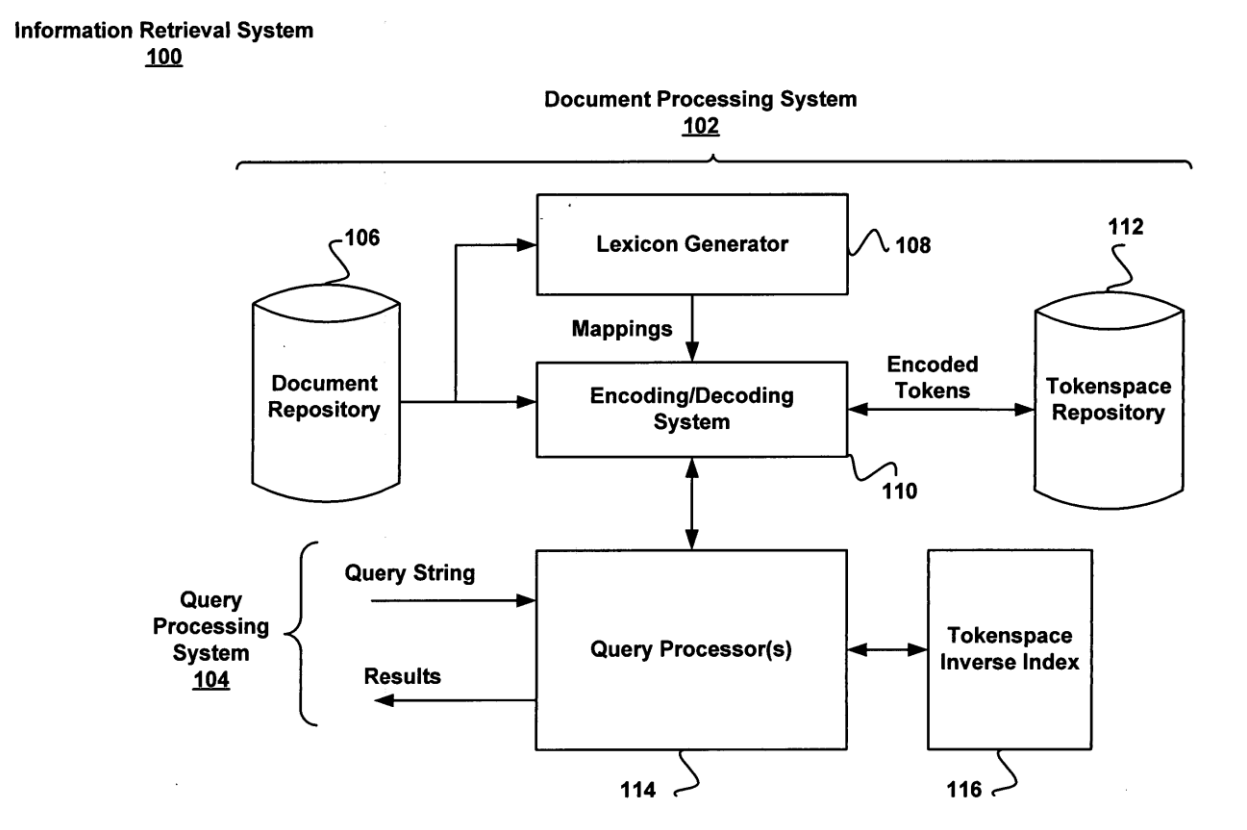


Yet The Question is Always How Long Should Content Be?



How Content is Understood by Google

Google has a more sophisticated and living system for doing the same thing.

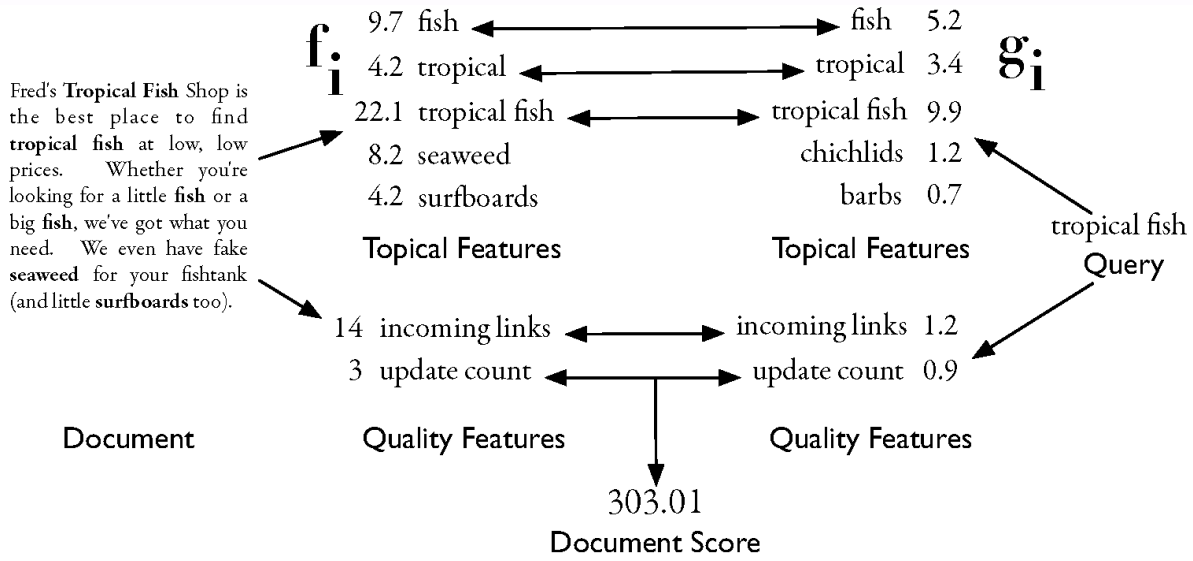


Content.

But we just “make great content?”

Content Scoring and Storing

Search engines can score documents based on a variety of features. Google's file system suggests that they have the capability to store multiple versions of any document over time.



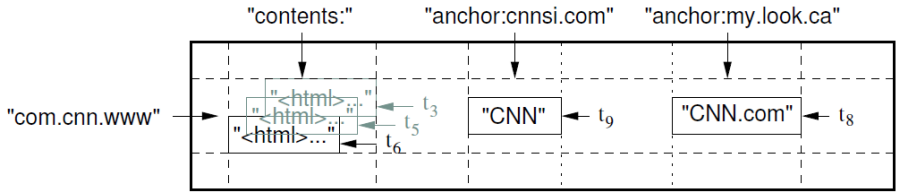


Figure 1: A slice of an example table that stores Web pages. The row name is a reversed URL. The contents column family contains the page contents, and the anchor column family contains the text of any anchors that reference the page. CNN's home page is referenced by both the Sports Illustrated and the MY-look home pages, so the row contains columns named anchor:cnnsi.com and anchor:my.look.ca. Each anchor cell has one version; the contents column has three versions, at timestamps t_3, t_5, and t_6.

Google's Scoring Functions

There's more than one scoring function. Google scores content and links a variety of different ways and then chooses the best results.

There is not just one “algorithm."

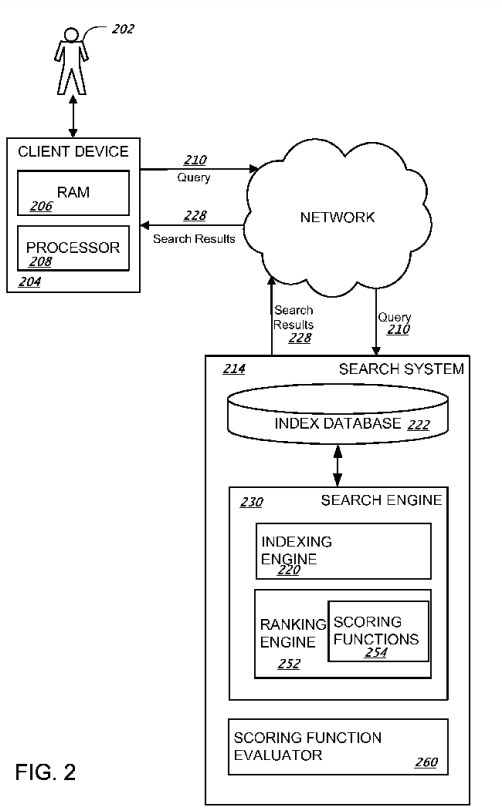


FIG. 2

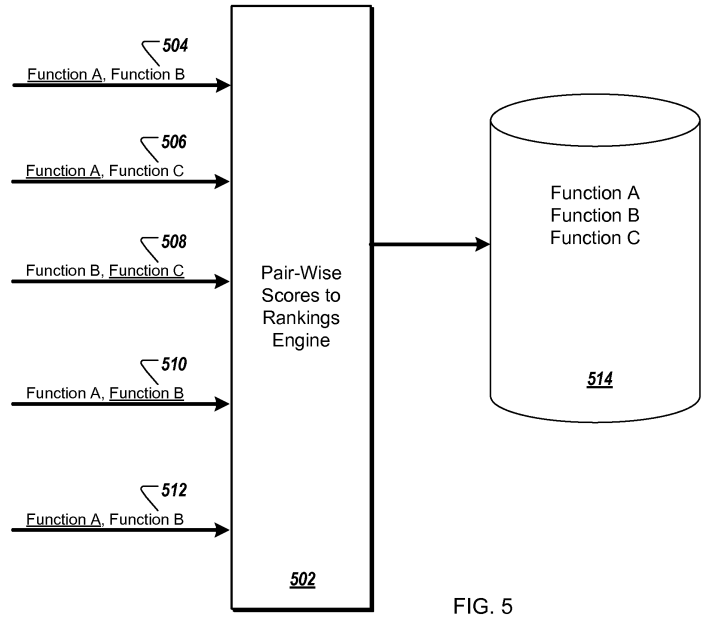
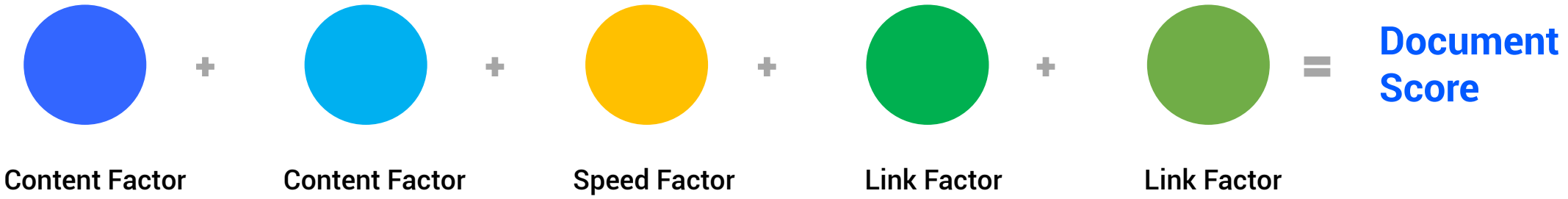
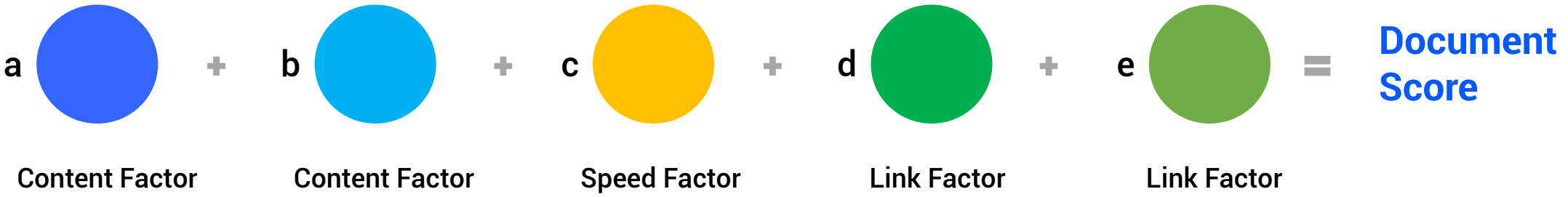


FIG. 5

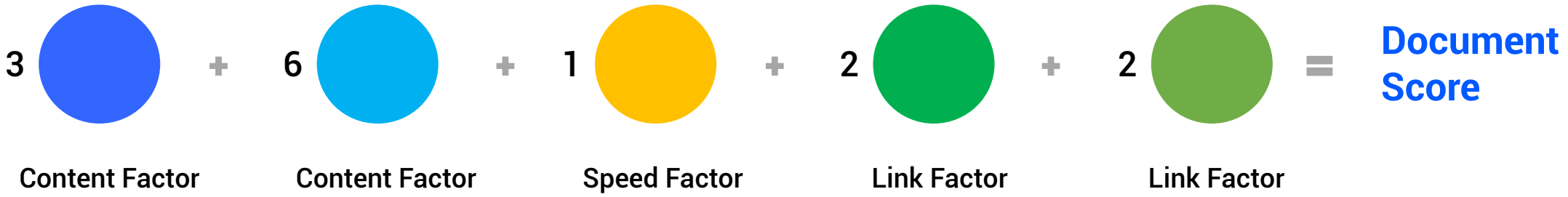
Document Scoring Simplified



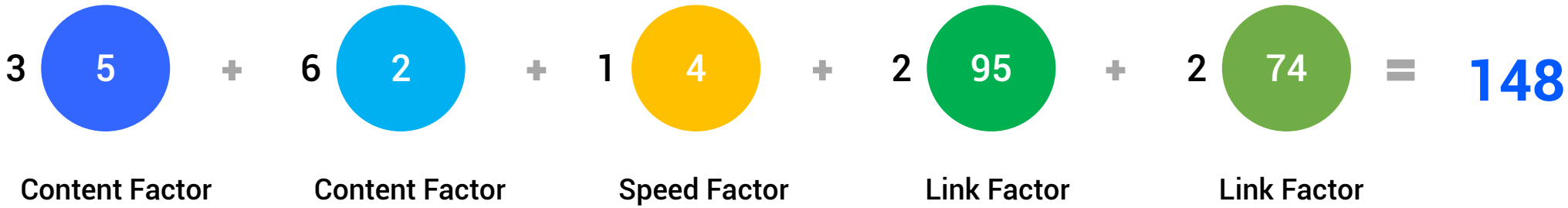
Each Component to the Equation Has a Weight



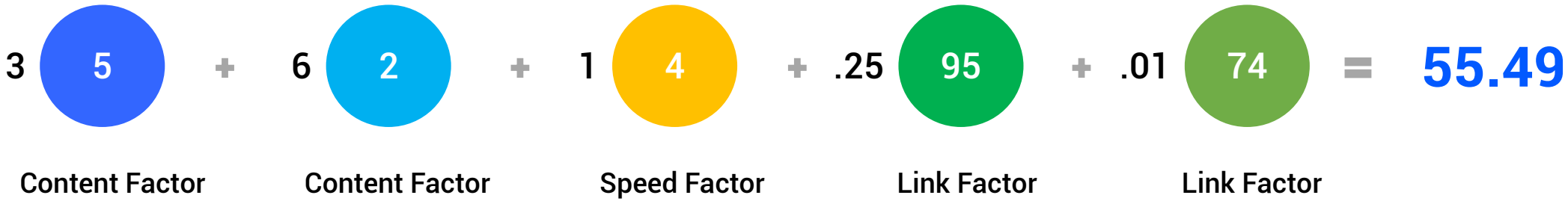
The Weights May Look Like This



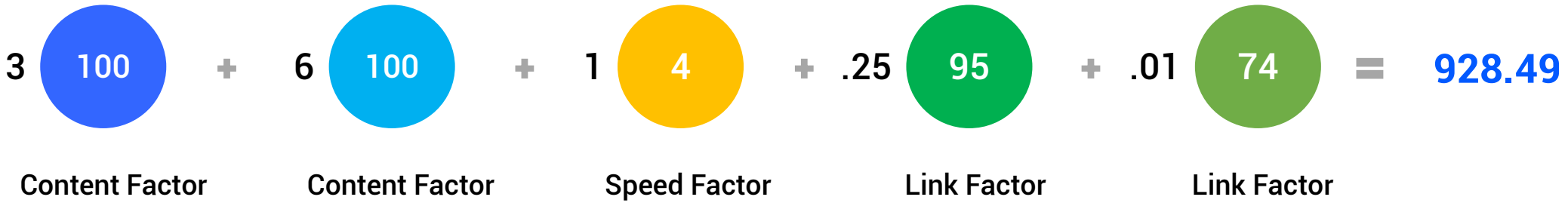
This is What SEOs Do



So, Then, Google Turns Down the Weights on Links



When Really, Content Optimization May Go Further



Google Constantly Evaluates and Experiments

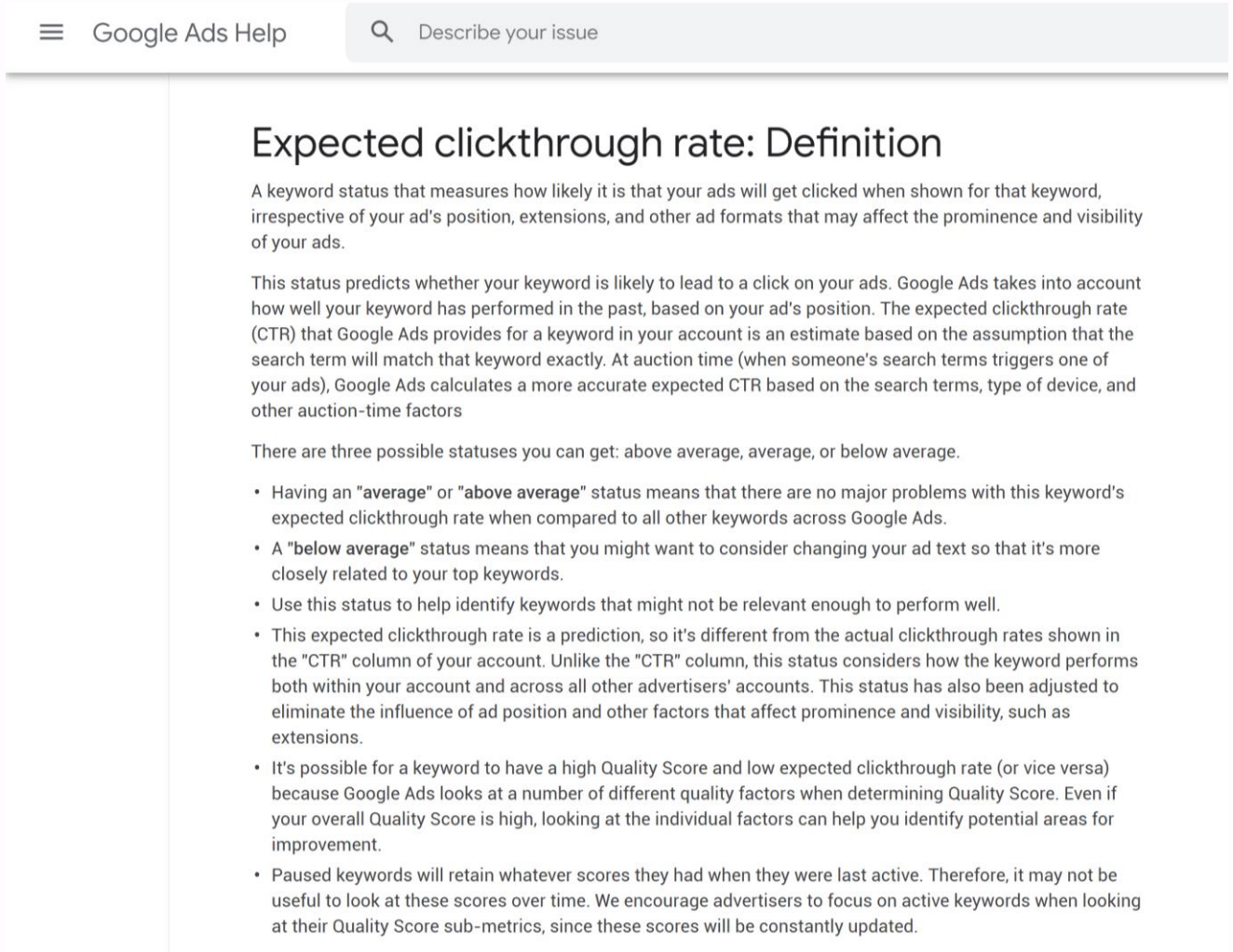
DS: Last month, we had the Quality Update. How is Google assessing the quality? How do clicks factor in?

GI: We use clicks in different ways. The main things that we use clicks for evaluation and experimentation. There are many, many people who are trying to induce noise in clicks. Rand Fishkin, for example is experimenting with clicks. Using clicks directly in ranking would be a mistake. In personalized results, if you search for apple, we would most likely serve you a disambiguation box. We have to figure out if you mean the company or the food. Then, we'd look at the click you made.

There is Evidence of CTR Expectations

Granted, the screenshot is from the Paid side, but this suggests that Google has at least considered an expected CTR. This is also consistent with the idea of evaluating a successful search based on click activity.

The hypothesis being that if a result does not consistently meet the expected CTR, it should be demoted.



Google Ads Help

Describe your issue

Expected clickthrough rate: Definition

A keyword status that measures how likely it is that your ads will get clicked when shown for that keyword, irrespective of your ad's position, extensions, and other ad formats that may affect the prominence and visibility of your ads.

This status predicts whether your keyword is likely to lead to a click on your ads. Google Ads takes into account how well your keyword has performed in the past, based on your ad's position. The expected clickthrough rate (CTR) that Google Ads provides for a keyword in your account is an estimate based on the assumption that the search term will match that keyword exactly. At auction time (when someone's search terms triggers one of your ads), Google Ads calculates a more accurate expected CTR based on the search terms, type of device, and other auction-time factors

There are three possible statuses you can get: above average, average, or below average.

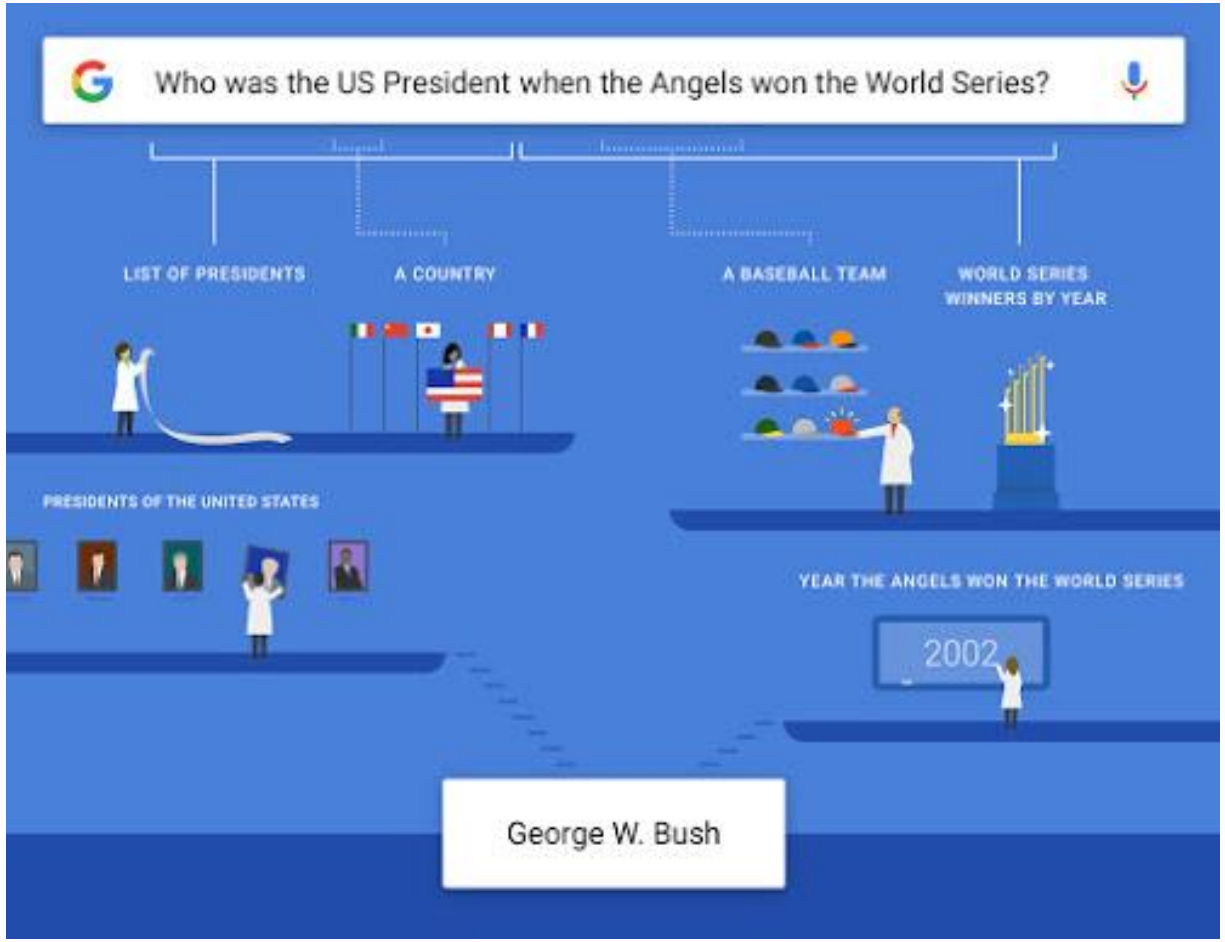
- Having an "**average**" or "**above average**" status means that there are no major problems with this keyword's expected clickthrough rate when compared to all other keywords across Google Ads.
- A "**below average**" status means that you might want to consider changing your ad text so that it's more closely related to your top keywords.
- Use this status to help identify keywords that might not be relevant enough to perform well.
- This expected clickthrough rate is a prediction, so it's different from the actual clickthrough rates shown in the "CTR" column of your account. Unlike the "CTR" column, this status considers how the keyword performs both within your account and across all other advertisers' accounts. This status has also been adjusted to eliminate the influence of ad position and other factors that affect prominence and visibility, such as extensions.
- It's possible for a keyword to have a high Quality Score and low expected clickthrough rate (or vice versa) because Google Ads looks at a number of different quality factors when determining Quality Score. Even if your overall Quality Score is high, looking at the individual factors can help you identify potential areas for improvement.
- Paused keywords will retain whatever scores they had when they were last active. Therefore, it may not be useful to look at these scores over time. We encourage advertisers to focus on active keywords when looking at their Quality Score sub-metrics, since these scores will be constantly updated.

Experiments.

But we don't A/B test our metadata?

Queries Become Entities

Google has explicitly indicated that they break a query into a series of entities first.



Attributes of Entities Inform Context

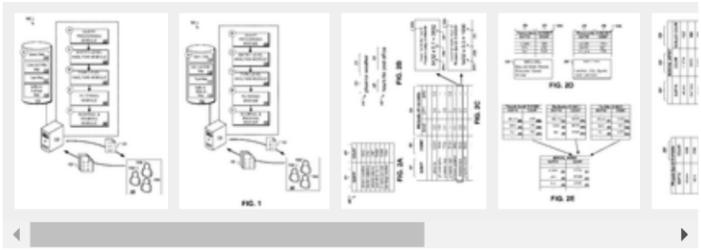
Google builds context around the query leveraging the information about the entity to inform what documents to consider.

Identifying and ranking attributes of entities

Abstract

Methods, systems, and apparatus, including computer programs encoded on a computer storage medium, for query analysis. Queries are identified in query data, and an entity-descriptive portion and a suffix are determined in each query. Query counts are determined for a number of times that the respective queries occur in the query data. Based on the query counts, an entity-level count is estimated, which represents a number of query submissions that include the particular suffix and are considered to refer to a first entity. The entity is determined to be a particular type of entity. A type-level count is determined, which represents a number of query submissions that include the first suffix and are estimated to refer to entities of the particular type of entity. A score is assigned to the particular suffix based on the entity-level count and the type-level count.

Images (7)



Classifications

G06F17/30

US9047278B1
United States

Download PDF Find Prior Art
Similar

Inventor: Benjamin J. Mann, Randolph G. Brown, John R. Provine, Vinicius J. Fortuna, Andrew W. Hogue

Current Assignee : Google LLC

Worldwide applications

2012 • US 2015 • US

Application US13/673,015 events

2012-11-09 • Application filed by Google LLC
2012-11-09 • Priority to US13/673,015
2013-05-09 • Assigned to GOOGLE INC.
2015-06-02 • Application granted
2015-06-02 • Publication of US9047278B1
2017-10-02 • Assigned to GOOGLE LLC

Entities are Extracted From Content to Inform Rankings

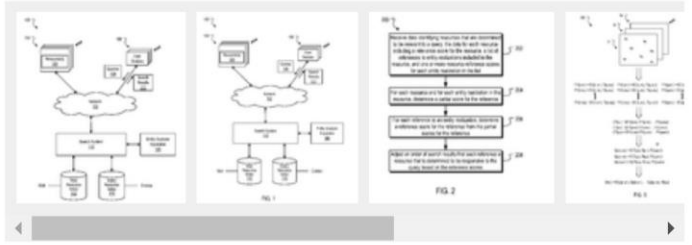
When Google is processing content, they extract ad score entities.

Ranking entity realizations for information retrieval

Abstract

Methods, systems, and apparatus, including computer programs encoded on a computer storage medium, for identifying and ranking entities for reference as search results. In one aspect, a method includes receiving data identifying resources that are relevant to a query. The data for each resource can include a relevance score, a list of references to entity realizations included in the resource, and for each reference to an entity realization in the list, one or more resource reference scores. For each resource and for each reference to an entity realization in the resource, a partial score for the reference can be determined from the resource reference scores for the reference and the relevance score for the resource. For each reference to an entity realization, a reference score for the reference is determined from each of the partial scores for the reference. Search results can be ranked based on the reference scores.

Images (7)



Classifications

G06F17/3053

US20150161127A1
United States

Download PDF | Find Prior Art
Similar

Inventor: Samuel C. Oates, Matthew K. Gray

Current Assignee : Google LLC

Worldwide applications

2013 • ~~US~~

Application US13/765,975 events

2012-02-13 • Priority to US201261598133P

2013-02-13 • Application filed by Google LLC

2013-02-13 • Priority to US13/765,975

2013-04-10 • Assigned to GOOGLE INC.

2015-06-11 • Publication of US20150161127A1

2019-04-02 • Application status is Abandoned

Entity Salience.

But our industry doesn't develop entity strategies?

Intent.

Intent for a given keyword can change over time.

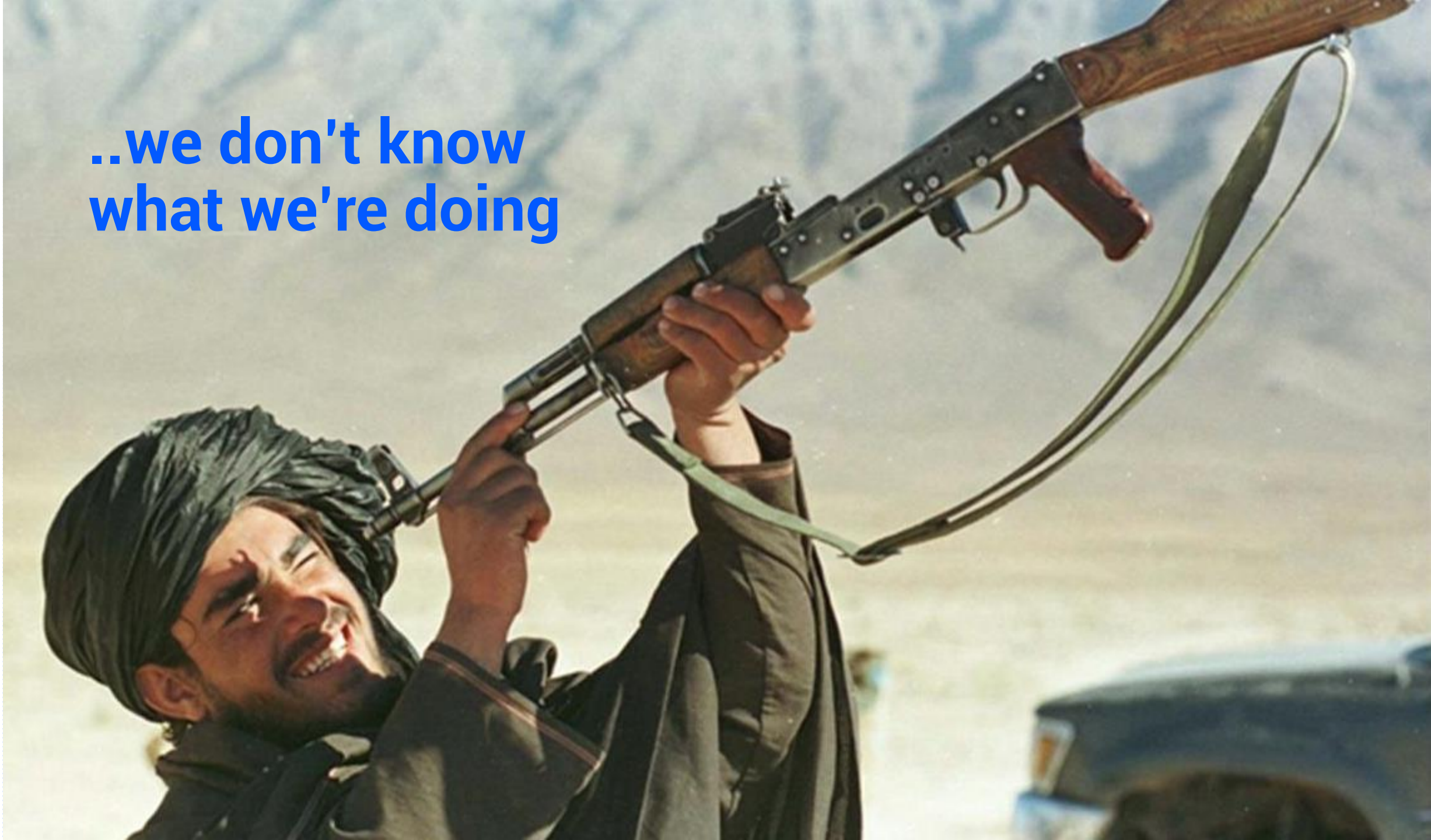
Intent.

Intent for a given keyword can change over time.

Yet we set and forget our keyword research?

Technical Content Optimization is a Multi-dimensional Word Game, but…

..we don't know what we're doing



Mind the Text Analysis Knowledge Gap

What is missing from the SEO skillset

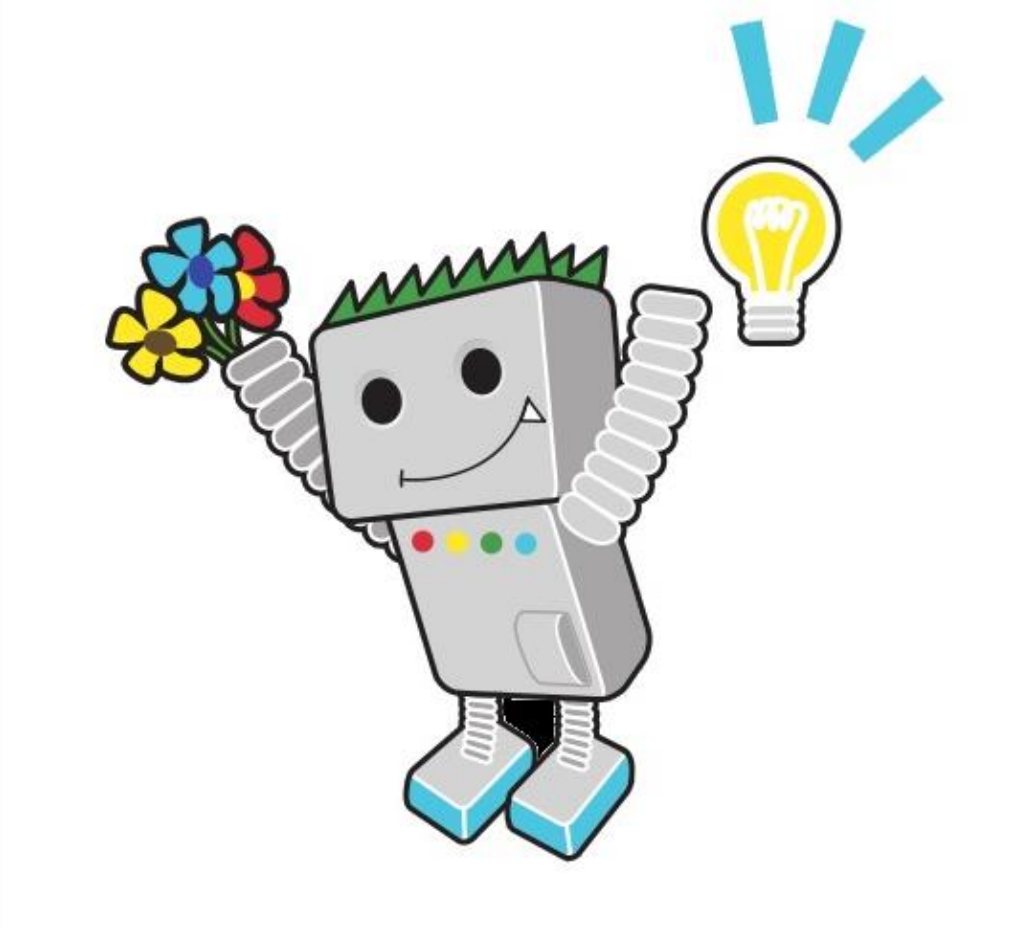
We're Entering Different Fields

We are no longer talking about SEO. Rather, we are talking about natural language processing, computational linguistics, information retrieval and graph theory. All things that conceptually underpin the computer science behind how search engines model topics in content.

Natural Language Processing Use Cases

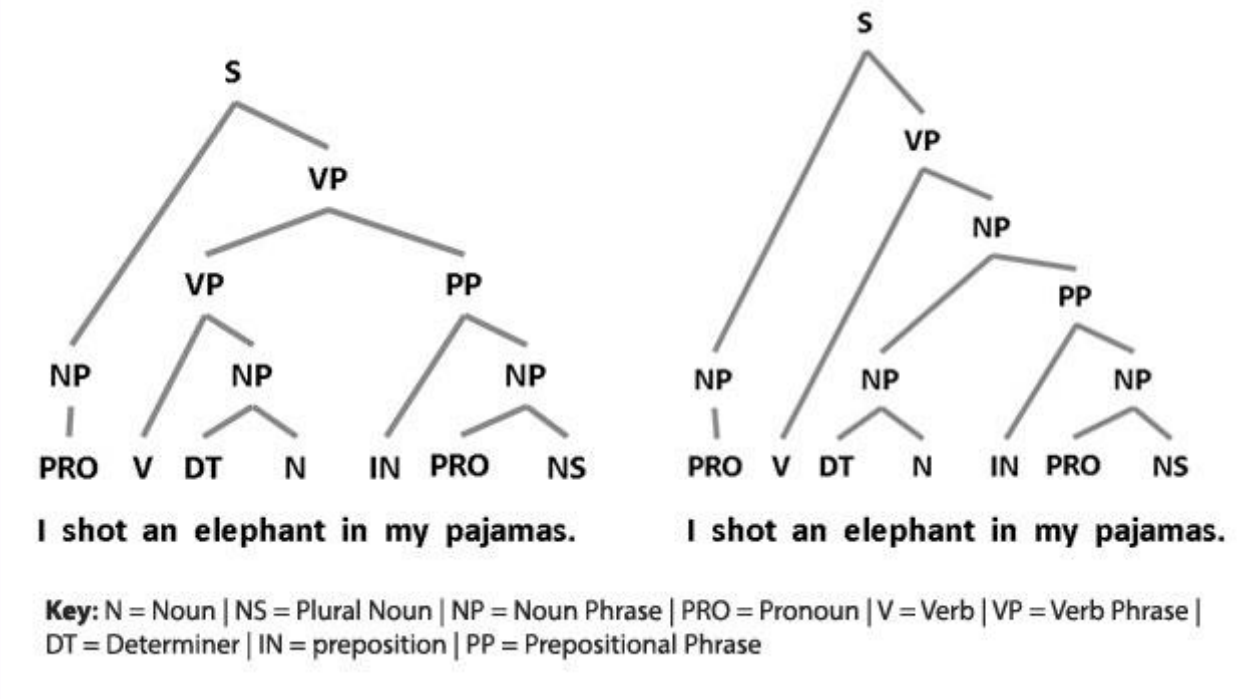
A “Corpus" is Built from Crawling

Googlebot crawls the web and builds a corpus of web documents. This is how the process begins.



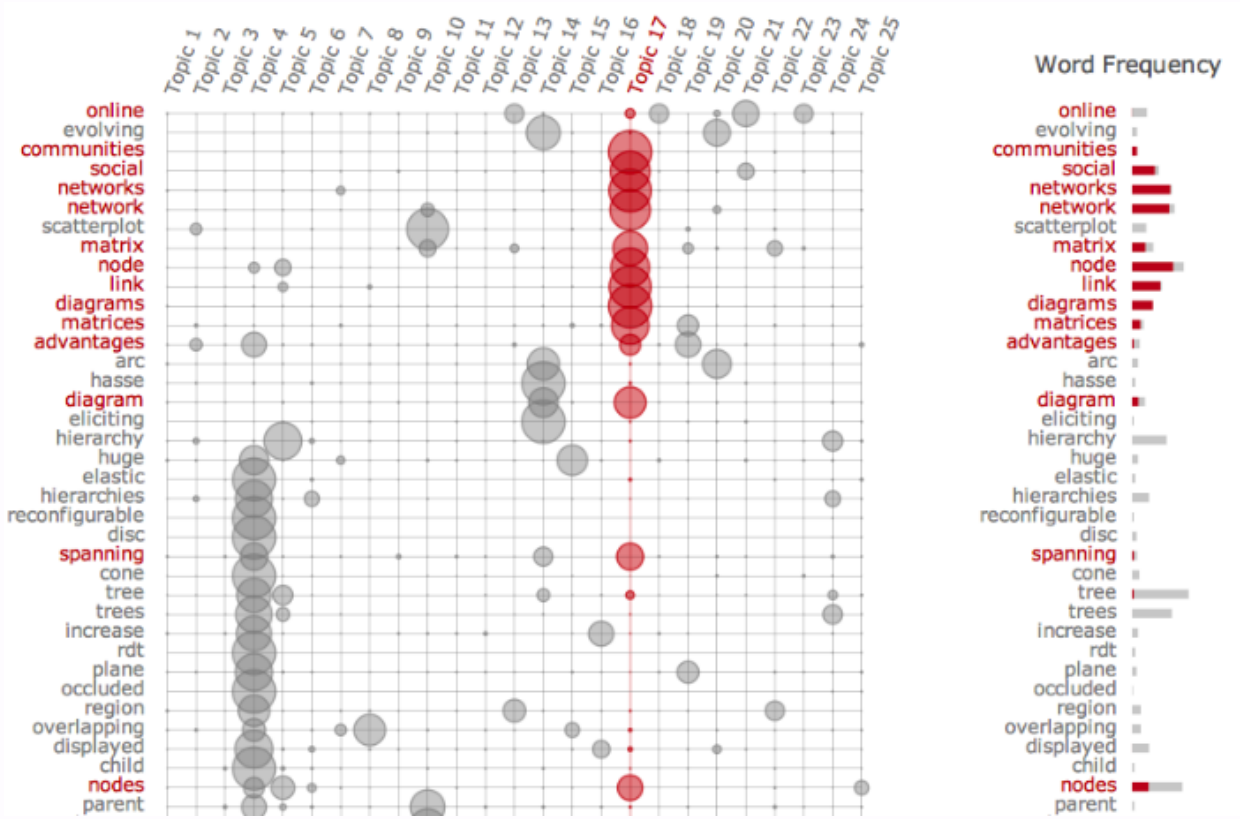
Documents are Parsed

Documents are broken down into words and then tagged with parts of speech.



Topics are Modeled

A series of transformations and analyses are done on the content to understand what they are about.



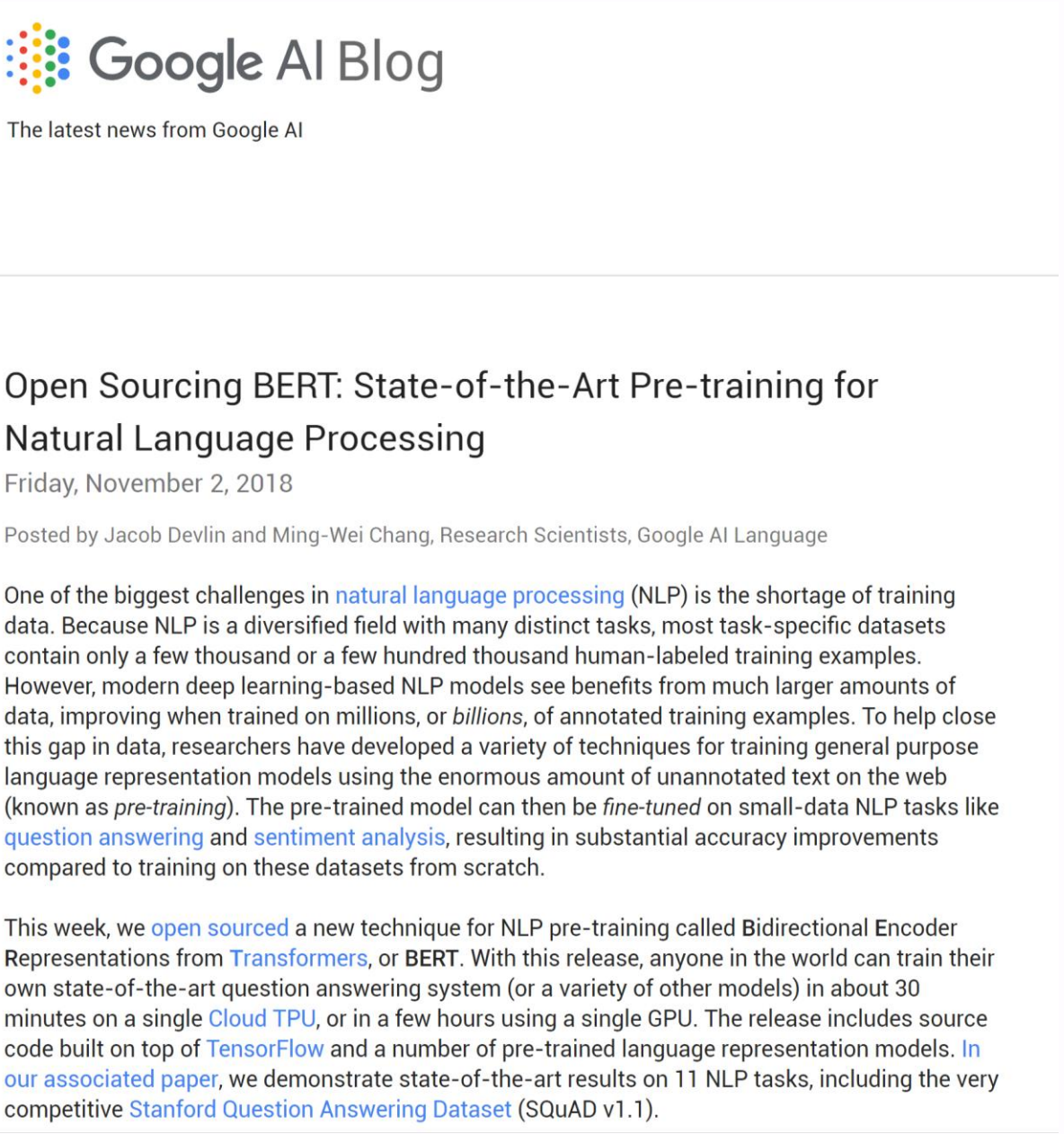
Language is Understood

Once the content is understood, and features are extracted, it can then be scored and used for rankings. This is how search engines develop their expectation of what should be featured in content.

Google Has Something Called BERT

BERT is Google's technique for pre-training NLP for classification, question answering and named entity recognition.

https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html



Google AI Blog

The latest news from Google AI

Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing

Friday, November 2, 2018

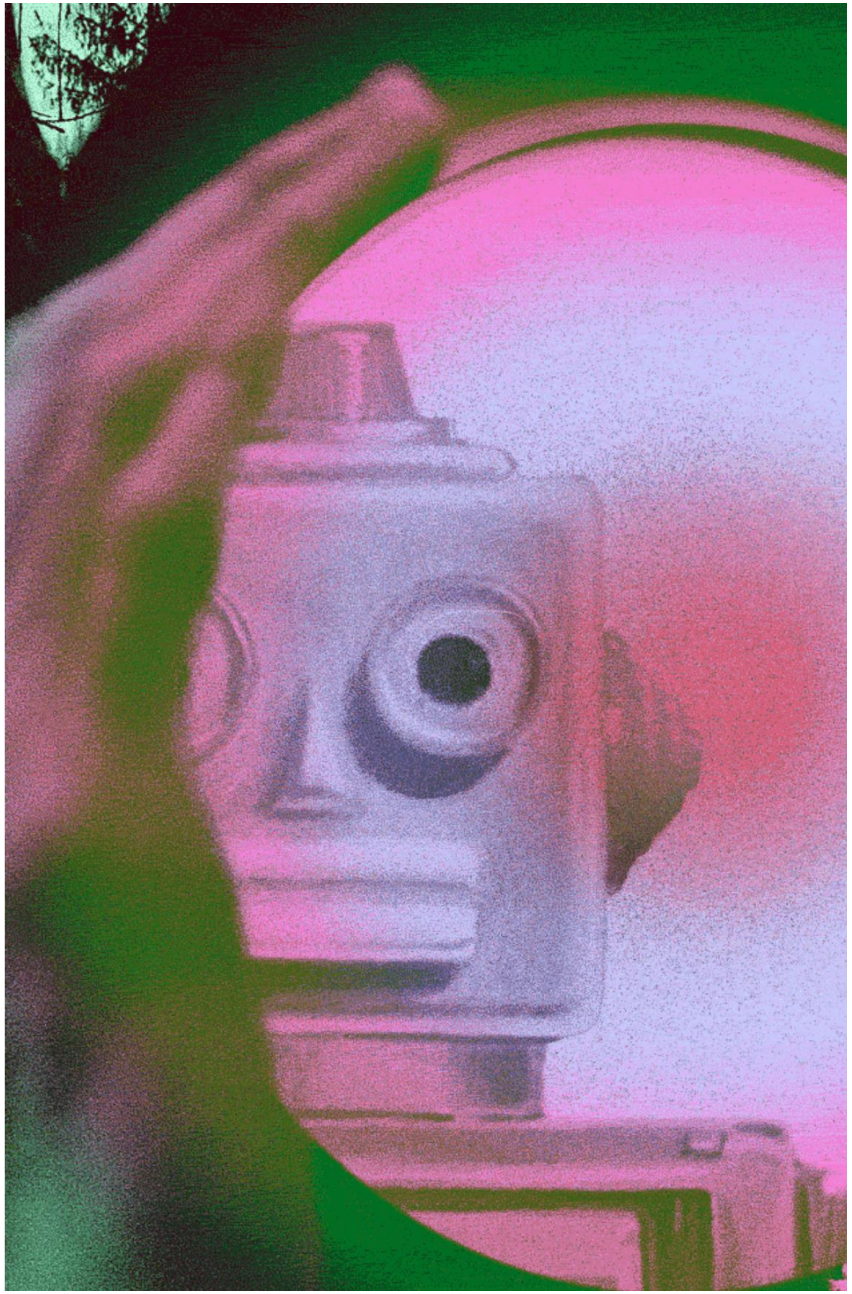
Posted by Jacob Devlin and Ming-Wei Chang, Research Scientists, Google AI Language

One of the biggest challenges in natural language processing (NLP) is the shortage of training data. Because NLP is a diversified field with many distinct tasks, most task-specific datasets contain only a few thousand or a few hundred thousand human-labeled training examples. However, modern deep learning-based NLP models see benefits from much larger amounts of data, improving when trained on millions, or *billions*, of annotated training examples. To help close this gap in data, researchers have developed a variety of techniques for training general purpose language representation models using the enormous amount of unannotated text on the web (known as *pre-training*). The pre-trained model can then be *fine-tuned* on small-data NLP tasks like question answering and sentiment analysis, resulting in substantial accuracy improvements compared to training on these datasets from scratch.

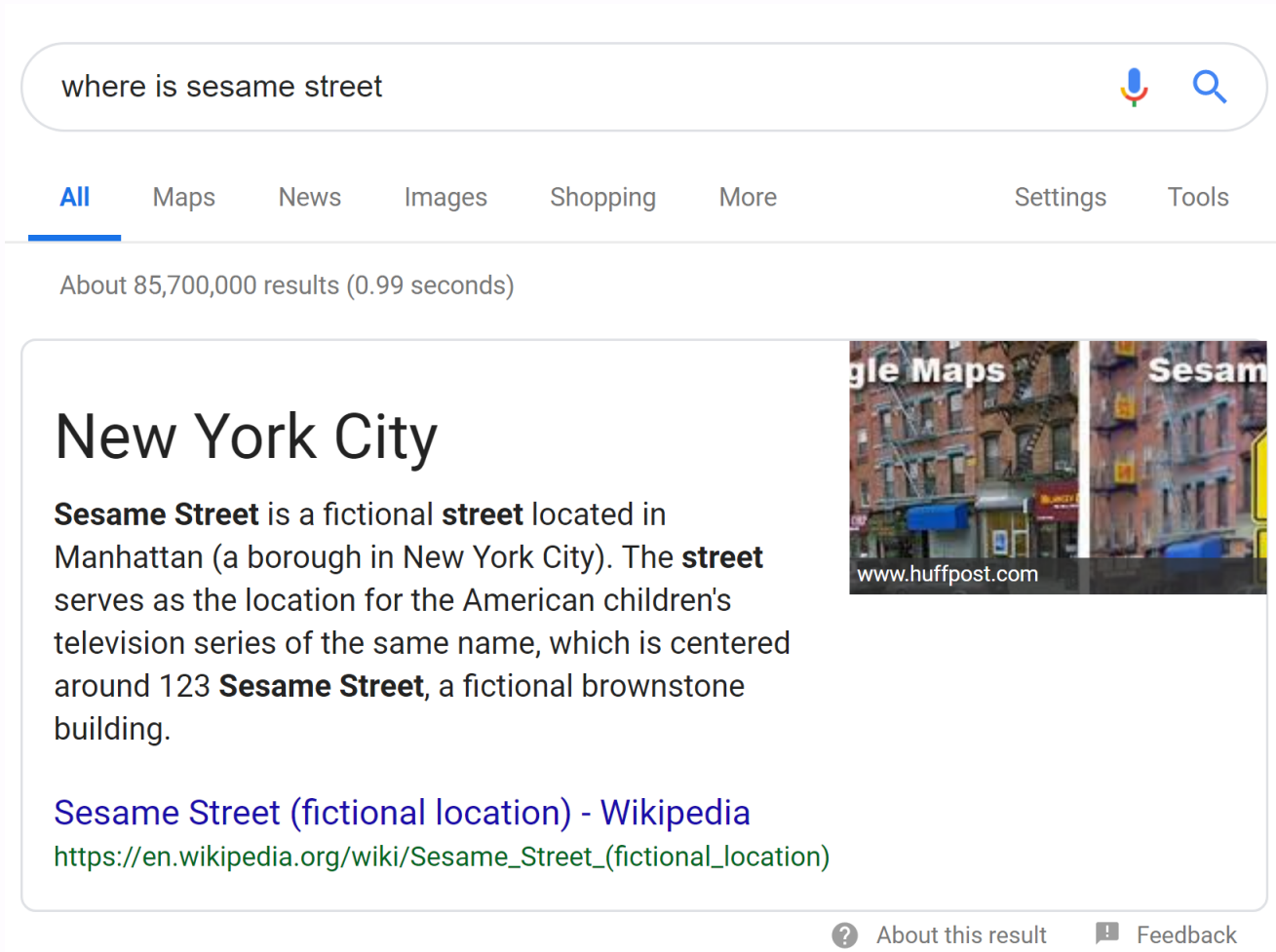
This week, we open sourced a new technique for NLP pre-training called **B**idirectional **E**ncoder **R**epresentations from Transformers, or **BERT**. With this release, anyone in the world can train their own state-of-the-art question answering system (or a variety of other models) in about 30 minutes on a single Cloud TPU, or in a few hours using a single GPU. The release includes source code built on top of TensorFlow and a number of pre-trained language representation models. In our associated paper, we demonstrate state-of-the-art results on 11 NLP tasks, including the very competitive Stanford Question Answering Dataset (SQuAD v1.1).

Finally, a Machine That Can Finish Your Sentence

Completing someone else's thought is not an easy trick for A.I. But new systems are starting to crack the code of natural language.



In other words, it's the thing that does this.



Learn More about BERT (and ELMo)

Jay Alammar has a great article with illustrations that walks you through BERT and some of the other pre-existing NLP mechanisms

http://jalammar.github.io/illustrated-bert/

Jay Alammar

Visualizing machine learning one concept at a time

Blog About

The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)

Discussions: Hacker News (98 points, 19 comments), Reddit r/MachineLearning (164 points, 20 comments)

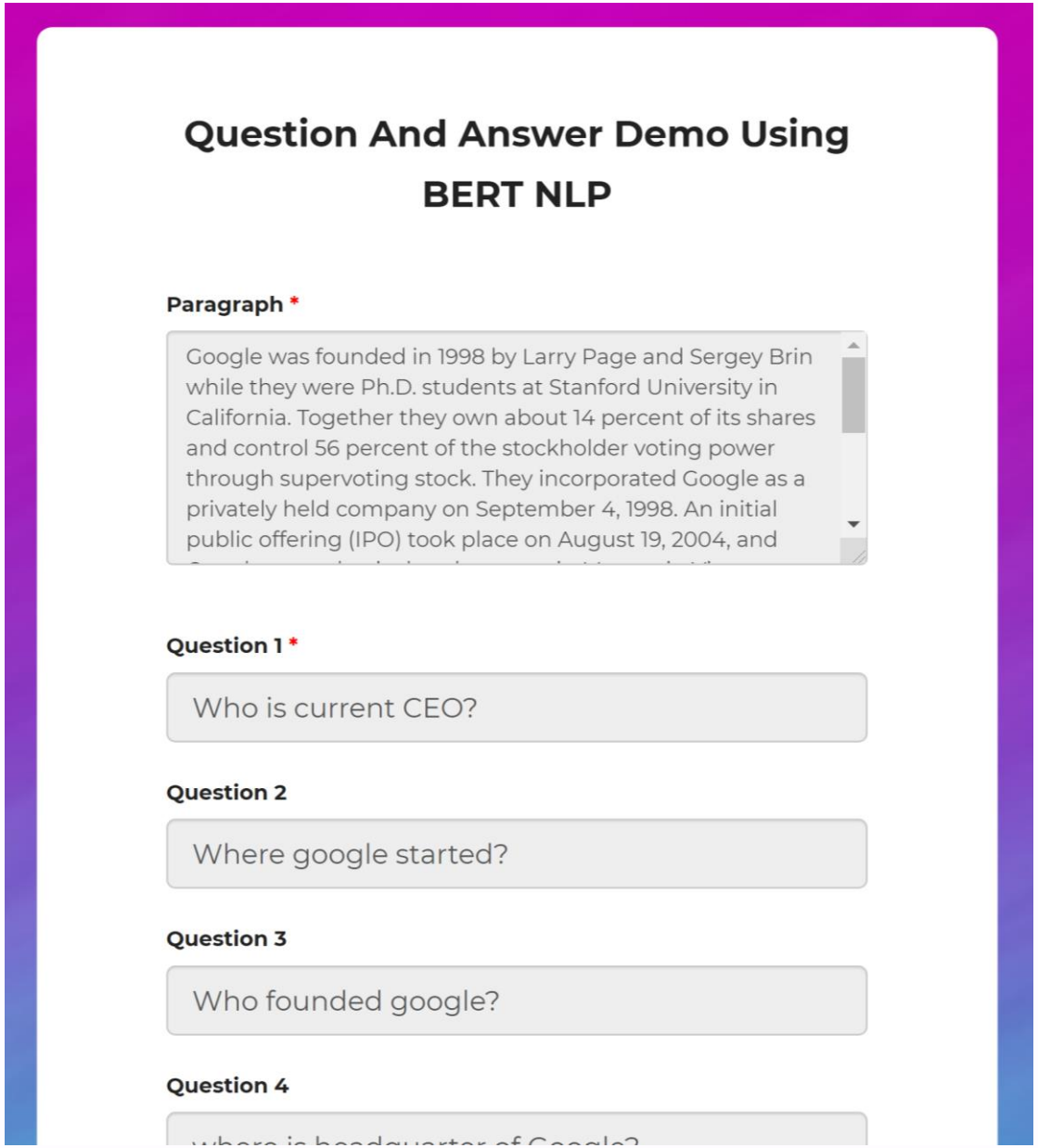
Translations: Chinese (Simplified), Persian

The year 2018 has been an inflection point for machine learning models handling text (or more accurately, Natural Language Processing or NLP for short). Our conceptual understanding of how best to represent words and sentences in a way that best captures underlying meanings and relationships is rapidly evolving. Moreover, the NLP community has been putting forward incredibly powerful components that you can freely download and use in your own models and pipelines (It's been referred to as NLP's ImageNet moment, referencing how years ago similar developments accelerated the development of machine learning in Computer Vision tasks).

BERT Q&A Demo

Get a preview of what Google *might* extract from a page for a given question with this BERT demo.

https://www.pragnakalp.com/demos/BERT-NLP-QnA-Demo/



By: Cyrus Shepard October 21st, 2014

More than Keywords: 7 Concepts of Advanced On-Page SEO

On-page SEO | Advanced SEO

The author's views are entirely his or her own (excluding the unlikely event of hypnosis) and may not always reflect the views of Moz.

255

160

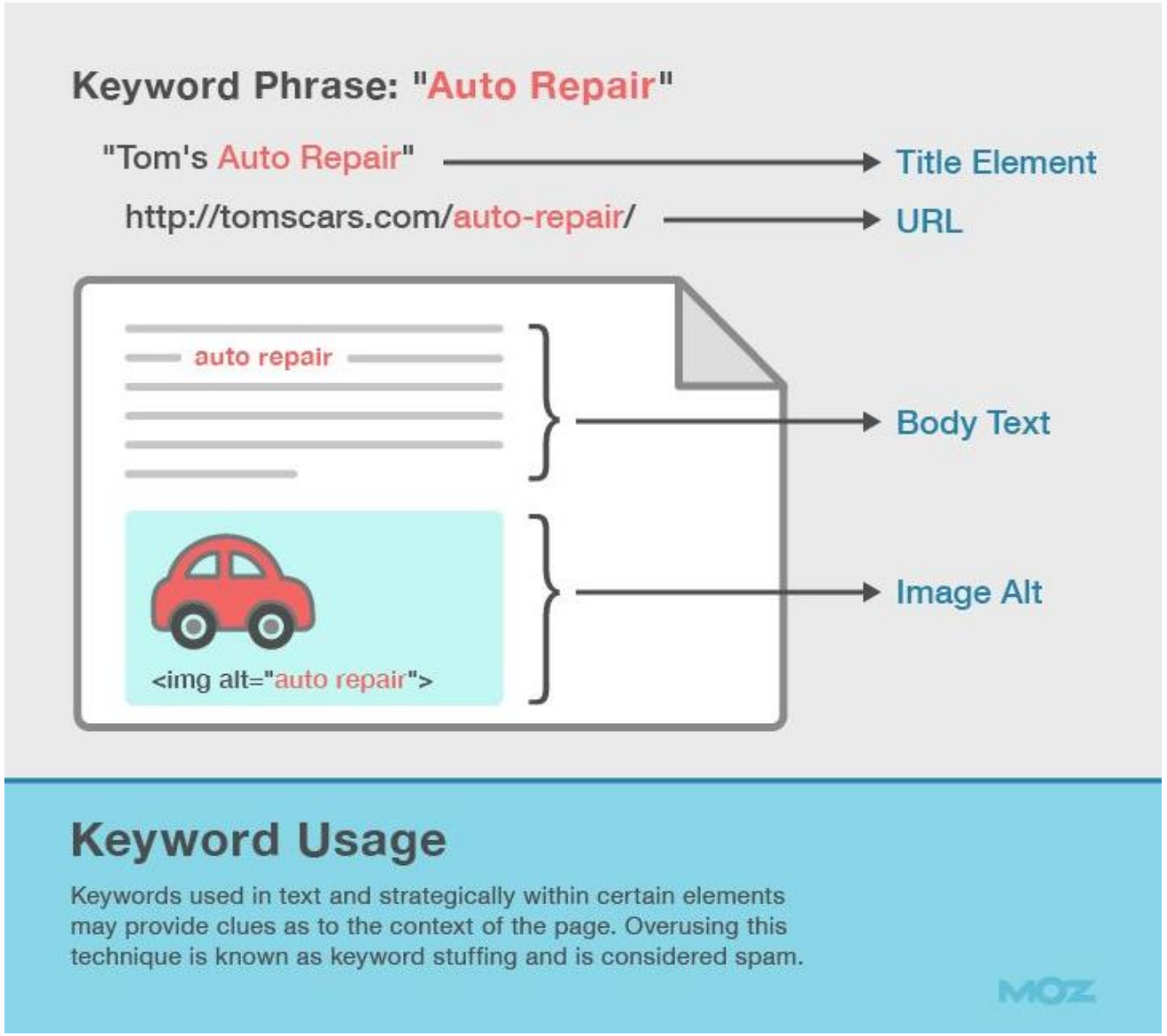
"What is this page about?"

As marketers, helping search engines answer that basic question is one of our most important tasks. Search engines can't read pages like humans can, so we incorporate **structure** and **clues** as to *what our content means*. This helps provide the **relevance** element of search engine optimization that

Keyword Usage

As you might imagine, leveraging Text Statistics to inform rankings indicates that you might want to use the keywords on the page.

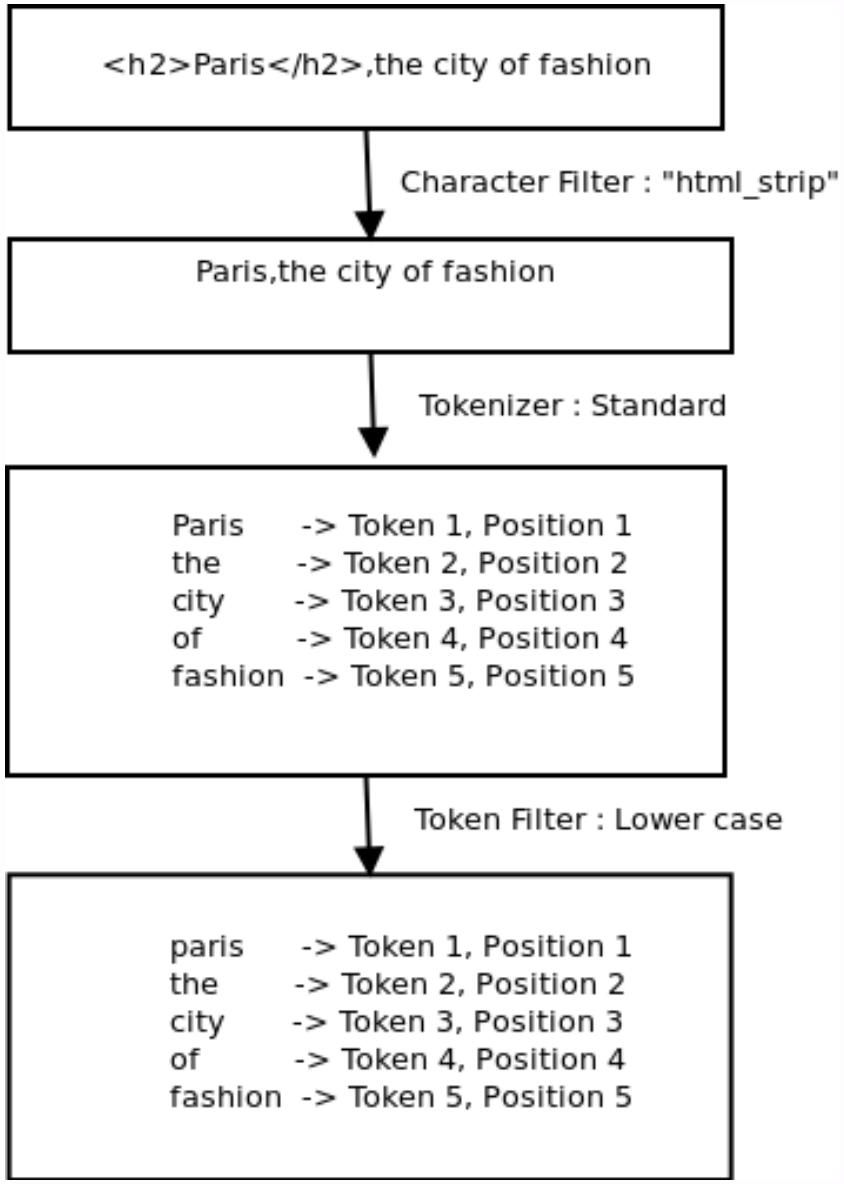
This is where having the opportunity to rank begins.



Tokenization

Search Engines break paragraphs into sentences and sentences into “tokens" or individual words.

This better positions content for statistical analysis



N-grams

N-grams are terms or phrases of N-length.

Full sentence	It does not, however, control whether an exaction is within Congress's power to tax.
Unigrams	"It"; "does"; "not,"; "however,"; "control"; "whether"; "an"; "exaction"; "is"; "within"; "Congress's"; "power"; "to"; "tax."
Bigrams	"It does"; "does not,"; "not, however,"; "however, control"; "control whether"; "whether an"; "an exaction"; "exaction is"; "is within"; "within Congress's"; "Congress's power"; "power to"; "to tax."
Trigrams	"It does not"; "does not, however"; "not, however, control"; "however, control whether"; "control whether an"; "whether an exaction"; "an exaction is"; "exaction is within"; "is within Congress's"; "within Congress's power"; "Congress's power to"; "power to tax."

Google AI Blog

The latest news from Google AI

All Our N-gram are Belong to You

Thursday, August 3, 2006

Posted by Alex Franz and Thorsten Brants, Google Machine Translation Team

Here at Google Research we have been using word n-gram models for a variety of R&D projects, such as statistical machine translation, speech recognition, spelling correction, entity detection, information extraction, and others. While such models have usually been estimated from training corpora containing at most a few billion words, we have been harnessing the vast power of Google's datacenters and distributed processing infrastructure to process larger and larger training corpora. We found that there's no data like more data, and scaled up the size of our data by one order of magnitude, and then another, and then one more - resulting in a training corpus of *one trillion words* from public Web pages.

We believe that the entire research community can benefit from access to such massive amounts of data. It will advance the state of the art, it will focus research in the promising direction of large-scale, data-driven approaches, and it will allow all research groups, no matter how large or small their computing resources, to play together. That's why we decided to share this enormous dataset with everyone. We processed 1,024,908,267,229 words of running text and are publishing the counts for all 1,176,470,663 five-word sequences that appear at least 40 times. There are 13,588,391 unique words, after discarding words that appear less than 200 times.

Watch for an announcement at the Linguistics Data Consortium (LDC), who will be distributing it soon, and then order your set of 6 DVDs. And let us hear from you - we're excited to hear what you will do with the data, and we're always interested in feedback about this dataset, or other potential datasets that might be useful for the research community.

Update (22 Sept. 2006): The LDC now has the data available in their catalog. The counts are as follows:

```
File sizes: approx. 24 GB compressed (gzip'ed) text files

Number of tokens:    1,024,908,267,229
Number of sentences:    95,119,665,584
Number of unigrams:         13,588,391
```

Google Books Ngram Viewer

Graph these comma-separated phrases: information retrieval case-insensitive

between 1800 and 2000 from the corpus English with smoothing of 3 . Search lots of books

Share Tweet

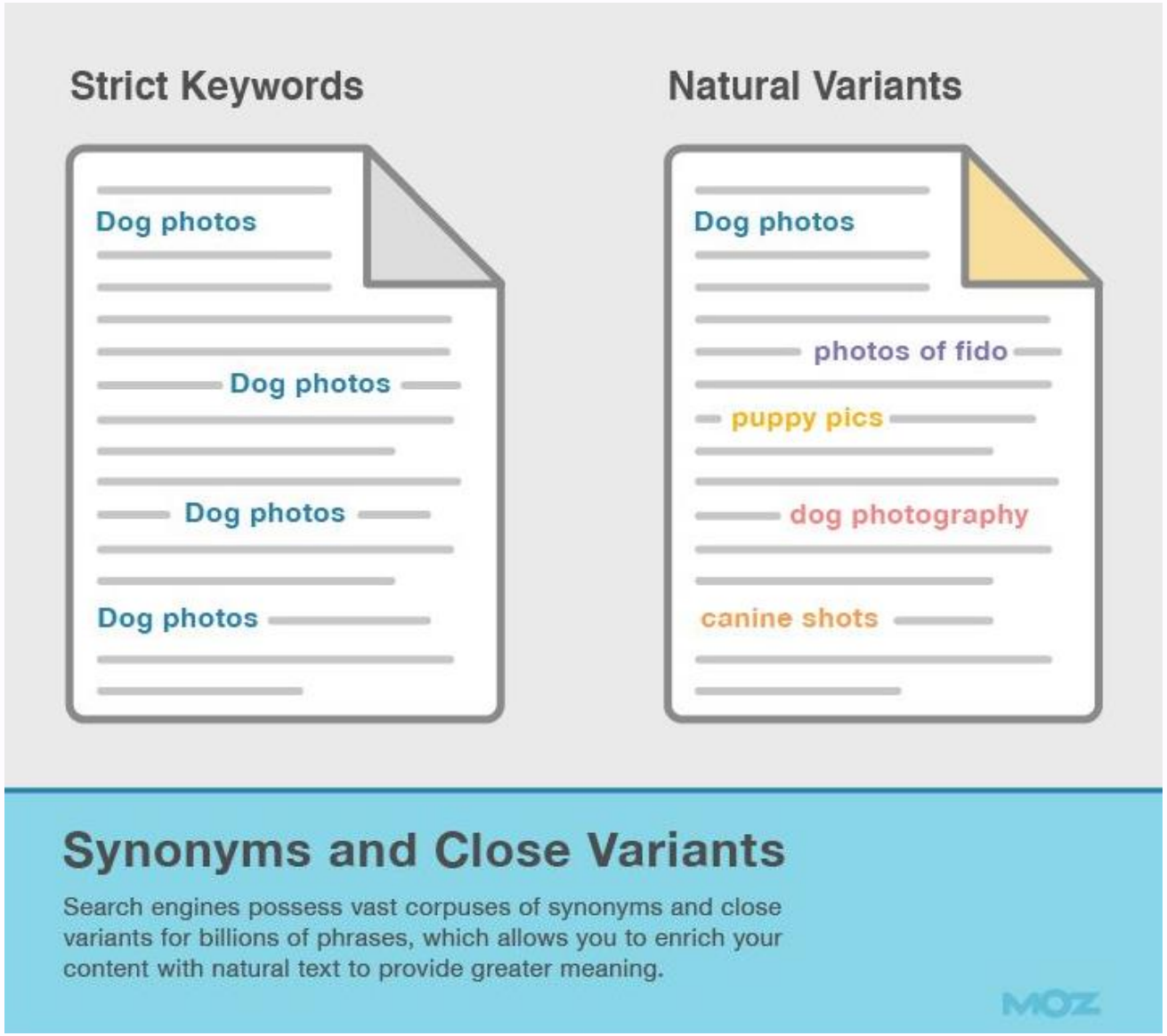
Embed Chart

Search in Google Books:

1800 - 1965 1966 - 1985 1986 - 1988 1989 - 1996 1997 - 2000 information retrieval English

Synonyms and Close Variants

Statistical relevance is computed using both synonyms and close variants of words.



Stemming

Search Engines break words into their stems to better determine relationships and relevance.

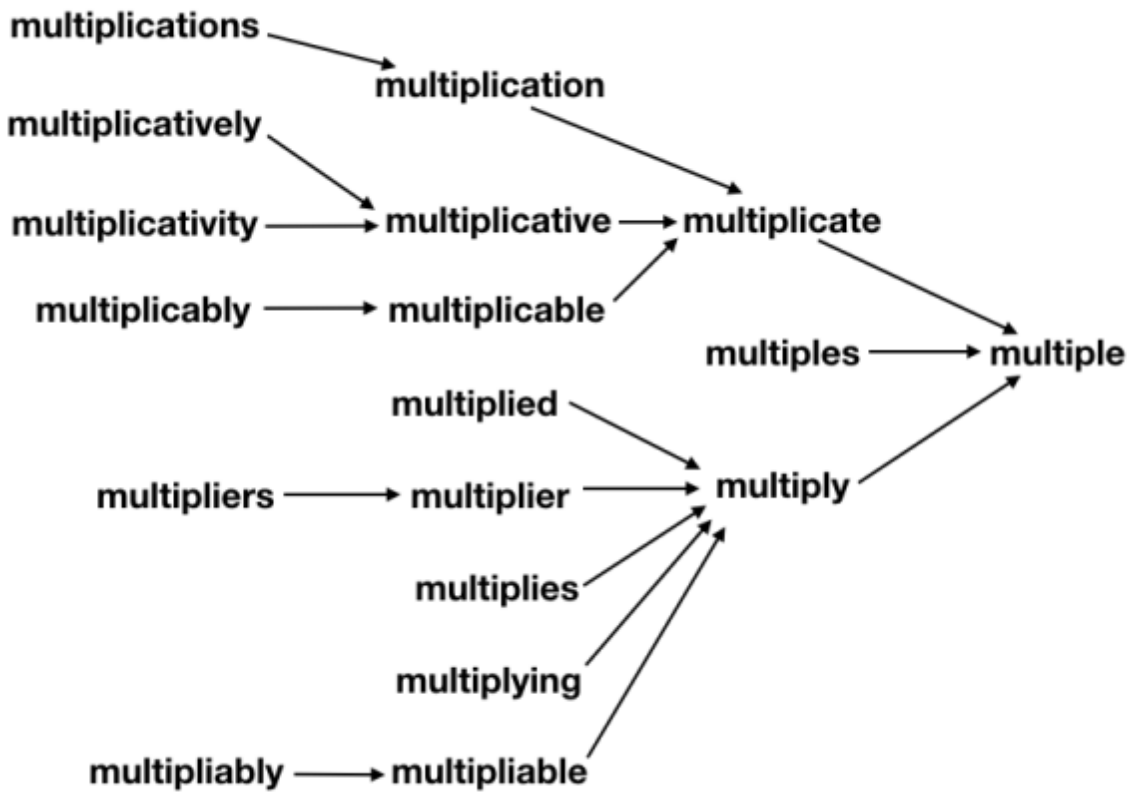
```
text <- "love loving lovingly loved lover lovely love"
text_tokens(text, stemmer = "en") # english stemmer
```

```
[[1]]
[1] "love"  "love"  "love"  "love"  "lover" "love"  "love"
```

Lemmatization

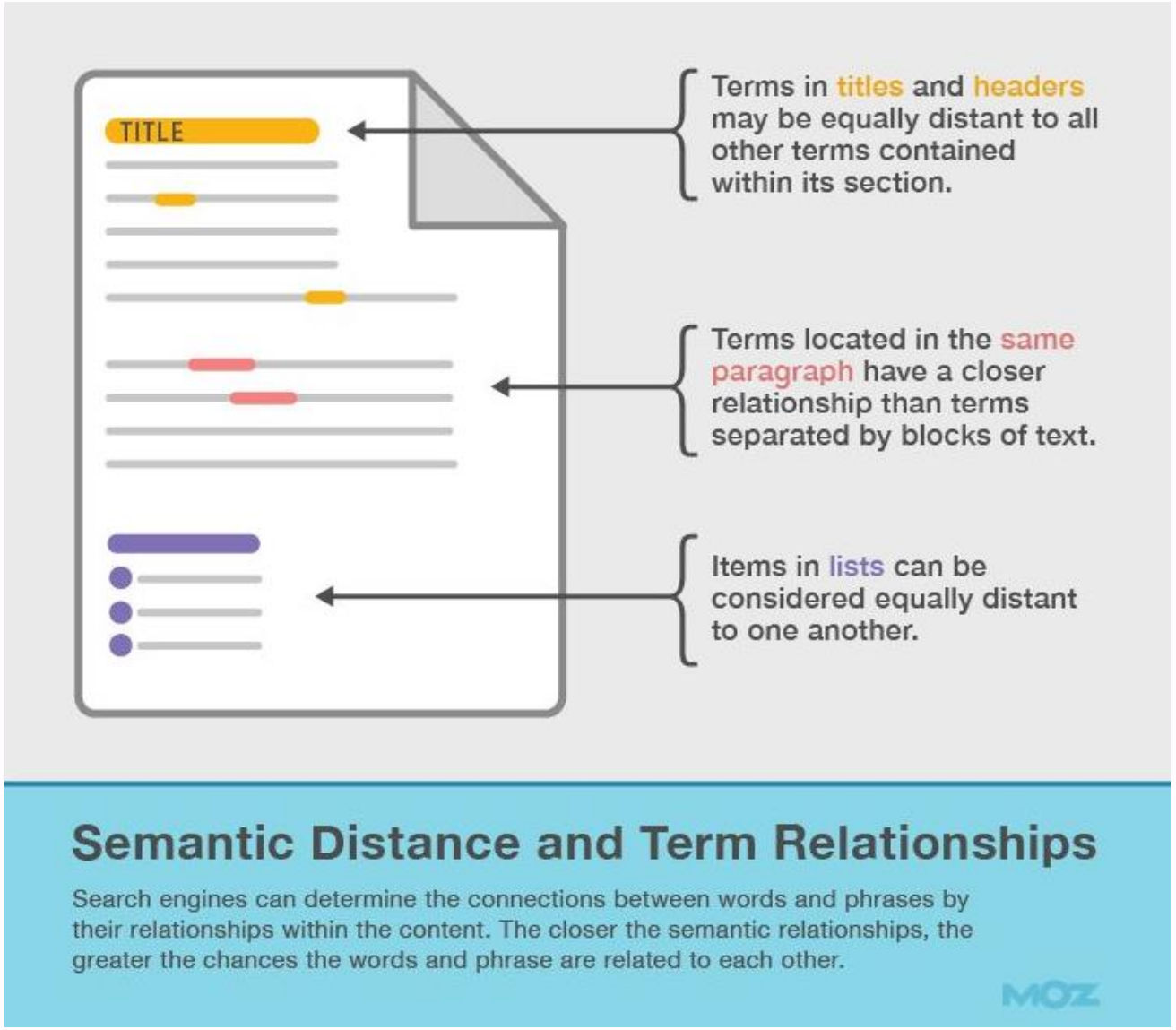
Similar to stemming, lemmatization is the grouping of inflected forms of the same word or idea.

For example, “better” and “good” have the same lemma.



Semantic Distance & Term Relationships

Search engines look for how physically close words are together to better determine their relationship in statistical models.



Latent Direchlet Allocation

LDA is a recursive algorithm for mapping keywords to topics.

A better way to think of it:

$$P(z = t|w) \quad \propto \quad (\alpha_t + n_{t|d}) \frac{\beta + n_{w|t}}{\beta V + n_{.|t}}$$

Chance of word is because of a topic

=

(Number of times the document already uses that topic a lot)

X

(Number of times that word has been in that topic)

SEOmoz

Document table with topics - 0:6 - Topic Extractor (Parallel LDA)

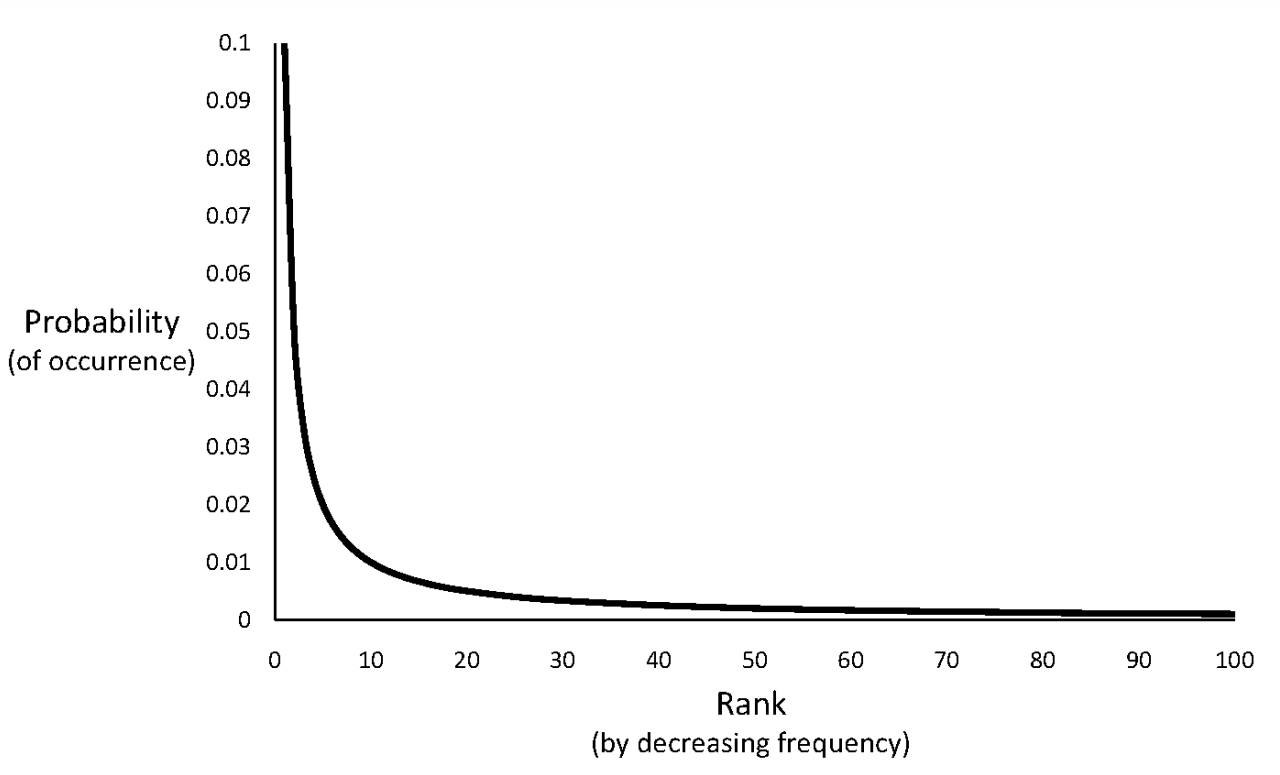
File Hilite Navigation View

Table "default" - Rows: 44834 | Spec - Columns: 13 | Properties | Flow Variables

Row ID	Document	Term	topic_0	topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	Assigne...
Row0	"SEO Tools: The Complete List (2019 Updat...	SEO[]	0.231	0	0	0	0	0	0.546	0.222	0	0	topic_6
Row1	"SEO Tools: The Complete List (2019 Updat...	Tools:[]	0.259	0	0	0	0	0	0.491	0.25	0	0	topic_6
Row2	"SEO Tools: The Complete List (2019 Updat...	The[]	0.056	0	0	0	0	0	0.722	0.222	0	0	topic_6
Row3	"SEO Tools: The Complete List (2019 Updat...	Complete[]	0.074	0	0	0	0	0	0.74	0.185	0	0	topic_6
Row4	"SEO Tools: The Complete List (2019 Updat...	List[]	0.139	0	0	0	0	0	0.694	0.167	0	0	topic_6
Row5	"SEO Tools: The Complete List (2019 Updat...	(2019[]	0.241	0	0	0	0	0	0.62	0.139	0	0	topic_6
Row6	"SEO Tools: The Complete List (2019 Updat...	Update)[]	0.157	0	0	0	0	0	0.666	0.176	0	0	topic_6
Row7	"SEO Tools: The Complete List (2019 Updat...	SECockpit[]	0.287	0	0	0	0	0	0.472	0.241	0	0	topic_6
Row8	"SEO Tools: The Complete List (2019 Updat...	is[]	0.204	0	0	0	0	0	0.639	0.157	0	0	topic_6
Row9	"SEO Tools: The Complete List (2019 Updat...	probably[]	0.213	0	0	0	0	0	0.676	0.111	0	0	topic_6
Row10	"SEO Tools: The Complete List (2019 Updat...	most[]	0.111	0	0	0	0	0	0.74	0.148	0	0	topic_6
Row11	"SEO Tools: The Complete List (2019 Updat...	advanced[]	0.194	0.009	0	0	0	0	0.518	0.278	0	0	topic_6
Row12	"SEO Tools: The Complete List (2019 Updat...	keyword[]	0.167	0	0	0	0	0	0.676	0.157	0	0	topic_6
Row13	"SEO Tools: The Complete List (2019 Updat...	research[]	0.093	0	0	0	0	0	0.629	0.278	0	0	topic_6
Row14	"SEO Tools: The Complete List (2019 Updat...	tool[]	0.222	0	0	0	0	0	0.713	0.065	0	0	topic_6
Row15	"SEO Tools: The Complete List (2019 Updat...	on[]	0.083	0	0	0	0	0	0.685	0.231	0	0	topic_6
Row16	"SEO Tools: The Complete List (2019 Updat...	the[]	0.148	0	0	0	0	0	0.602	0.25	0	0	topic_6
Row17	"SEO Tools: The Complete List (2019 Updat...	market.[]	0.342	0	0	0	0	0	0.565	0.093	0	0	topic_6
Row18	"SEO Tools: The Complete List (2019 Updat...	Yes,[]	0.241	0	0	0	0	0	0.583	0.176	0	0	topic_6
Row19	"SEO Tools: The Complete List (2019 Updat...	you[]	0.093	0	0	0	0	0	0.777	0.13	0	0	topic_6
Row20	"SEO Tools: The Complete List (2019 Updat...	enter[]	0.102	0	0	0	0	0	0.62	0.278	0	0	topic_6
Row21	"SEO Tools: The Complete List (2019 Updat...	a[]	0.102	0	0	0	0	0	0.648	0.25	0	0	topic_6
Row22	"SEO Tools: The Complete List (2019 Updat...	seed[]	0.241	0	0	0	0	0	0.611	0.148	0	0	topic_6
Row23	"SEO Tools: The Complete List (2019 Updat...	and[]	0.194	0	0	0	0	0	0.657	0.148	0	0	topic_6
Row24	"SEO Tools: The Complete List (2019 Updat...	get[]	0.241	0	0	0	0	0	0.537	0.222	0	0	topic_6
Row25	"SEO Tools: The Complete List (2019 Updat...	suggestions[]	0.148	0	0	0	0	0	0.685	0.167	0	0	topic_6
Row26	"SEO Tools: The Complete List (2019 Updat...	like[]	0.167	0	0	0	0	0	0.731	0.102	0	0	topic_6
Row27	"SEO Tools: The Complete List (2019 Updat...	any[]	0.046	0	0	0	0	0	0.611	0.343	0	0	topic_6
Row28	"SEO Tools: The Complete List (2019 Updat...	other[]	0.093	0	0	0	0	0	0.509	0.352	0.046	0	topic_6
Row29	"SEO Tools: The Complete List (2019 Updat...	tool.[]	0.194	0	0	0	0	0	0.602	0.204	0	0	topic_6
Row30	"SEO Tools: The Complete List (2019 Updat...	But[]	0.213	0	0	0	0	0	0.62	0.12	0.046	0	topic_6
Row31	"SEO Tools: The Complete List (2019 Updat...	can[]	0.241	0	0	0	0	0	0.648	0.111	0	0	topic_6
Row32	"SEO Tools: The Complete List (2019 Updat...	do[]	0.148	0	0	0	0	0	0.676	0.176	0	0	topic_6
Row33	"SEO Tools: The Complete List (2019 Updat...	so[]	0.213	0	0	0	0	0	0.657	0.13	0	0	topic_6
Row34	"SEO Tools: The Complete List (2019 Updat...	much[]	0.333	0	0	0	0	0	0.555	0.111	0	0	topic_6
Row35	"SEO Tools: The Complete List (2019 Updat...	more[]	0.148	0	0	0	0	0	0.759	0.093	0	0	topic_6
Row36	"SEO Tools: The Complete List (2019 Updat...	than[]	0.213	0	0	0	0	0	0.611	0.176	0	0	topic_6
Row37	"SEO Tools: The Complete List (2019 Updat...	list[]	0.065	0	0	0	0	0	0.759	0.176	0	0	topic_6

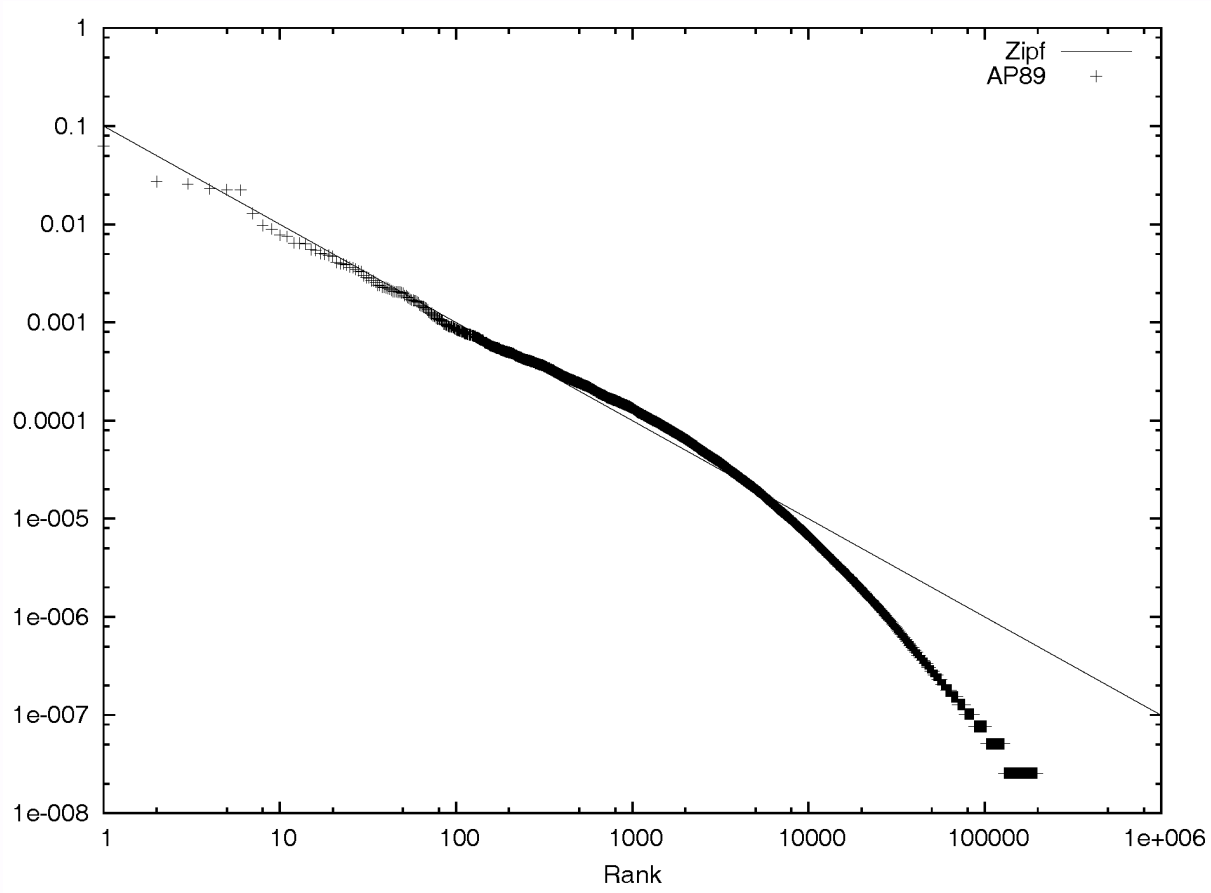
Zipf's Law

Zipf's law is a theory that words will be similarly distributed across documents in the same corpus (or document set).



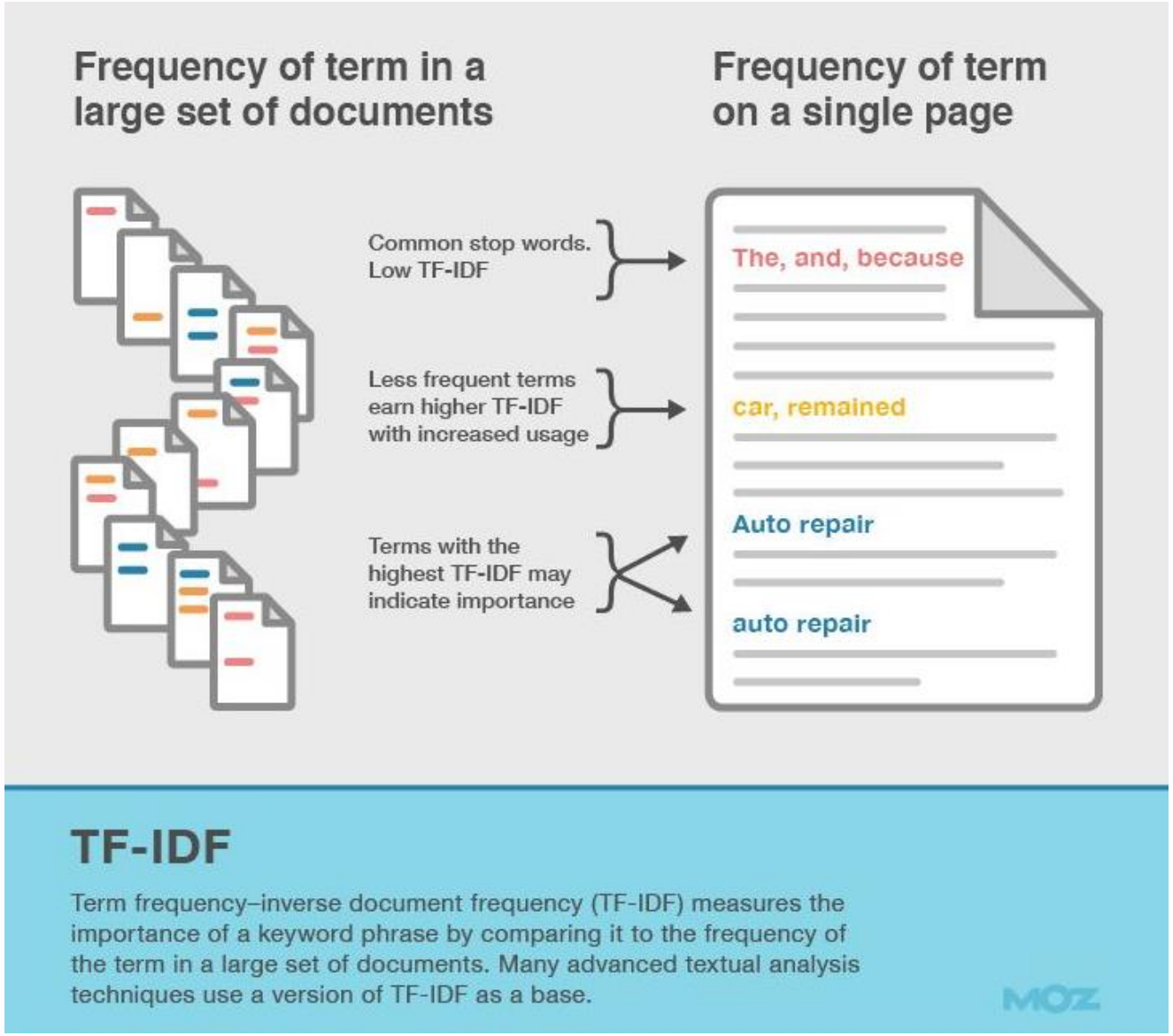
Zipf's Law Applied

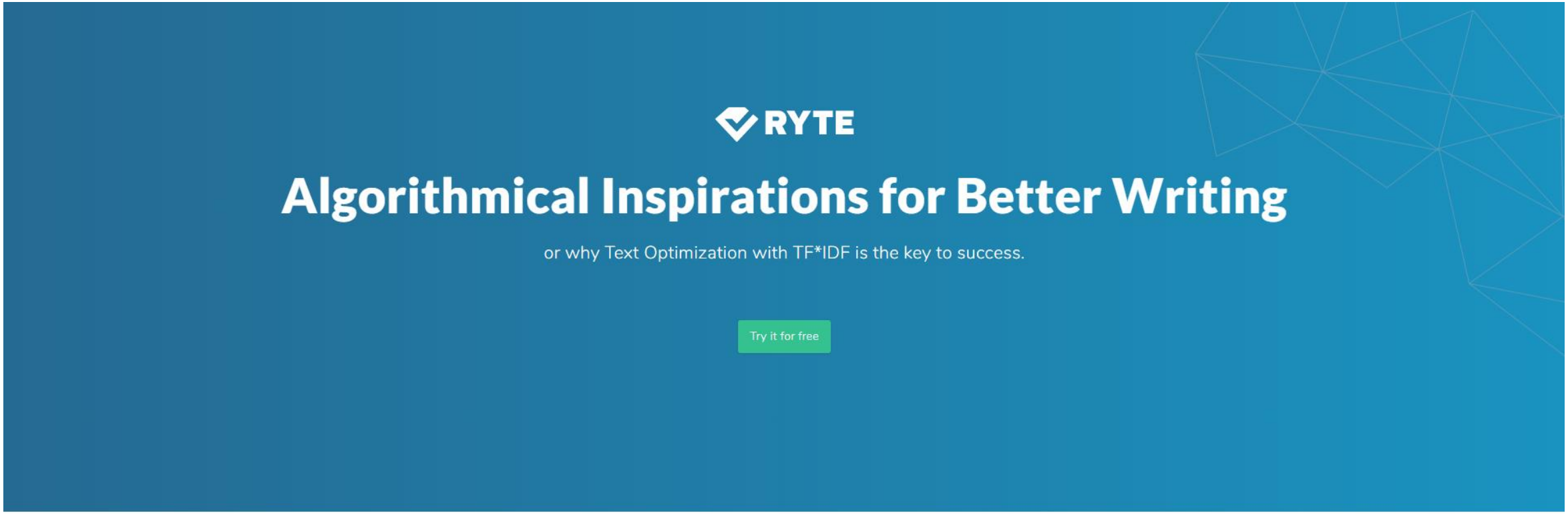
When run on an actual dataset, zipf's law tends to hold up in high rankings.



Term Frequency Inverse Document Frequency

Term frequency, inverse document frequency (TF-IDF) identifies the importance of keywords with respect to other keywords in documents in a corpus.



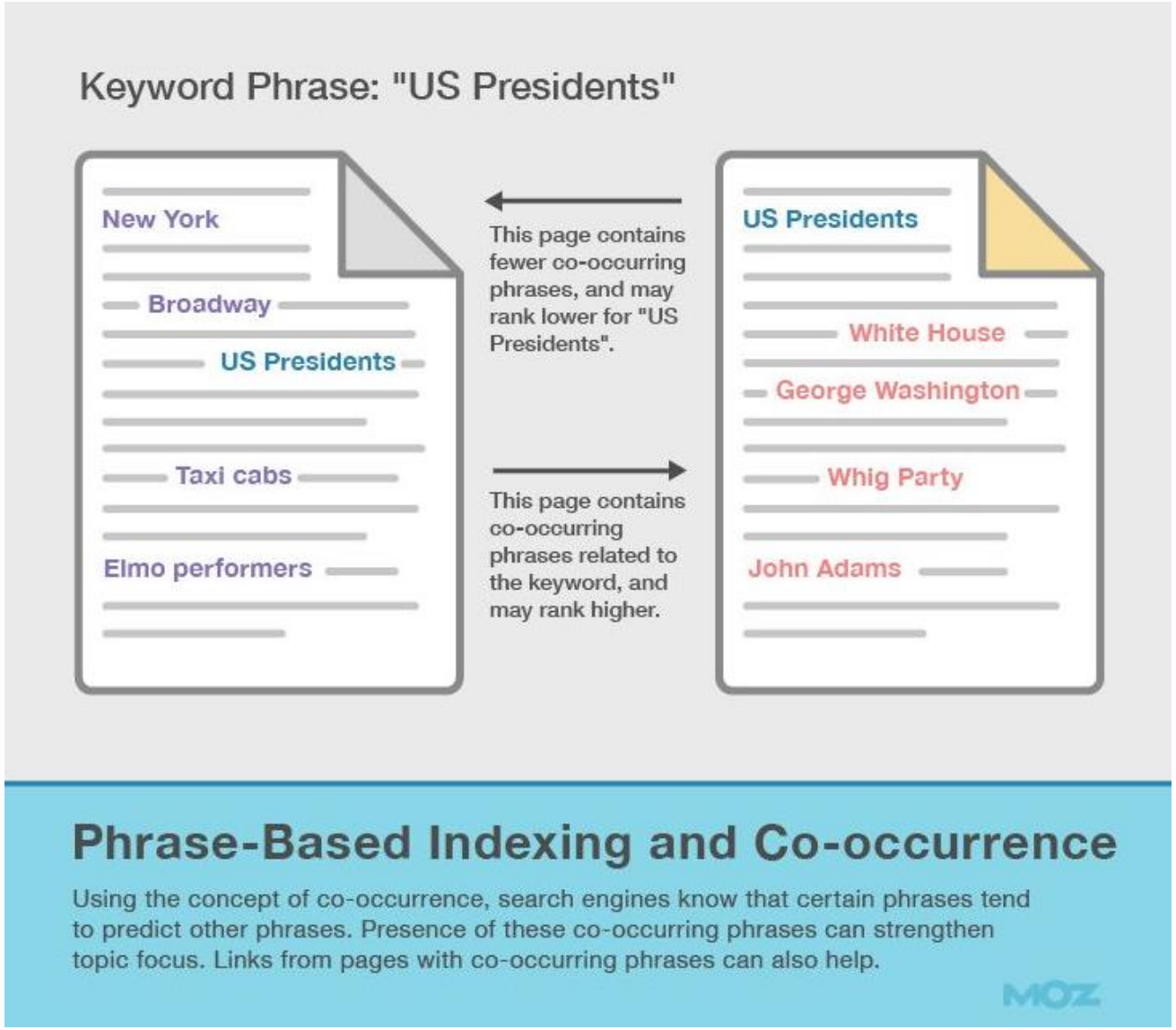


Search Engine Optimization has been around for many years. So far, however, our focus has been primarily on rankings and backlinks. We have monitored and exchanged links, bought links and built links by ourselves. But for years now, the search engines have been trying to convince us of a slightly different approach: "You need good and relevant content that is especially valuable for your users". And so we started to write content.

Phrase-based Indexing & Co-occurence

Google Specifically looks for keywords that co-occur with a target keyword in a document set.

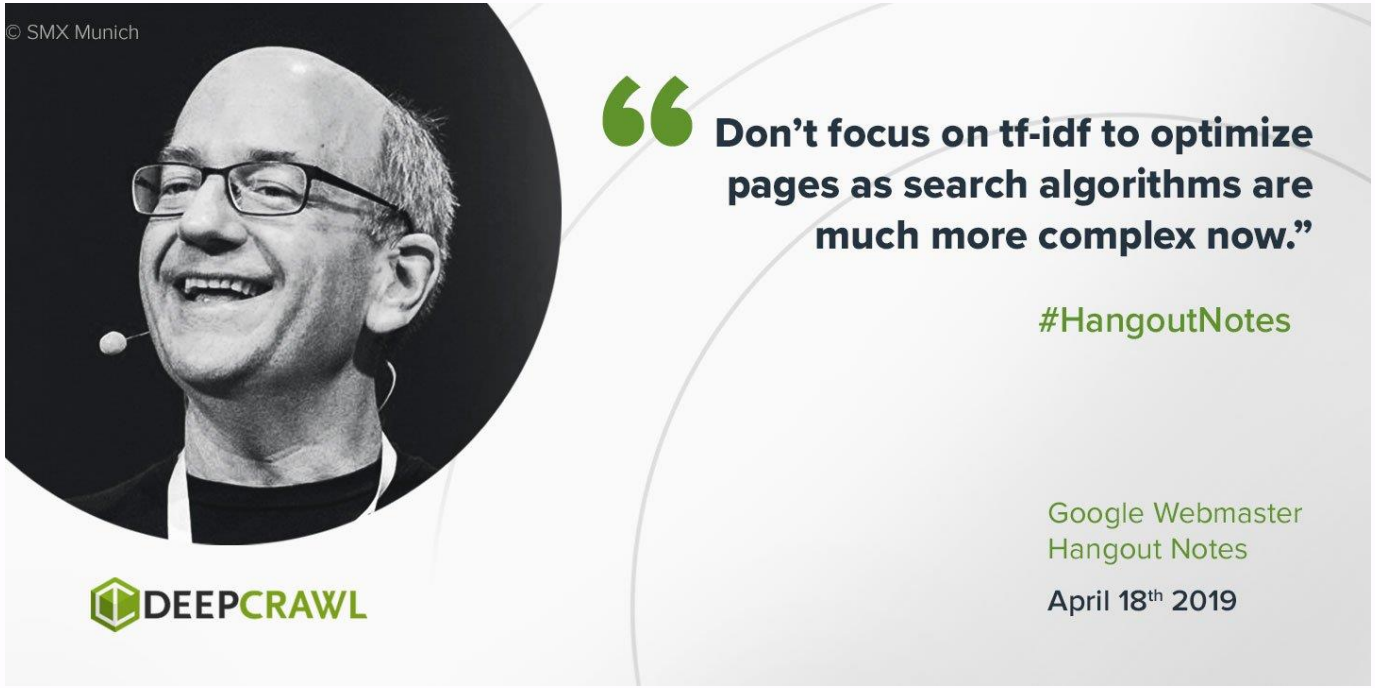
Usage of those keywords is an indication of relevance for subsequent documents.



But You Heard TF-IDF is Old And You Shouldn't Focus On it…

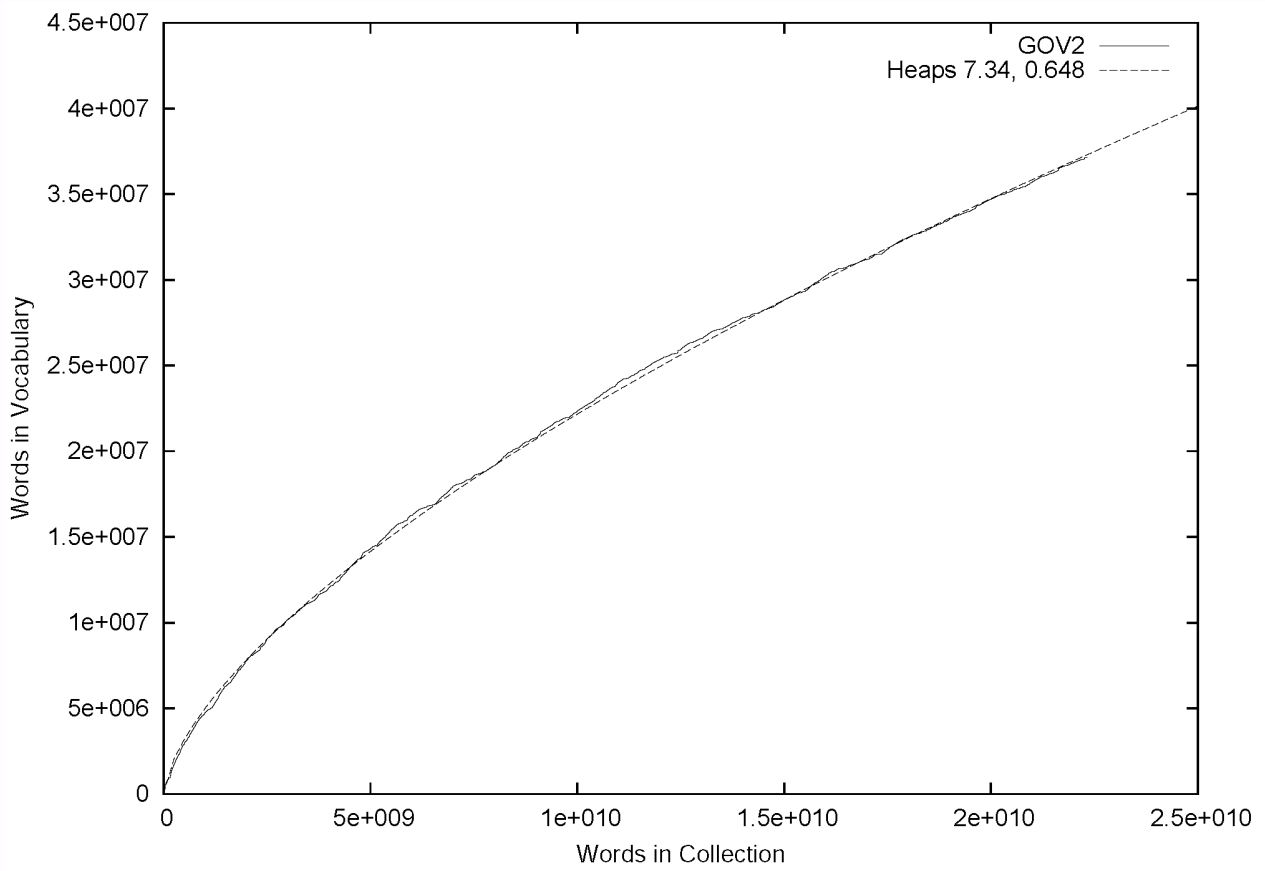
I actually agree with this, you should not focus on TF-IDF in isolation.

I suspect that Google is using a series of text analysis functions to calculate relevance, but **the age of the technique does not make it any less valuable.**



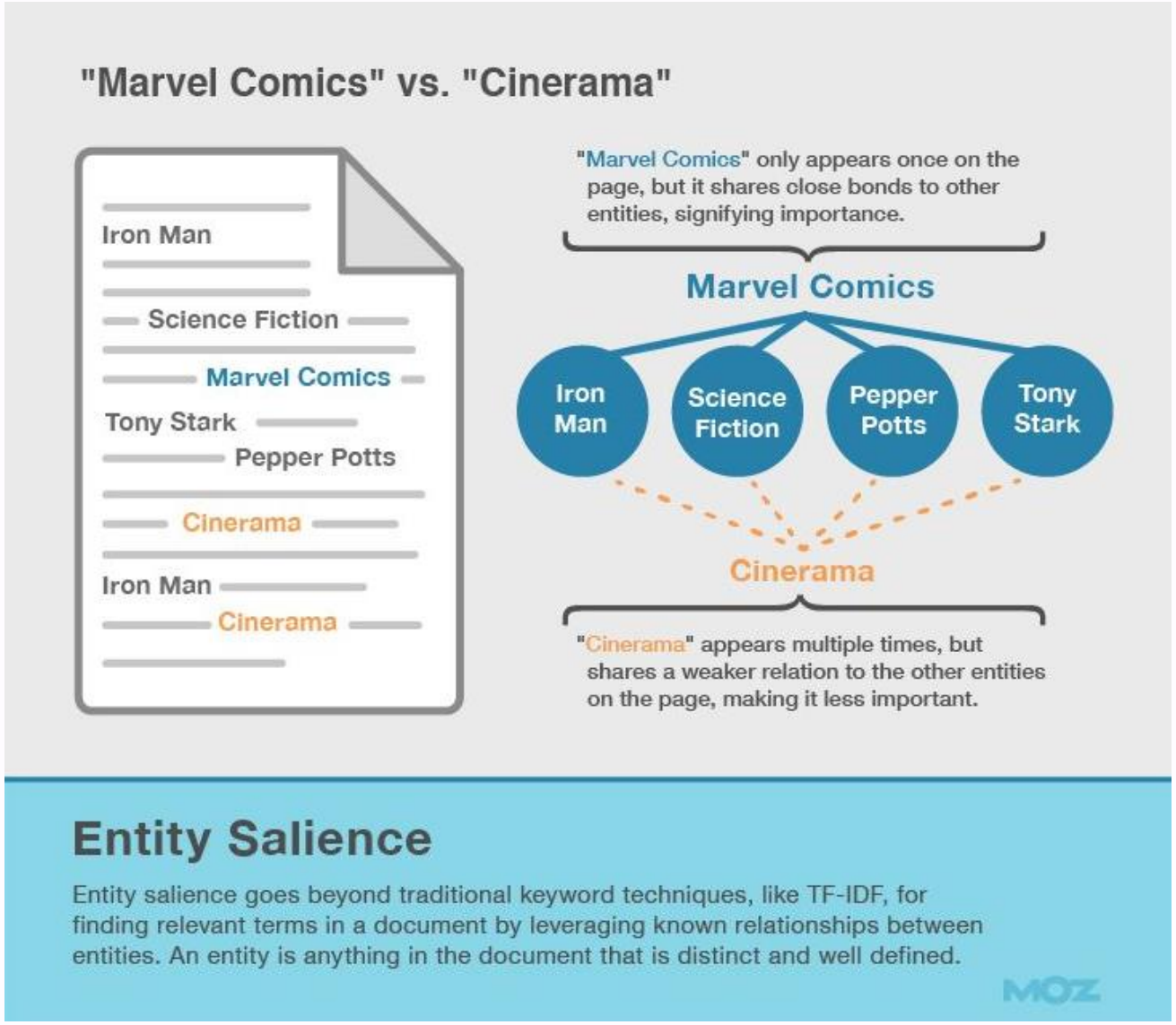
Heap's Law

Heaps law indicates that vocabulary within a corpus grows at a predictable rate.



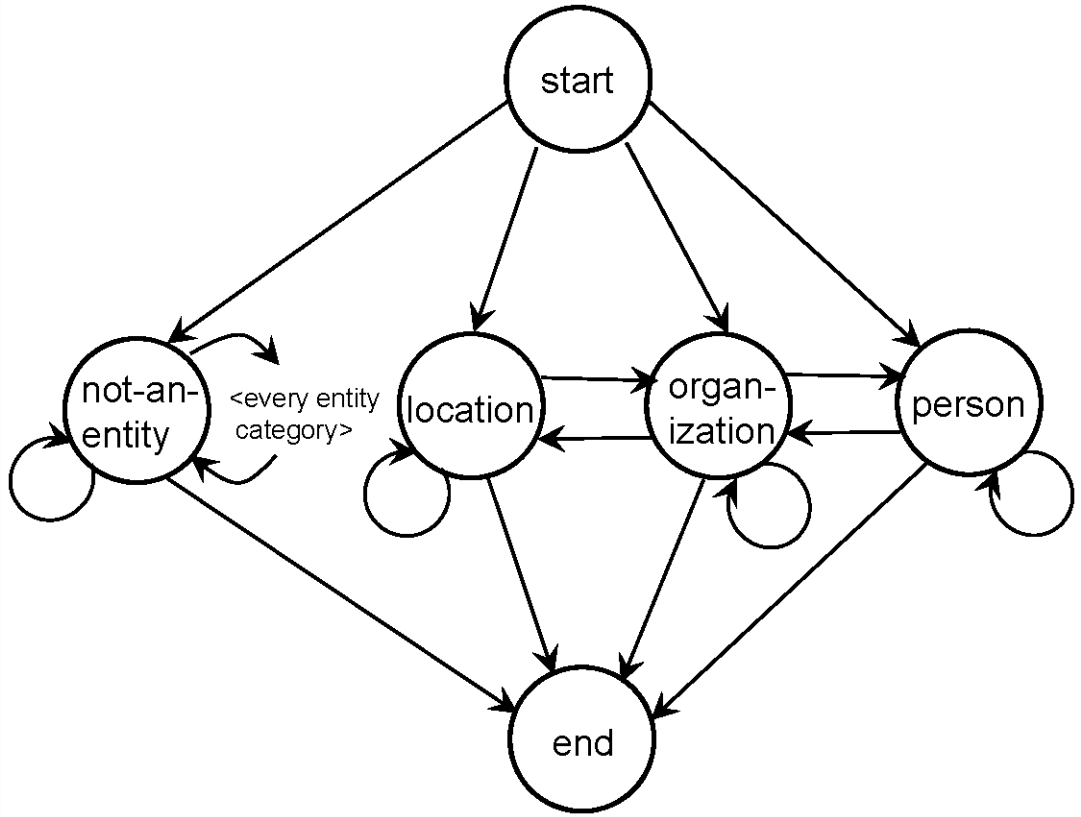
Entity Salience

Search engines leverage the features of the entity to understand the context in which the entity is referenced. Implicit features of the entity may be used to understand the document.



Hidden Markov Models

Hidden Markov Models allow search engines to extract implicit entities.



Entities are Also Used in Verification

Google double checks what you're talking about by using entities.

H/t @TomAnthonySEO

Knowledge-Based Trust: Estimating the Trustworthiness of Web Sources

Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang
Wilko Horn, Camillo Lugaresi, Shaohua Sun, Wei Zhang
Google Inc.
{lunadong|gabr|kpmurphy|vandang|wilko|camillol|sunsh|weizh}@google.com

ABSTRACT

The quality of web sources has been traditionally evaluated using *exogenous* signals such as the hyperlink structure of the graph. We propose a new approach that relies on *endogenous* signals, namely, the correctness of factual information provided by the source. A source that has few false facts is considered to be trustworthy.

The facts are automatically extracted from each source by information extraction methods commonly used to construct knowledge bases. We propose a way to distinguish errors made in the extraction process from factual errors in the web source per se, by using joint inference in a novel multi-layer probabilistic model.

We call the trustworthiness score we computed *Knowledge-Based Trust (KBT)*. On synthetic data, we show that our method can reliably compute the true trustworthiness levels of the sources. We then apply it to a database of 2.8B facts extracted from the web, and thereby estimate the trustworthiness of 119M webpages. Manual evaluation of a subset of the results confirms the effectiveness of the method.

1. INTRODUCTION

"Learning to trust is one of life's most difficult tasks."
– Isaac Watts.

Quality assessment for web sources[1] is of tremendous importance in web search. It has been traditionally evaluated using exogenous signals such as hyperlinks and browsing history. However, such signals mostly capture how popular a webpage is. For example, the gossip websites listed in [16] mostly have high PageRank scores [4], but would not generally be considered reliable. Conversely, some less popular websites nevertheless have very accurate information.

In this paper, we address the fundamental question of estimating how trustworthy a given web source is. Informally, we define the trustworthiness or *accuracy* of a web source as the probability that it contains the correct value for a fact (such as Barack Obama's nationality), assuming that it mentions any value for that fact. (Thus we do not penalize sources that have few facts, so long as they are correct.)

We propose using *Knowledge-Based Trust (KBT)* to estimate source trustworthiness as follows. We extract a plurality of facts from many pages using information extraction techniques. We then jointly estimate the correctness of these facts and the accuracy of the sources using inference in a probabilistic model. Inference is an iterative process, since we believe a source is accurate if its facts are correct, and we believe the facts are correct if they are extracted from an accurate source. We leverage the redundancy of information on the web to break the symmetry. Furthermore, we show how to initialize our estimate of the accuracy of sources based on authoritative information, in order to ensure that this iterative process converges to a good solution.

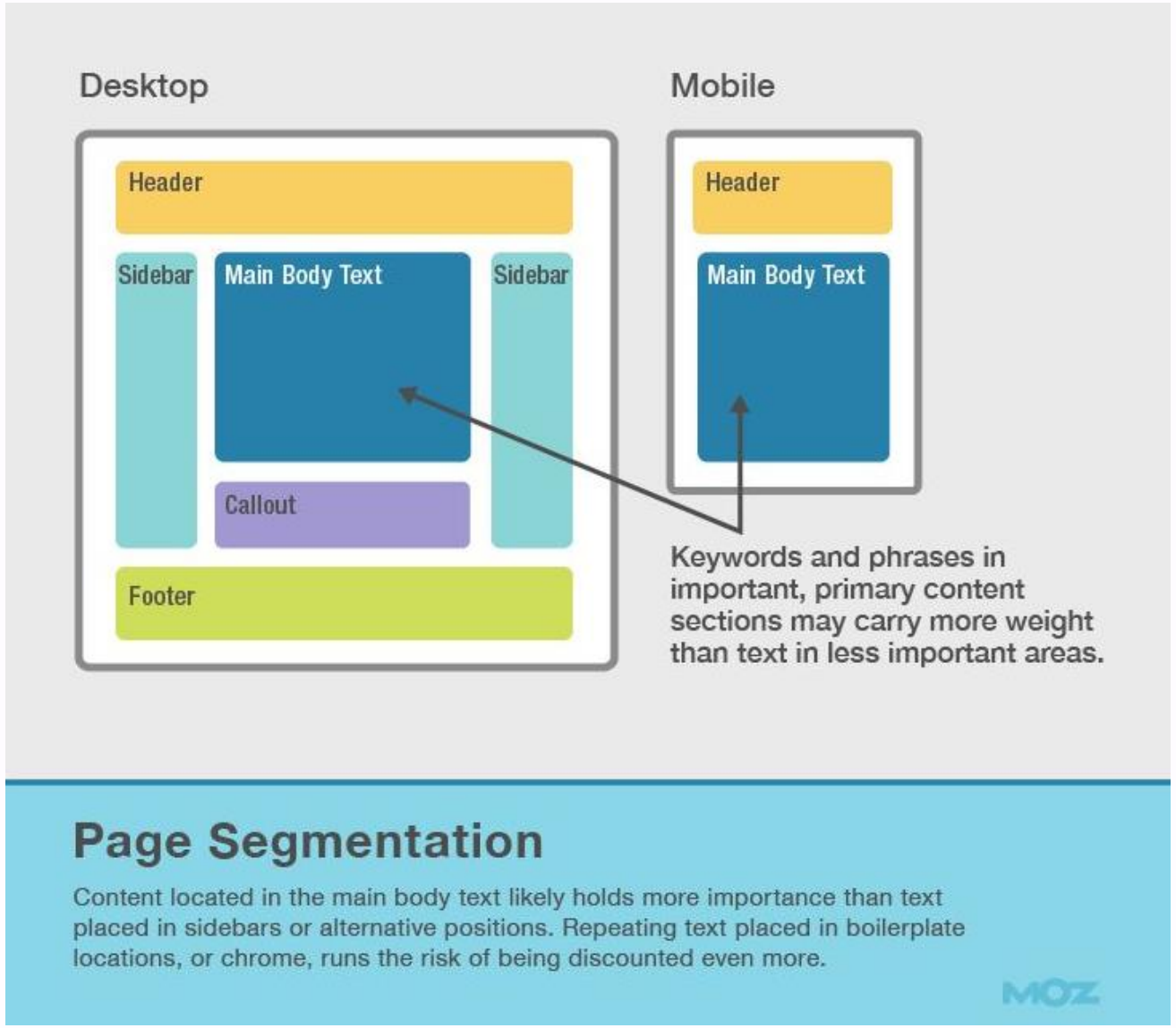
The fact extraction process we use is based on the *Knowledge Vault* (KV) project [10]. KV uses 16 different information extraction systems to extract (subject, predicate, object) *knowledge triples* from webpages. An example of such a triple is *(Barack Obama, nationality, USA)*. A subject represents a real-world entity, identified by an ID such as *mid*s in *Freebase* [2]; a predicate is predefined in *Freebase*, describing a particular attribute of an entity; an object can be an entity, a string, a numerical value, or a date.

The facts extracted by automatic methods such as KV may be wrong. One method for estimating if they are correct or not was described in [11]. However, this earlier work did not distinguish between factual errors on the page and errors made by the extraction system. As shown in [11], extraction errors are far more prevalent than source errors. Ignoring this distinction can cause us to incorrectly distrust a website.

Another problem with the approach used in [11] is that it estimates the reliability of each webpage independently. This can cause problems when data are sparse. For example, for more than one billion webpages, KV is only able to extract a single triple (other extraction systems have similar limitations). This makes it difficult to reliably estimate the trustworthiness of such sources

Page Segmentation

Modern Search engines determine both prominence of content and review the visual hierarchy to determine how valuable the content is to the page.



How it all works.

Search engines are recursive.

What we're missing.

We need to understand Google's statistical expectation of content and add it to our content's requirements.

What Should We Be Doing?

Let's Get Actionable

Step 1 - Research & Create:

Create content that fulfills expectations

We need to create or optimize content that balances user needs and search engine expectations.

This requires researching the query space based on what is currently ranking in Google's corpus.

Step 2 - Build Relevant Links:

Build Links with Content Parity

Building links needs to be a more detailed process. You need to only queue up link prospects that are topically relevant.

These are the links that will propel you further than links from random unrelated content.

Step 3 - Test & Optimize:

Measure performance and experiment with metadata

You need to continually review the performance of your content against that of the average performance in your space.

If you are performing below average, this an opportunity to perform experiments on your metadata.

The "Perfectly" Optimized Page

(for the example keyword phrase "chocolate donuts")

Page Title: Chocolate Donuts | Mary's Bakery

Meta Description: Mary's Bakery's chocolate donuts are possibly the most delicious, perfectly formed, flawlessly chocolately donuts ever made.

H1 Headline:
Chocolate Donuts from Mary's Bakery

Image Filename:
chocolate-donuts.jpg

Photo of Donuts (with Alt Attribute): Chocolate Donuts

Body Text:

chocolate donuts

donuts

chocolate donuts

donuts

chocolate

chocolate donuts

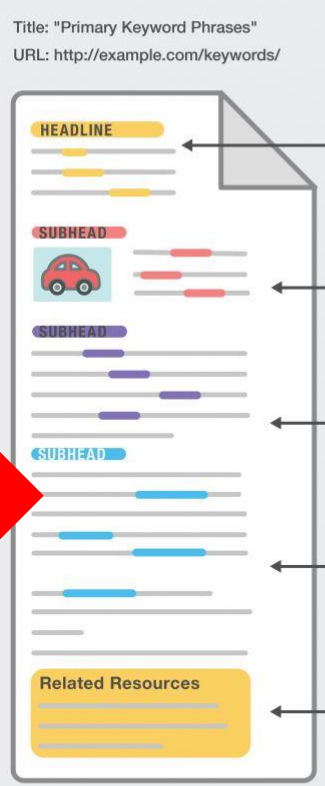
chocolate

chocolate donuts

Page URL: http://marysbakery.com/chocolate-donuts

On-page Topic Targeting for SEO

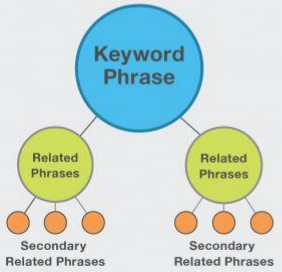
When creating content, we organize **keyword phrases** and **ideas** with **related concepts** in order to provide search engines with *better topical signals*, with the goal to achive higher rankings.



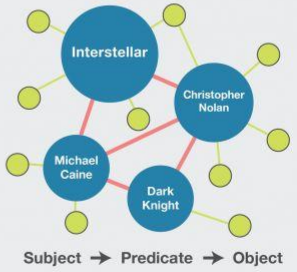
Related Keyword Phrases

1. **Location**
   Keywords placed in important areas like **titles**, **headlines**, and **primary content** sections
2. **Frequency**
   Important **phrases** and their **variants** may appear more frequently than average
3. **Distance**
   Related phrases arranged **close together** or connected by **HTML elements**
4. **Links**
   Use topically relevant **outlinks** and **citations** to high-quality resources
5. **Supplemental Content**
   Resources which may contain information or **links** to other pages of your site.

6. **Keywords and Related Phrases**



7. **Entities and Schema**



Source: http://moz.com/blog/on-page-topic-seo

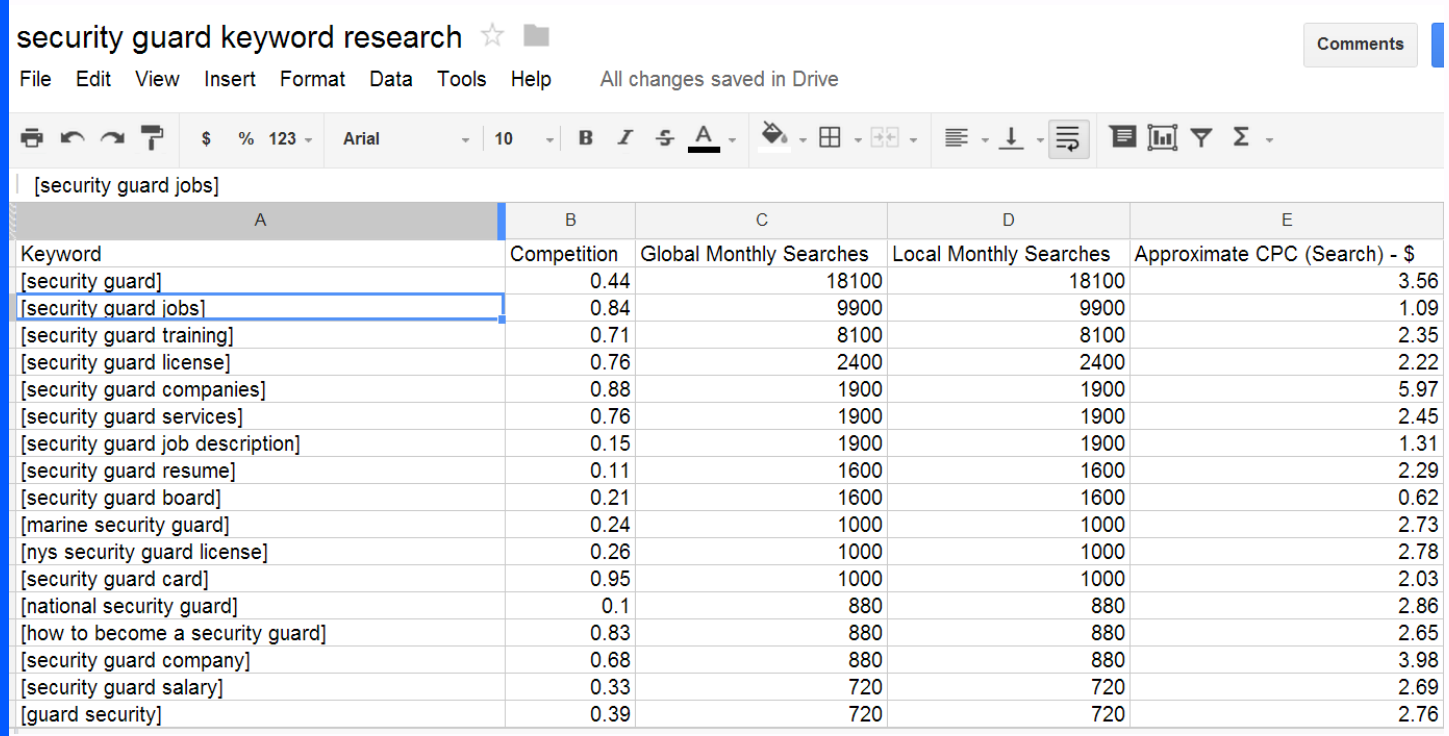
MOZ

We need to do more of Column B

Our industry talks a lot about the concepts in the column B, but in practice we do not have enough SEO tools to facilitate the latter.

It Starts with Keyword Research

This cannot be the extent of your keyword research in 2019.



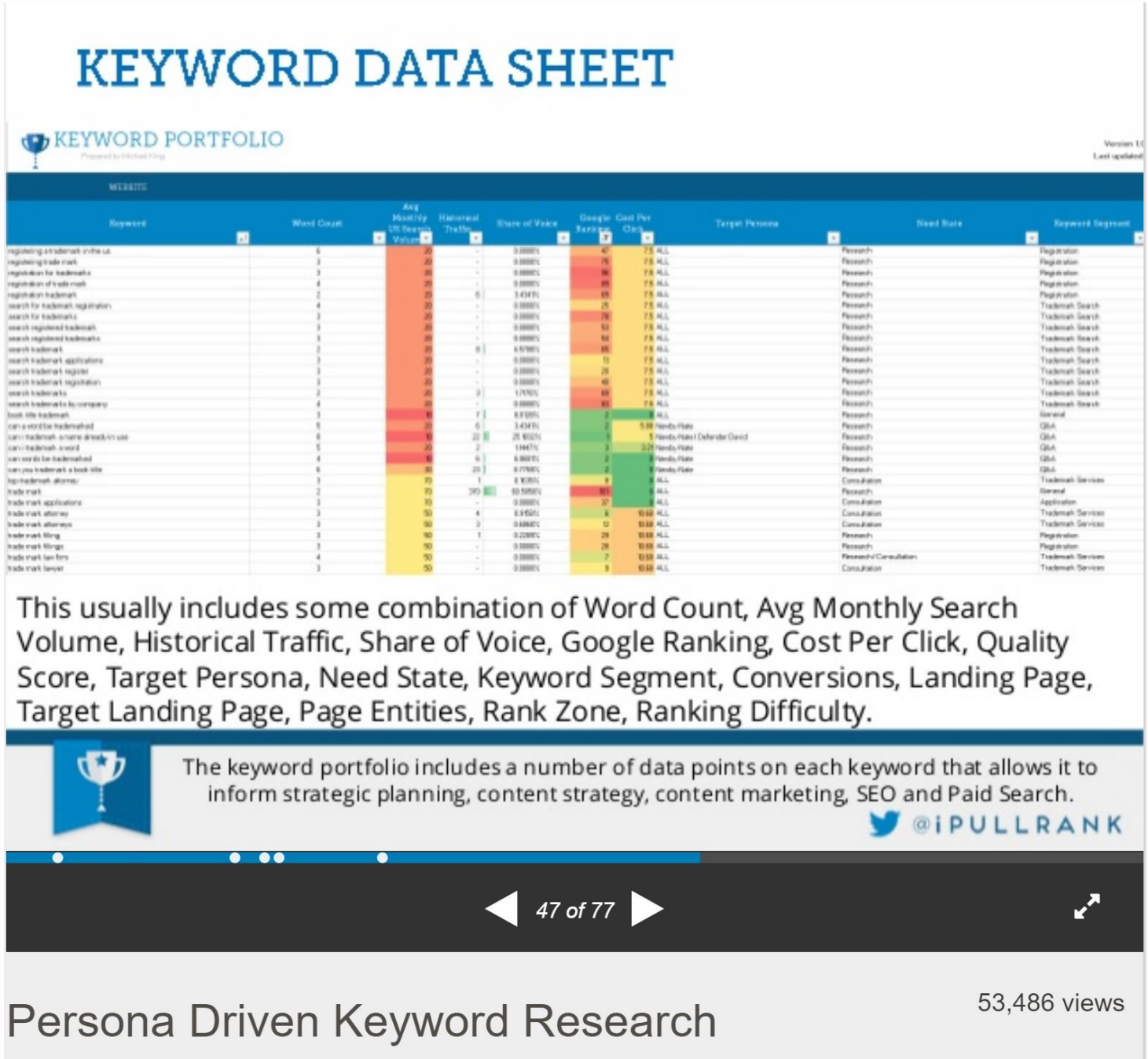
security guard keyword research

Comments

File Edit View Insert Format Data Tools Help All changes saved in Drive

[security guard jobs]

	A	B	C	D	E
	Keyword	Competition	Global Monthly Searches	Local Monthly Searches	Approximate CPC (Search) - $
	[security guard]	0.44	18100	18100	3.56
	[security guard jobs]	0.84	9900	9900	1.09
	[security guard training]	0.71	8100	8100	2.35
	[security guard license]	0.76	2400	2400	2.22
	[security guard companies]	0.88	1900	1900	5.97
	[security guard services]	0.76	1900	1900	2.45
	[security guard job description]	0.15	1900	1900	1.31
	[security guard resume]	0.11	1600	1600	2.29
	[security guard board]	0.21	1600	1600	0.62
	[marine security guard]	0.24	1000	1000	2.73
	[nys security guard license]	0.26	1000	1000	2.78
	[security guard card]	0.95	1000	1000	2.03
	[national security guard]	0.1	880	880	2.86
	[how to become a security guard]	0.83	880	880	2.65
	[security guard company]	0.68	880	880	3.98
	[security guard salary]	0.33	720	720	2.69
	[guard security]	0.39	720	720	2.76



Let's Take it back to 2013

At minimum you should be pulling, keywords, search volume, ranking difficulty, co-occurring keywords, target personas, need states, entities.

Check out my Persona-driven Keyword research deck for how to pull most of this.

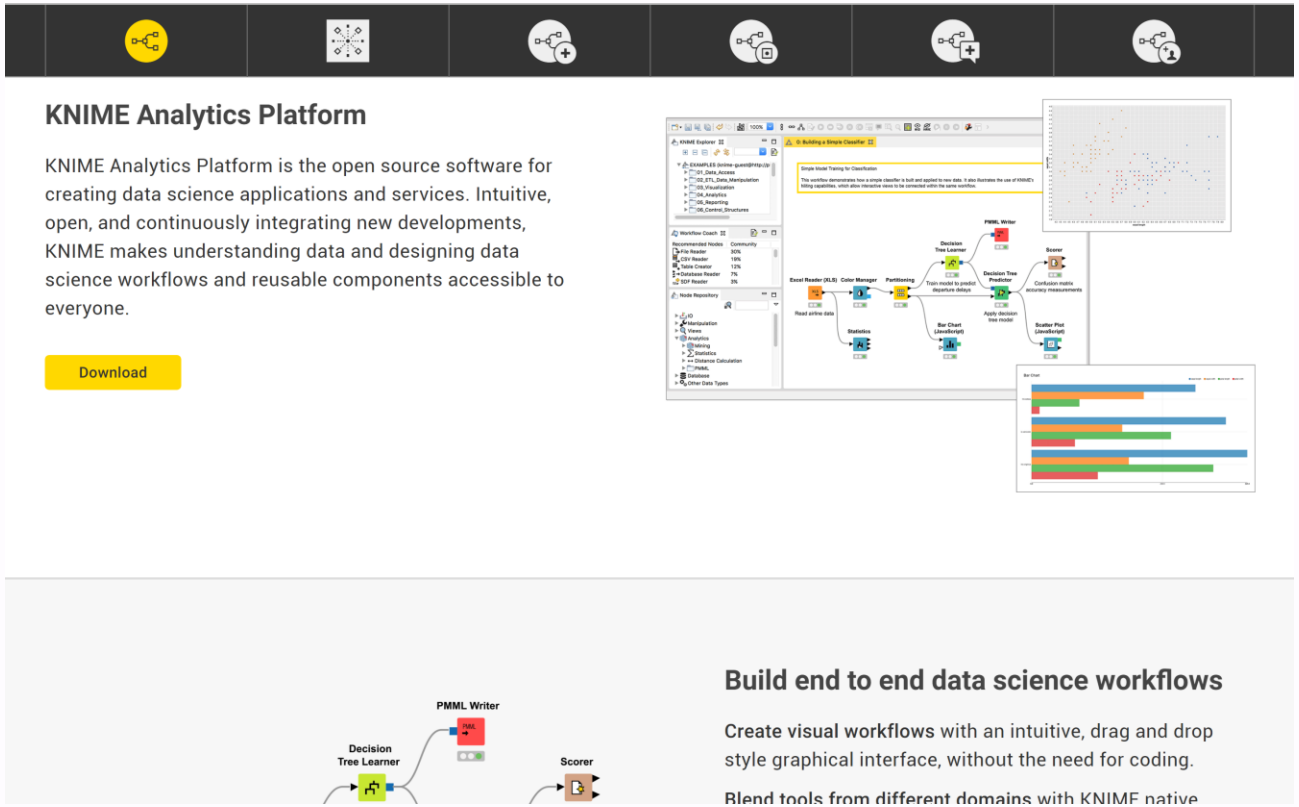
https://www.slideshare.net/ipullrank/persona-driven-keyword-research/

The goal.

The ultimate goal is to determine a series of co-occurring terms and entities to incorporate into your content for each target keyword.

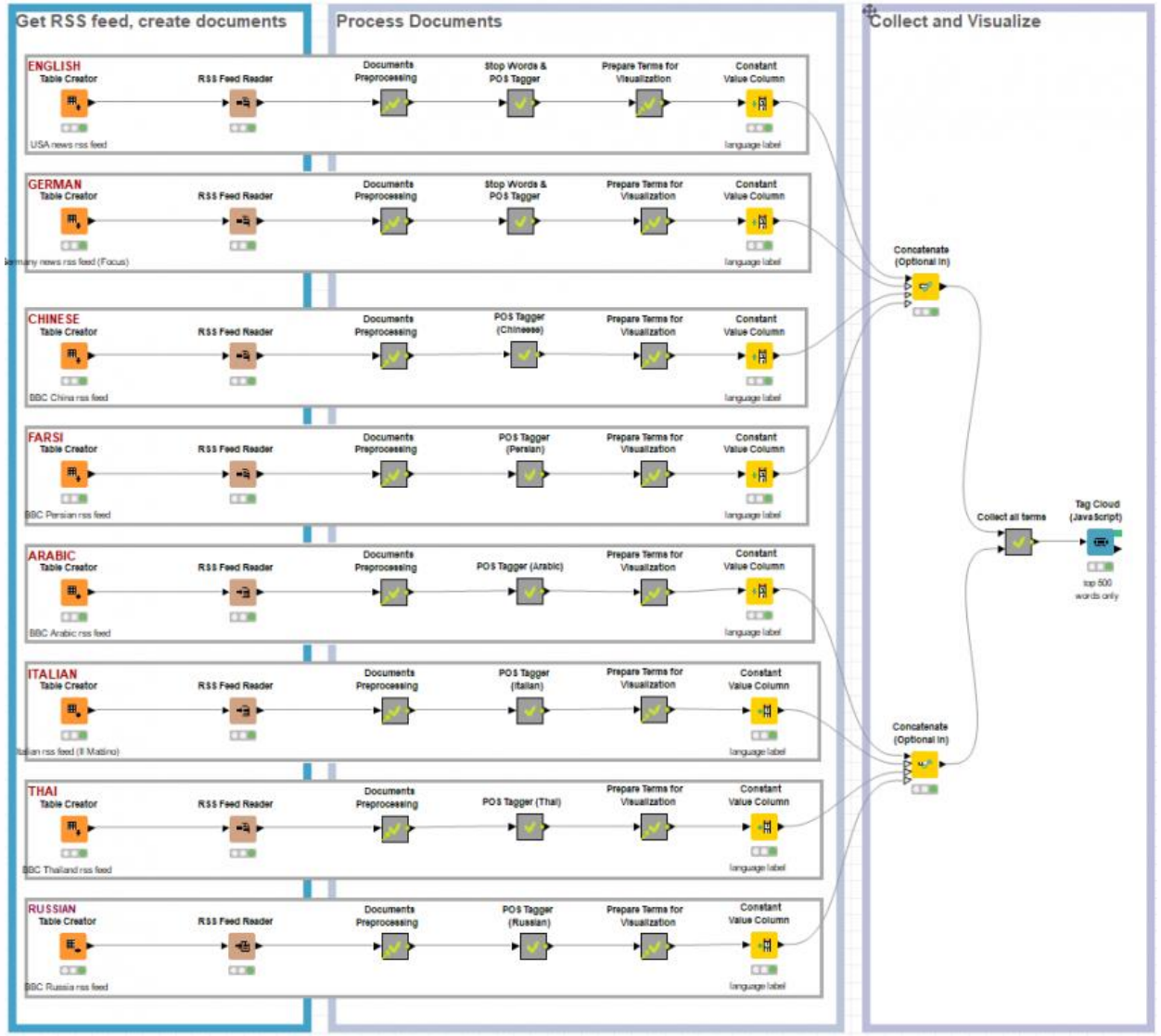
Meet Knime Analytics

https://www.knime.com/knime-software/knime-analytics-platform



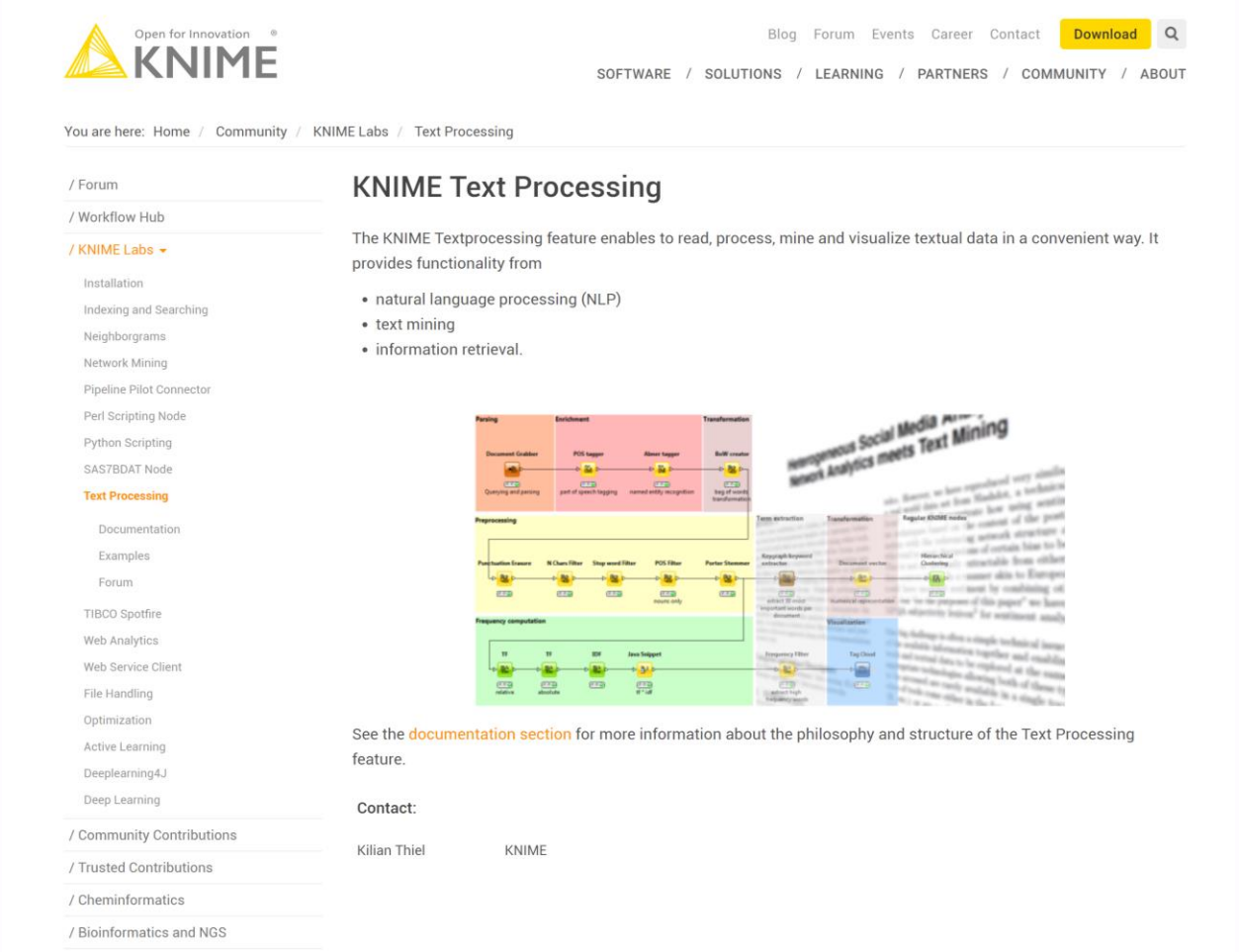
Knime Allows for Drag & Drop Text Analysis

More than just text analysis, Knime has a series of features for.



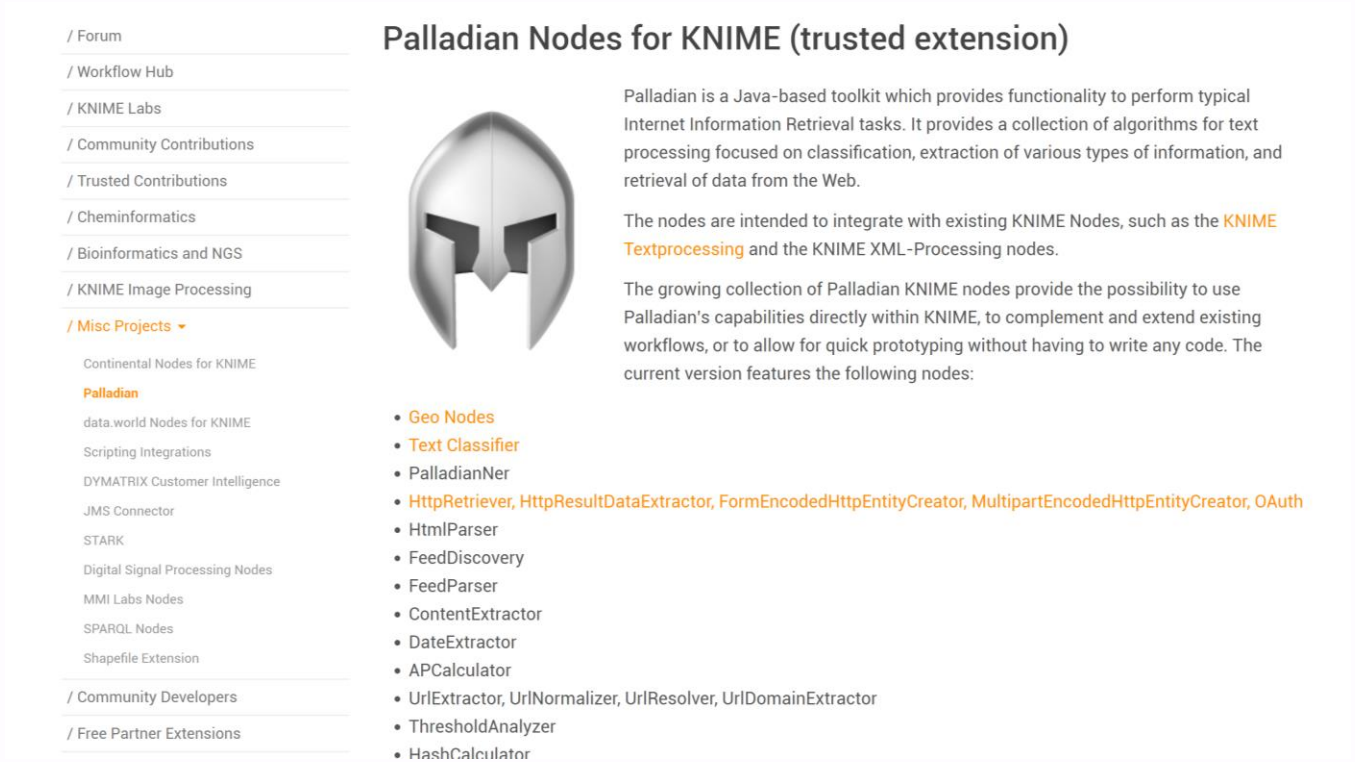
TextProcessing Extension

https://www.knime.com/knime-text-processing



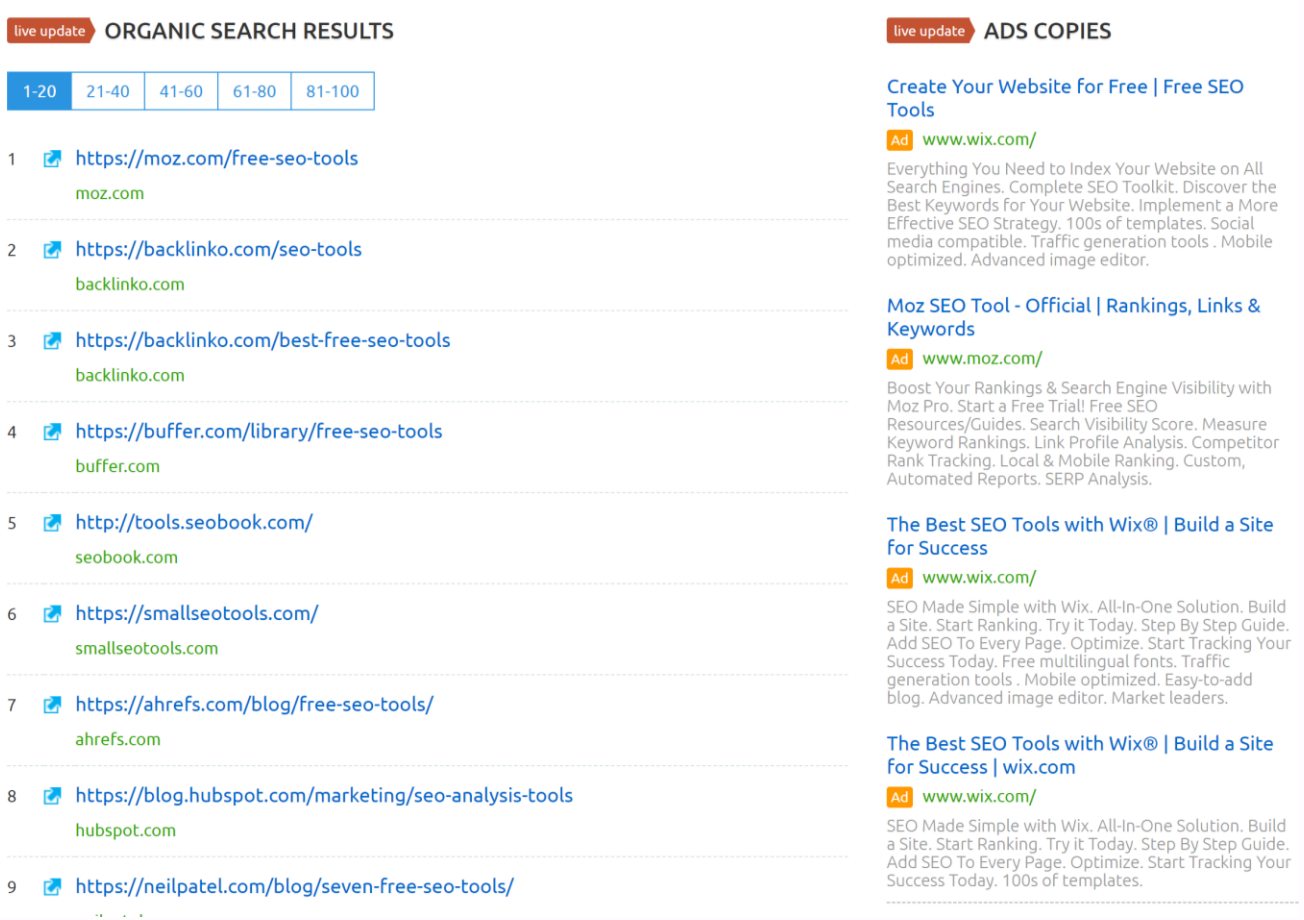
Palladian HTTP Extension

https://www.knime.com/community/palladian



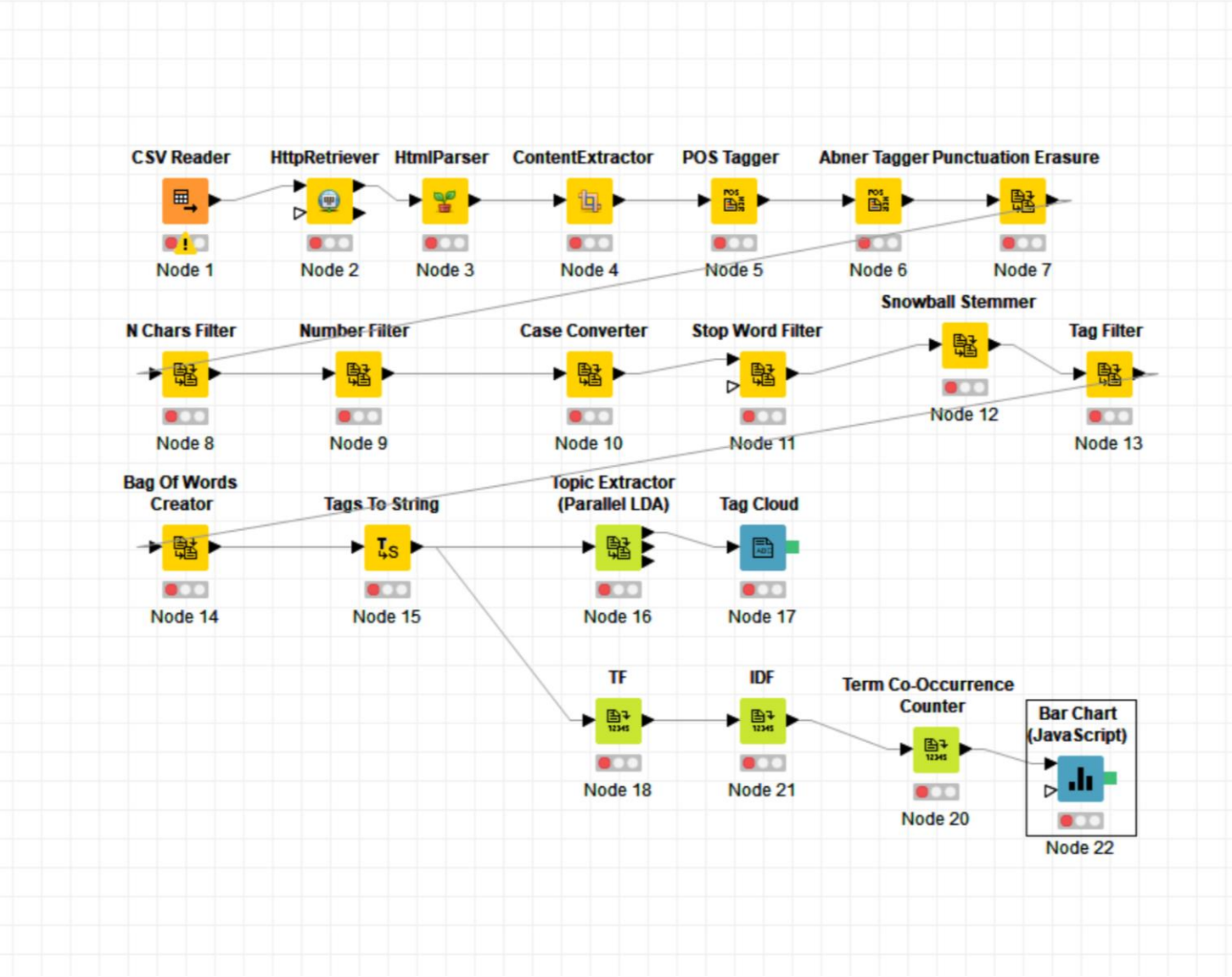
Download a “Corpus” for a Keyword

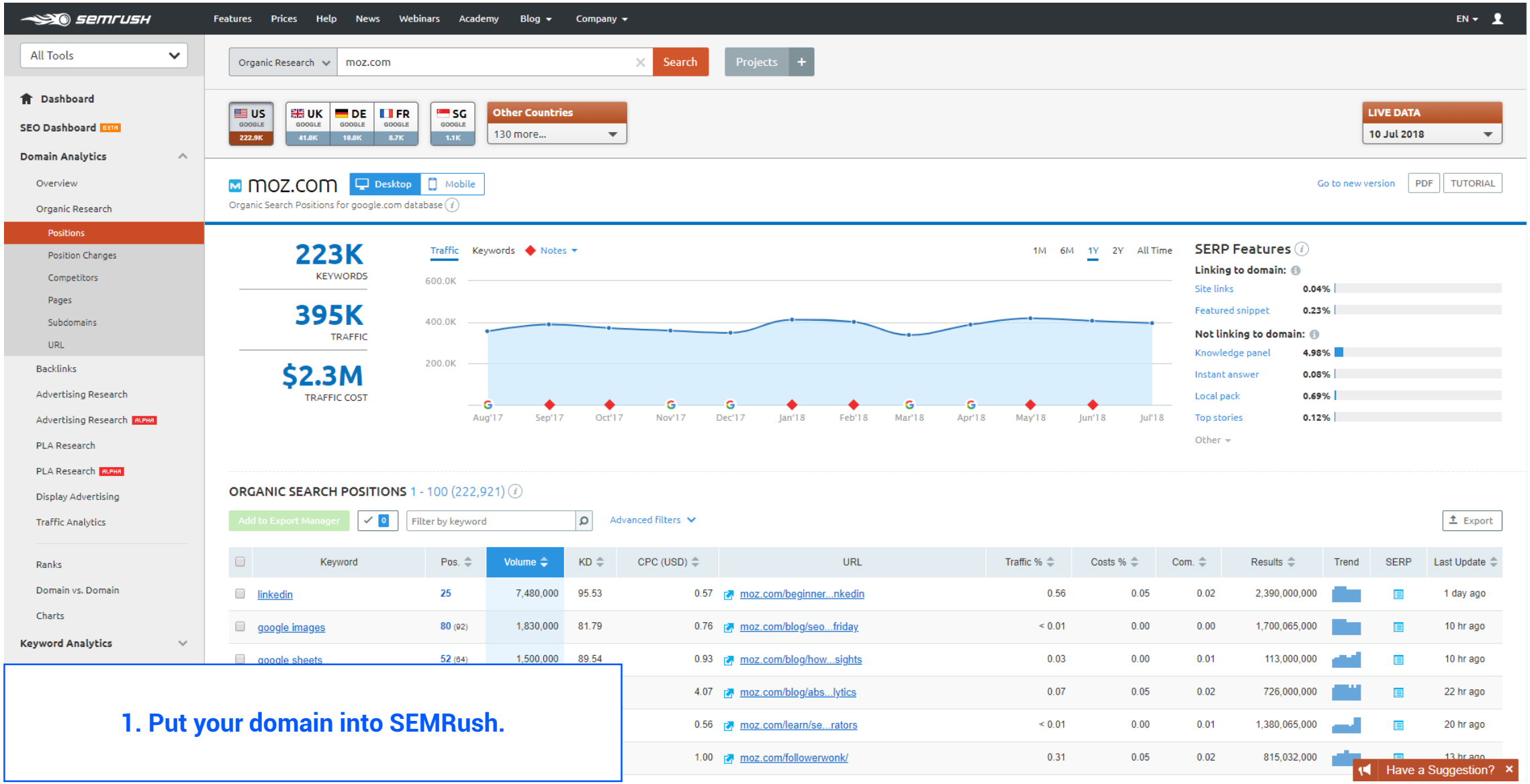
Ideally, we'd review all of the results for a given keyword to achieve parity with Google, but in this case we'll download the top 100 results for “seo tools” from SEMRush.



Here's a Starter Workflow

This workflow allows you to get started with LDA and computes Term Frequency.





1. Put your domain into SEMRush.

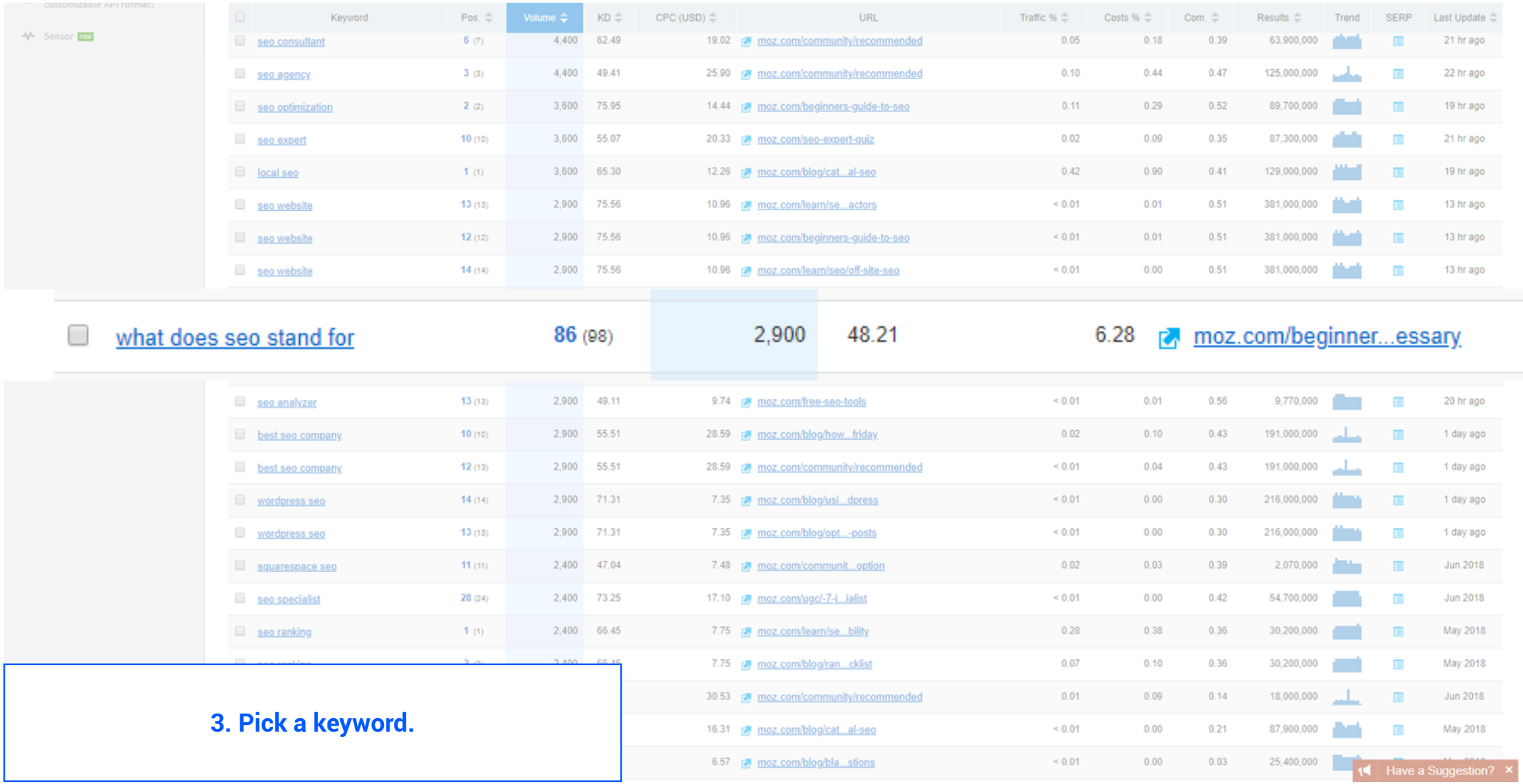
customizable API format?

Sensor new

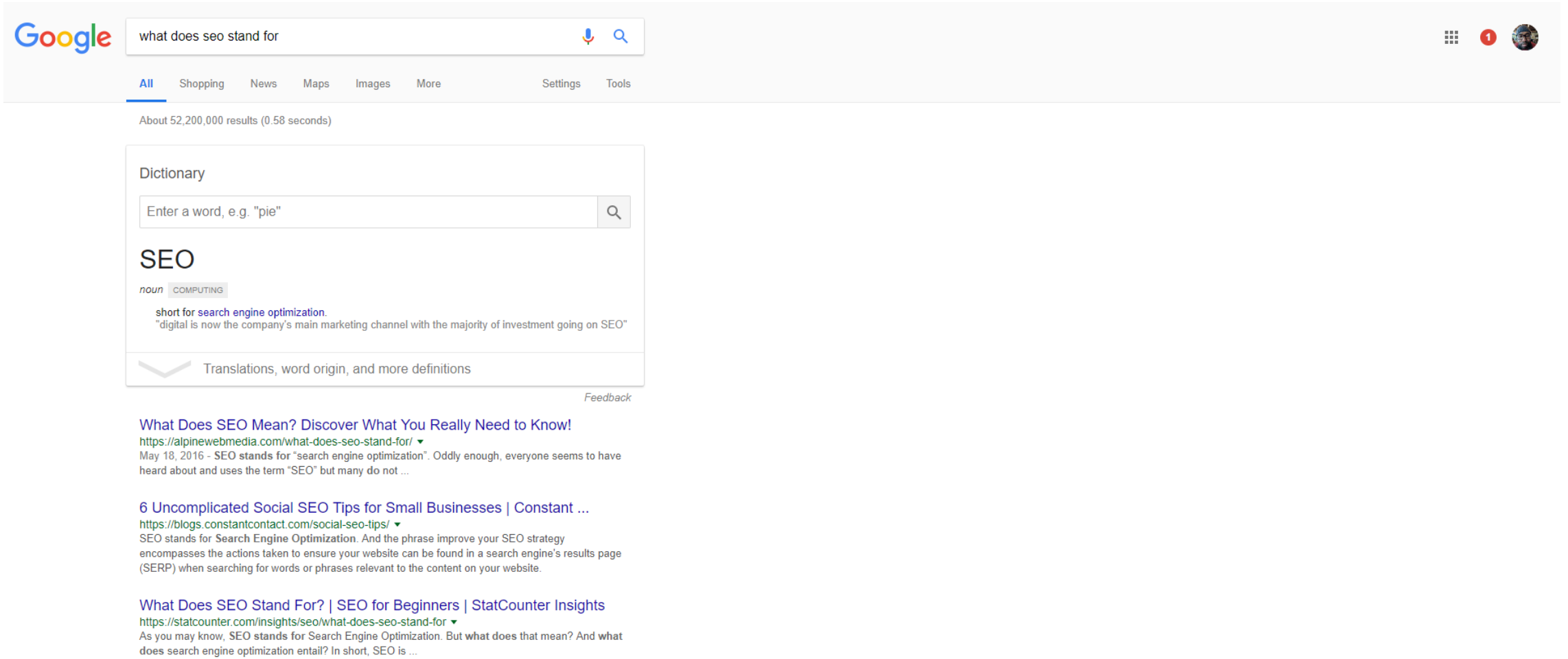
Keyword	Pos.	Volume	KD	CPC (USD)	URL	Traffic %	Costs %	Com.	Results	Trend	SERP	Last Update
seo consultant	6 (7)	4,400	62.49	19.02	moz.com/community/recommended	0.05	0.18	0.39	63,900,000			21 hr ago
seo agency	3 (3)	4,400	49.41	25.90	moz.com/community/recommended	0.10	0.44	0.47	125,000,000			22 hr ago
seo optimization	2 (2)	3,600	75.95	14.44	moz.com/beginners-guide-to-seo	0.11	0.29	0.52	89,700,000			19 hr ago
seo expert	10 (10)	3,600	55.07	20.33	moz.com/seo-expert-quiz	0.02	0.09	0.35	87,300,000			21 hr ago
local seo	1 (1)	3,600	65.30	12.26	moz.com/blog/cat...al-seo	0.42	0.90	0.41	129,000,000			19 hr ago
seo website	13 (13)	2,900	75.56	10.96	moz.com/learn/se...actors	< 0.01	0.01	0.51	381,000,000			13 hr ago
seo website	12 (12)	2,900	75.56	10.96	moz.com/beginners-guide-to-seo	< 0.01	0.01	0.51	381,000,000			13 hr ago
seo website	14 (14)	2,900	75.56	10.96	moz.com/learn/seo/off-site-seo	< 0.01	0.00	0.51	381,000,000			13 hr ago
seo website	10 (10)	2,900	75.56	10.96	moz.com/learn/seo/on-site	0.02	0.04	0.51	381,000,000			13 hr ago
what does seo stand for	86 (98)	2,900	48.21	6.28	moz.com/beginner...essary	< 0.01	0.00	0.03	4,290,000			1 day ago
majestic seo	45	2,900	61.37	11.04	moz.com/communit...ic-seo	< 0.01	0.00	0.09	5,950,000			10 hr ago
seo analyzer	13 (13)	2,900	49.11	9.74	moz.com/free-seo-tools	< 0.01	0.01	0.56	9,770,000			20 hr ago
best seo company	10 (10)	2,900	55.51	28.59	moz.com/blog/how...friday	0.02	0.10	0.43	191,000,000			1 day ago
best seo company	12 (13)	2,900	55.51	28.59	moz.com/community/recommended	< 0.01	0.04	0.43	191,000,000			1 day ago
wordpress seo	14 (14)	2,900	71.31	7.35	moz.com/blog/usi...dpress	< 0.01	0.00	0.30	216,000,000			1 day ago
wordpress seo	13 (13)	2,900	71.31	7.35	moz.com/blog/opt...-posts	< 0.01	0.00	0.30	216,000,000			1 day ago
squarespace seo	11 (11)	2,400	47.04	7.48	moz.com/communit...option	0.02	0.03	0.39	2,070,000			Jun 2018
seo specialist	28 (24)	2,400	73.25	17.10	moz.com/ugc/-7-j...ialist	< 0.01	0.00	0.42	54,700,000			Jun 2018
seo ranking	1 (1)	2,400	66.45	7.75	moz.com/learn/se...bility	0.28	0.38	0.36	30,200,000			May 2018
				7.75	moz.com/blog/ran...cklist	0.07	0.10	0.36	30,200,000			May 2018
				30.53	moz.com/community/recommended	0.01	0.09	0.14	18,000,000			Jun 2018
				16.31	moz.com/blog/cat...al-seo	< 0.01	0.00	0.21	87,900,000			May 2018
				6.57	moz.com/blog/bla...stions	< 0.01	0.00	0.03	25,400,000			

2. See what you don't rank well for in organic research.

Have a Suggestion? ×



3. Pick a keyword.



4. Look at the SERP.

ALPINE WEB MEDIA

Web Design | Reviews | Portfolio | Web Services | Blog | Contact

May 18, 2016 By William Alexander 2 Comments

What does SEO stand for?

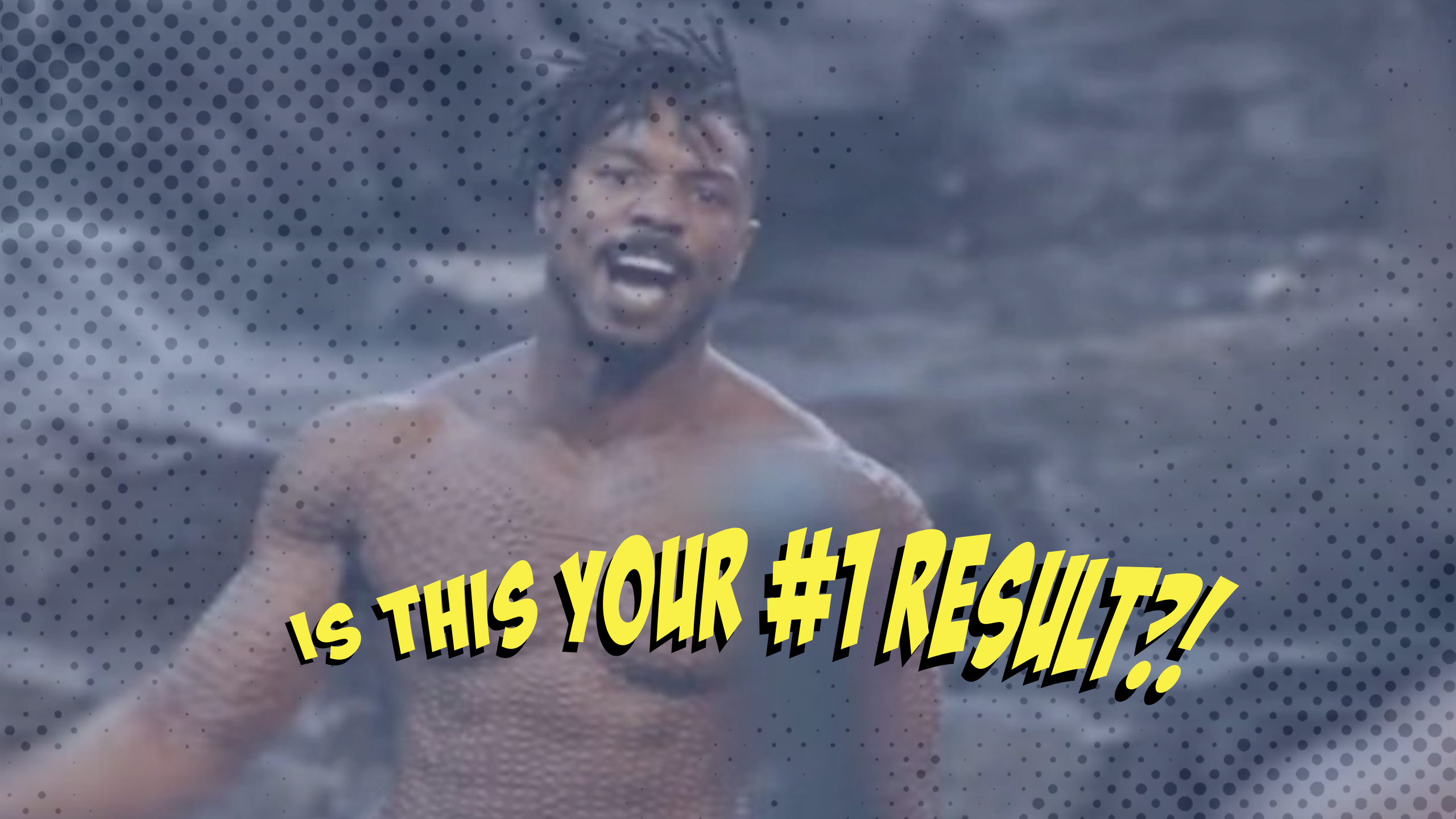
What does SEO mean, and what do I REALLY need to know about it?

3 Shares

me when I say:

erstand what SEO is and how to implement it properly.

Really Google?


IS THIS YOUR #1 RESULT?!

MOZ Products Blog About Learn SEO Moz Pro Moz Local Free SEO Tools Log in

CHAPTERS (CLICK TO ACCESS)

Tweet 8.7K Like Share

CHAPTER THREE WHY SEARCH ENGINE MARKETING IS NECESSARY

DOWNLOAD PDF

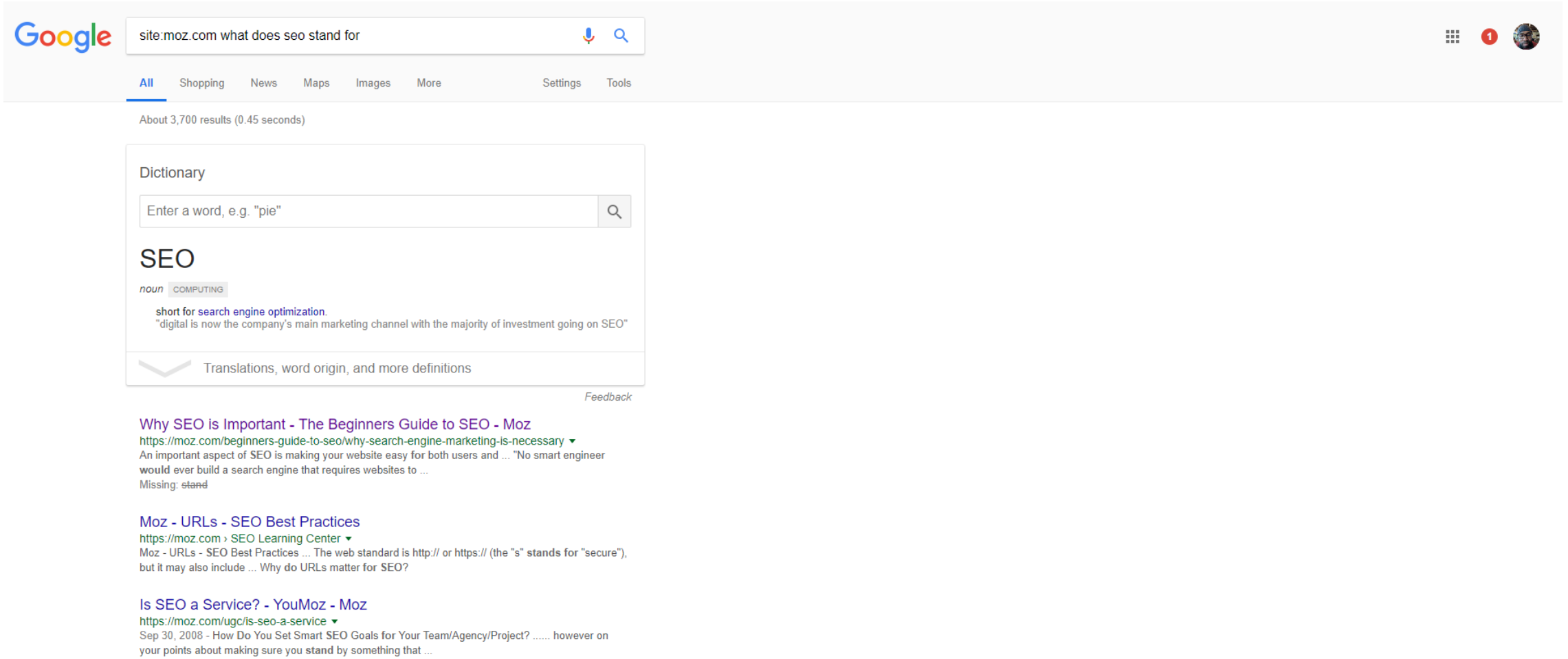
An important aspect of SEO is making your website easy for both users and search engine robots to understand. Although search engines have become increasingly sophisticated, they still can't see and understand a web page the same way a human can. SEO helps the engines figure out what each page is about, and how it may be useful for users.

A Common Argument Against SEO

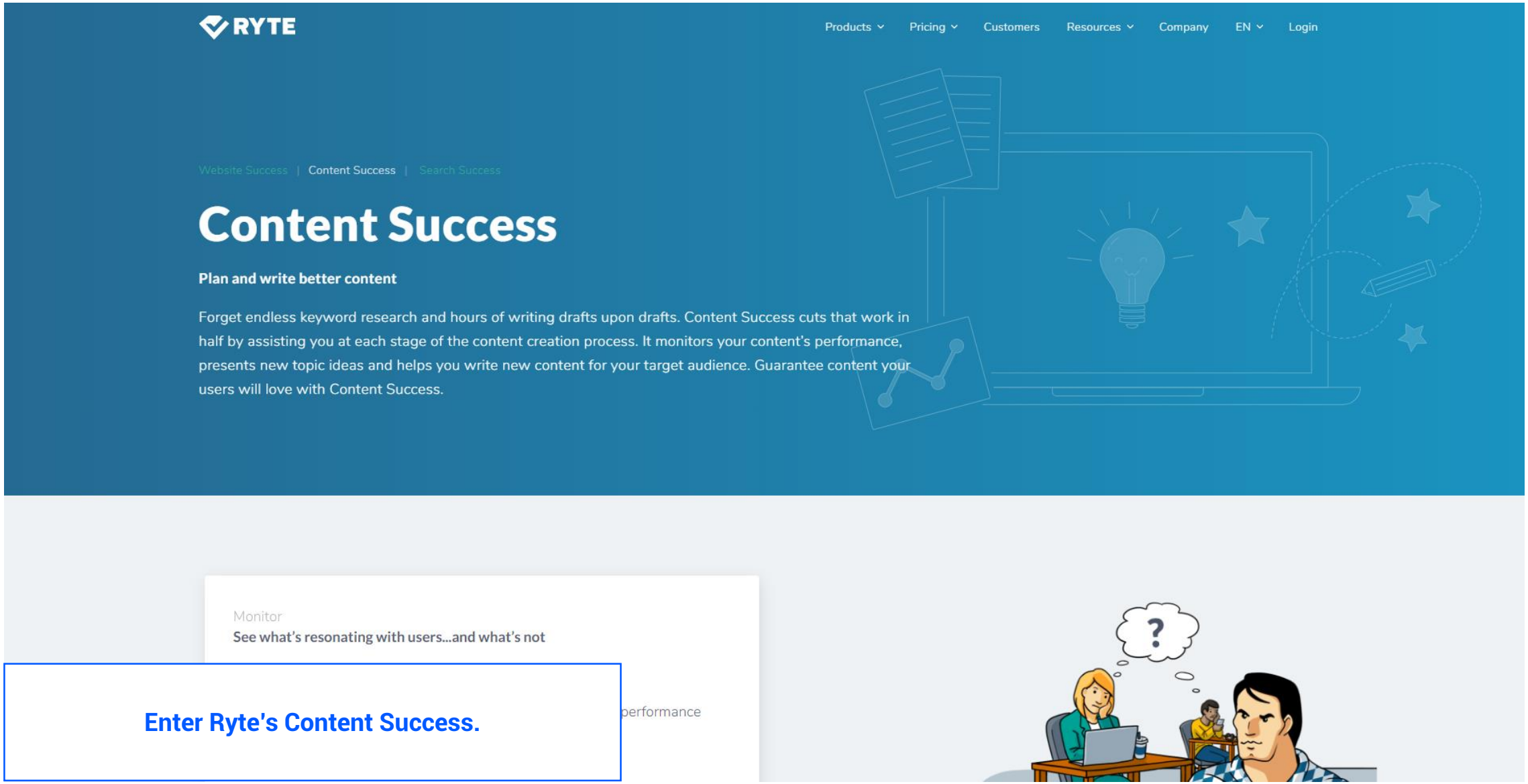
We frequently hear statements like this:

"No smart engineer would ever build a search engine that requires websites to follow certain rules or principles in order to be ranked or indexed. Anyone with half a brain would want a system that can crawl through any architecture, parse any amount of complex or imperfect code, and still find a way to return the most relevant results, not the ones that have been 'optimized' by unlicensed search marketing experts."

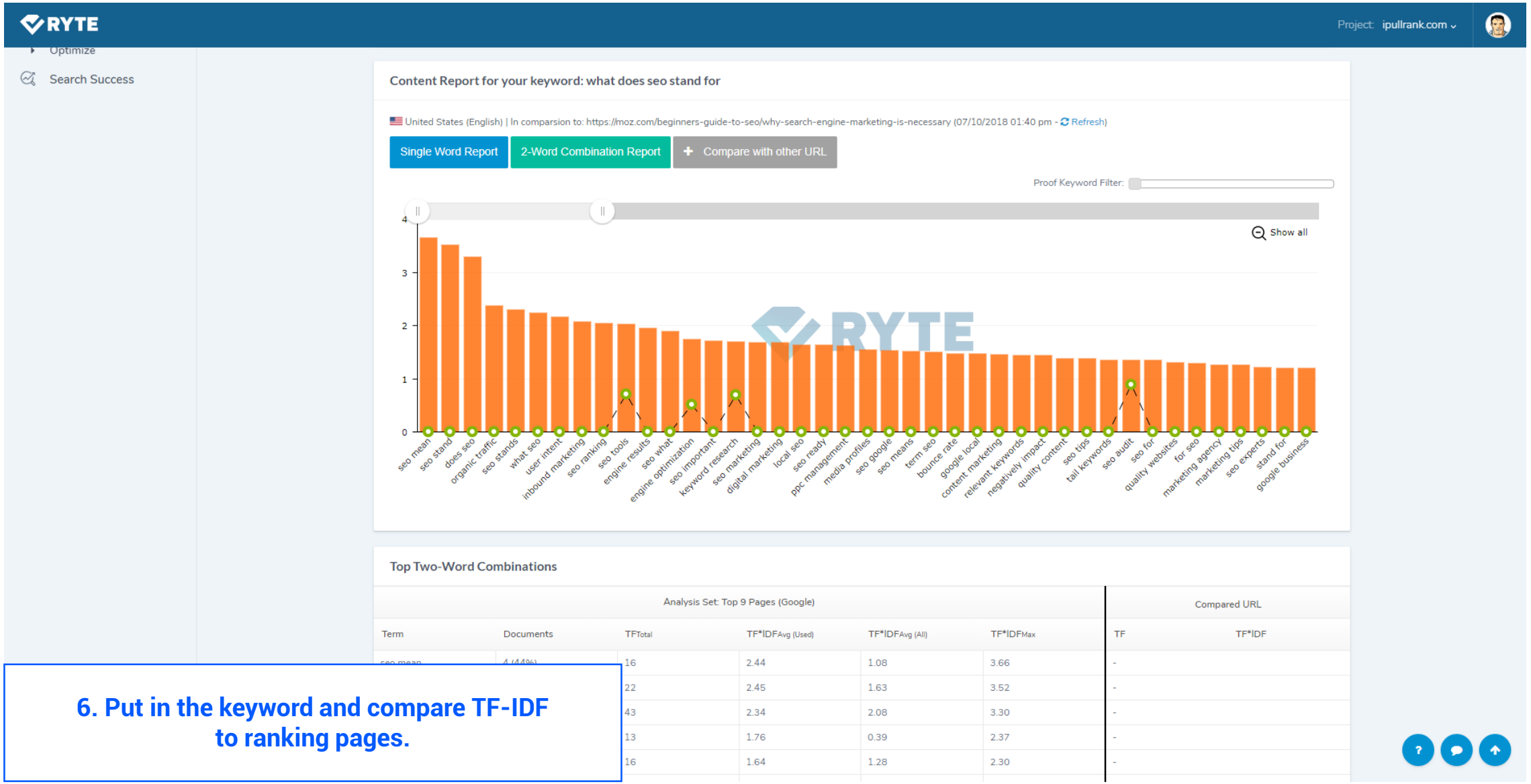
This is the page that ranks #86?!



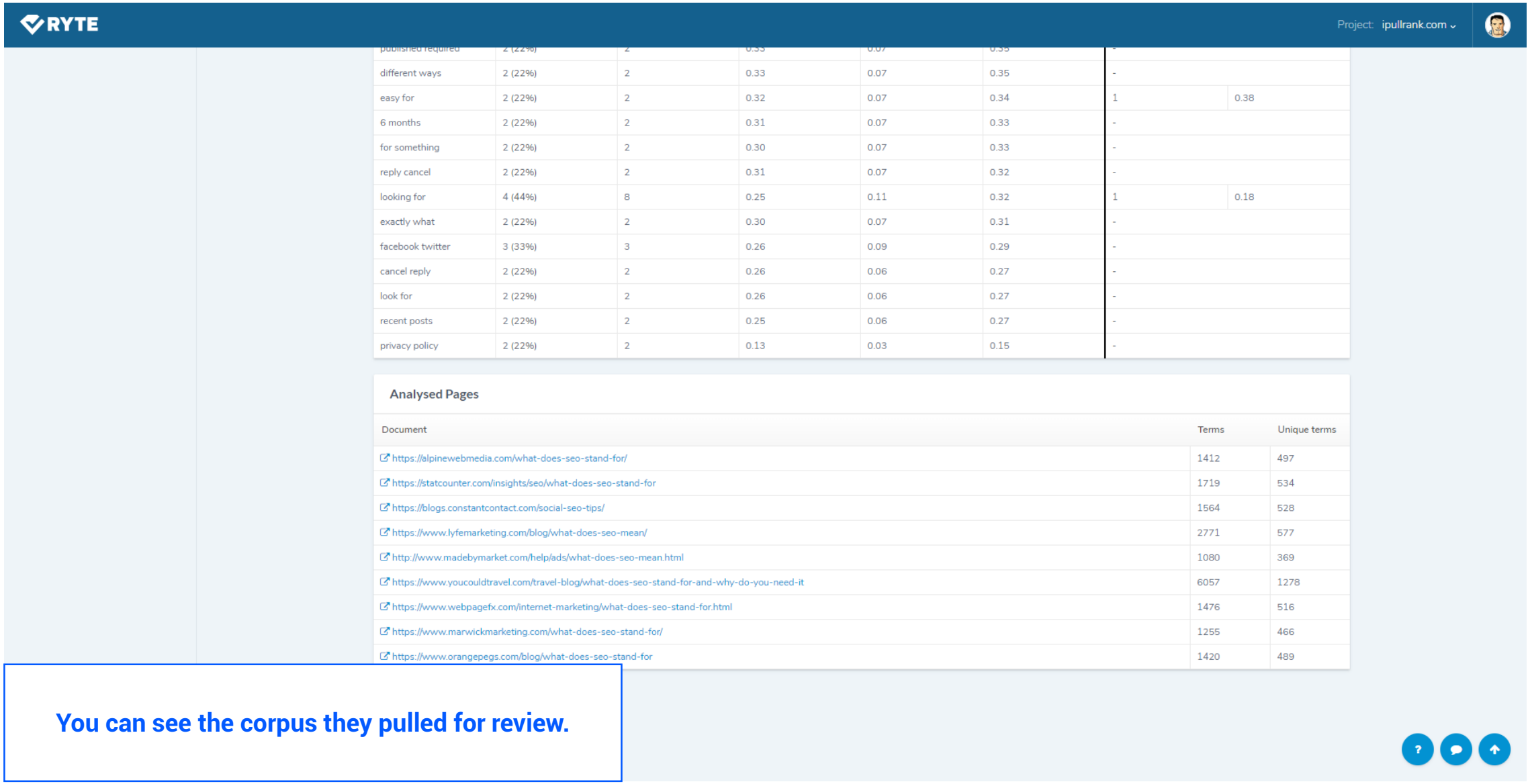
5. See what else could rank for our site.


RYTE
Products
Pricing
Customers
Resources
Company
EN
Login
Website Success | Content Success | Search Success
Content Success
Plan and write better content
Forget endless keyword research and hours of writing drafts upon drafts. Content Success cuts that work in half by assisting you at each stage of the content creation process. It monitors your content's performance, presents new topic ideas and helps you write new content for your target audience. Guarantee content your users will love with Content Success.
Monitor
See what's resonating with users...and what's not
performance
Enter Ryte's Content Success.

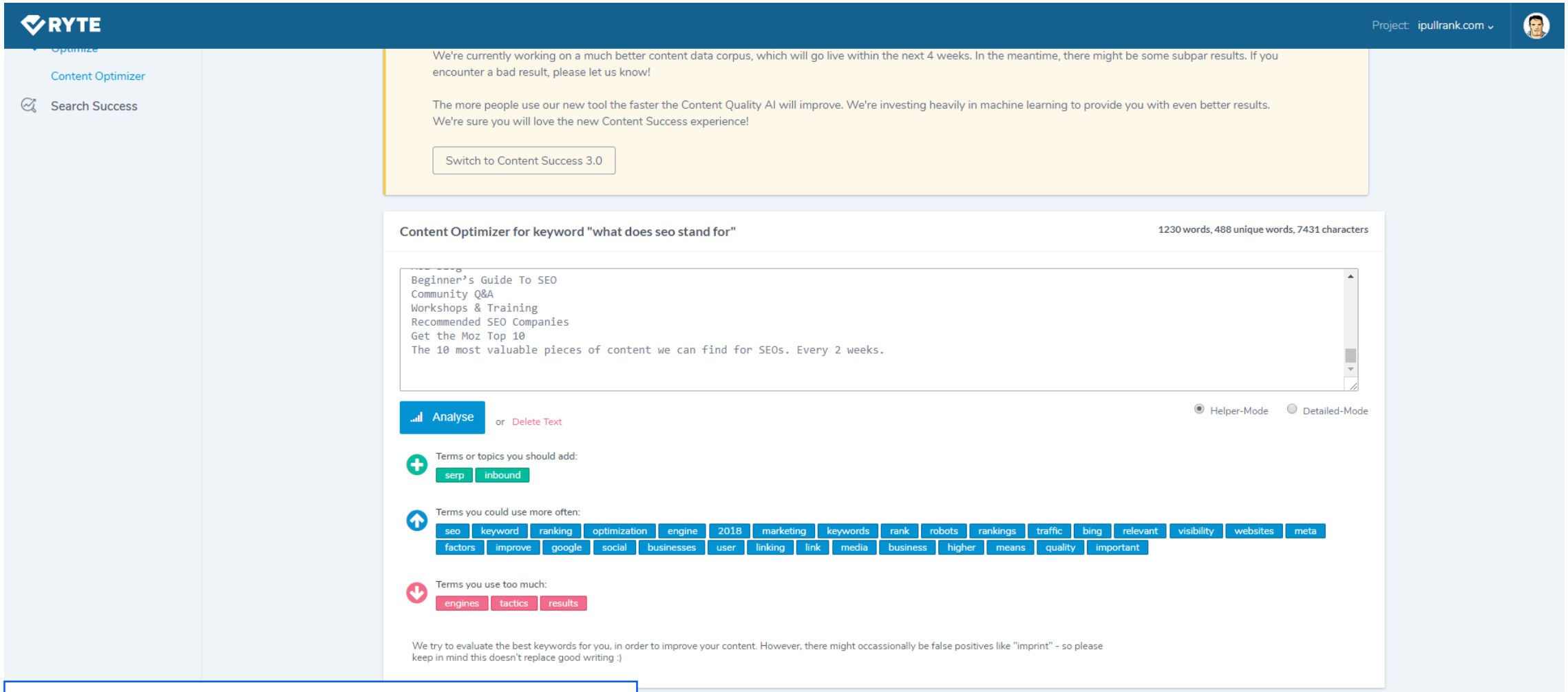
?



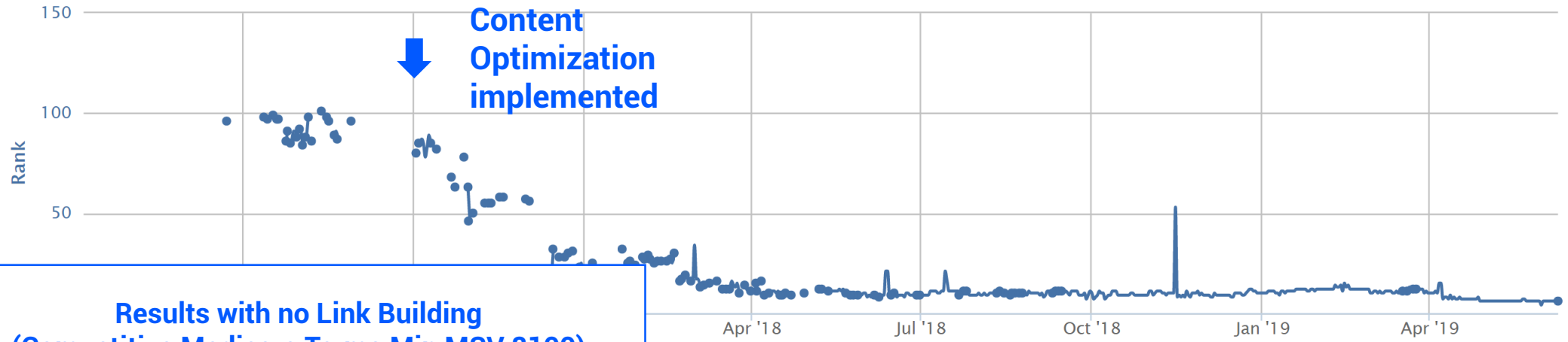
6. Put in the keyword and compare TF-IDF to ranking pages.



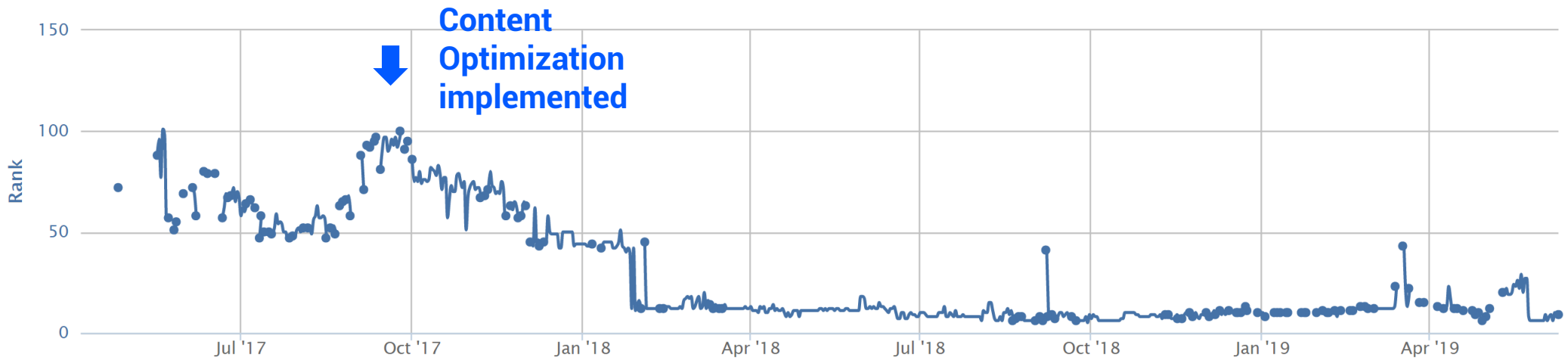
You can see the corpus they pulled for review.



7. Put the copy in. See what important keywords are missing and make edits.



Results with no Link Building
(Competitive Medicare Terms Min MSV 8100)



Results with No Link Building
(Competitive Medicare Terms Min MSV 8100)

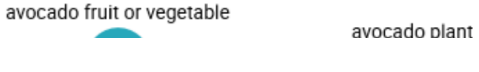
Agile Content Development

How do you make great revenue-driving online content?

Iterate, iterate, iterate. With Searchmetrics, **agile content development** backs the creative process for the first time with deep learning and the latest in data science.

Create lively content that is also optimized to meet user intent and stand atop the search page rankings. Drive predictable revenue gains using the new future of web content development.

Discovery

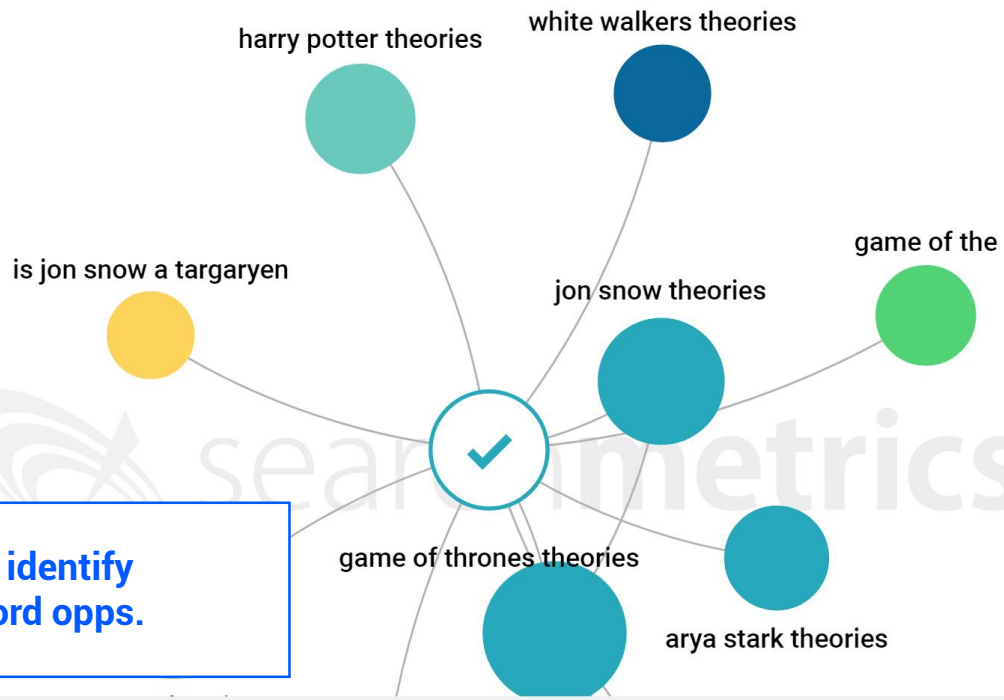


Topic Explorer

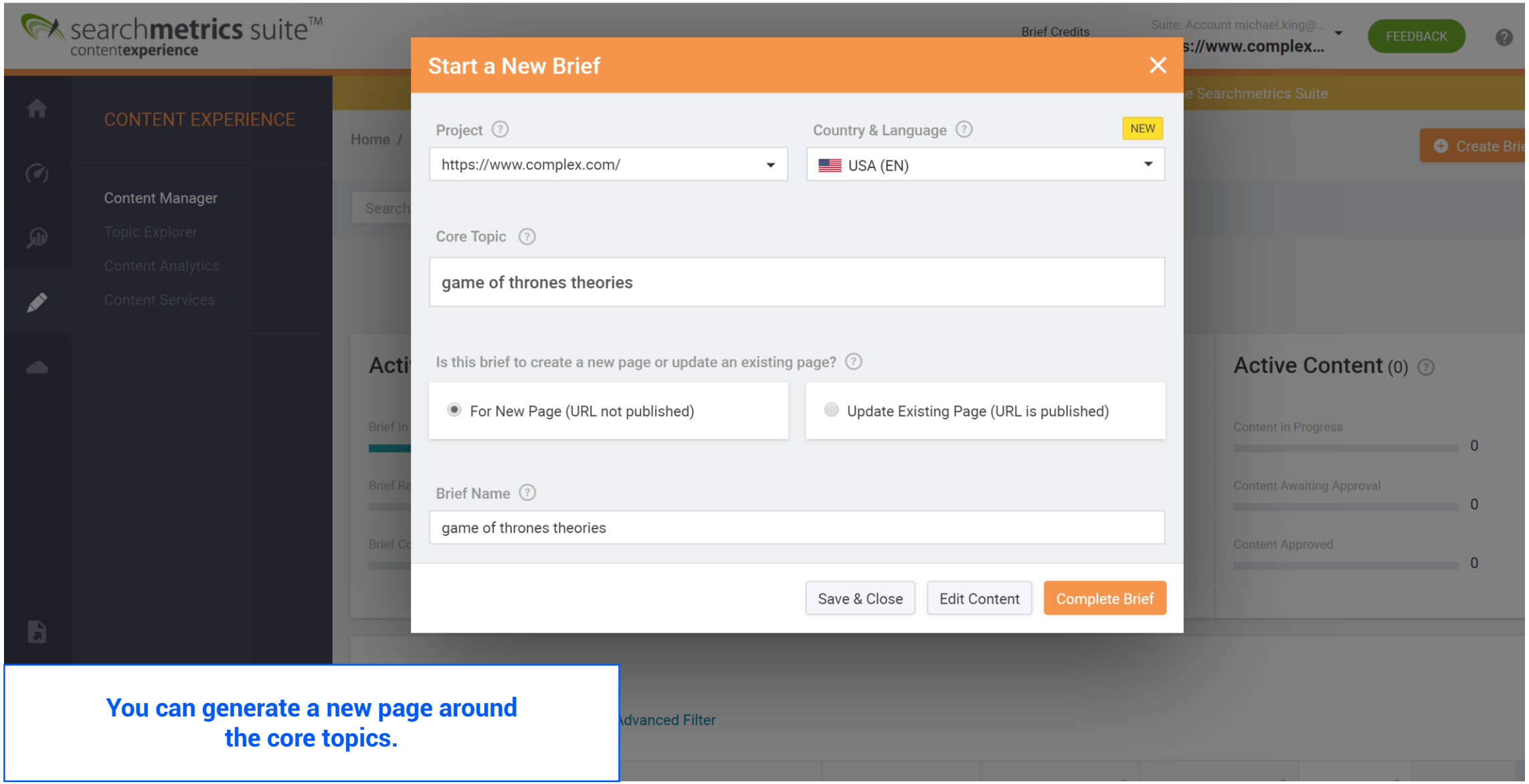
Semantic Associations | Rankings | Seasonality | Competitiveness | Search Intent

Topics Clustered by Semantic Proximity

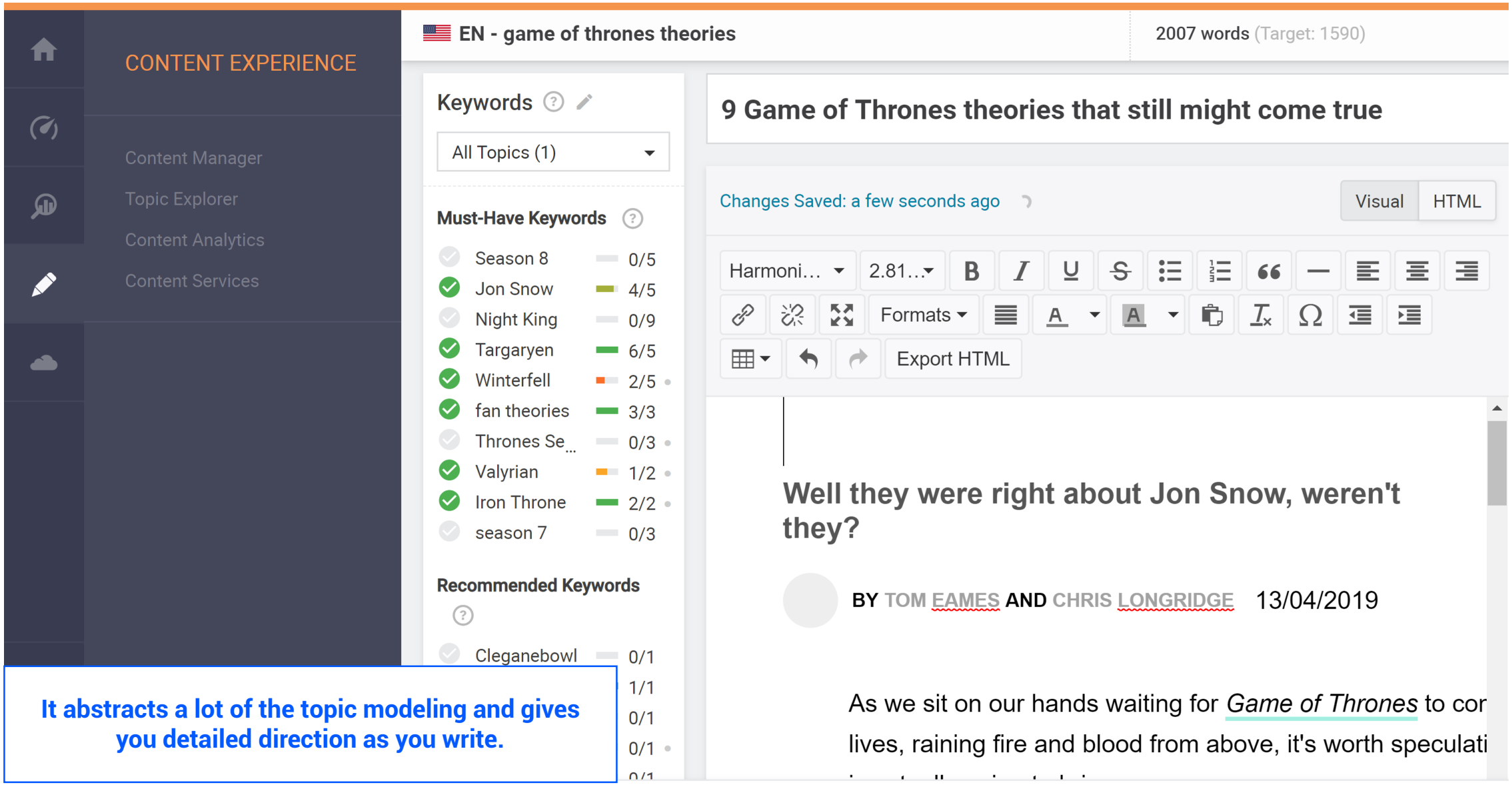
Add Topic



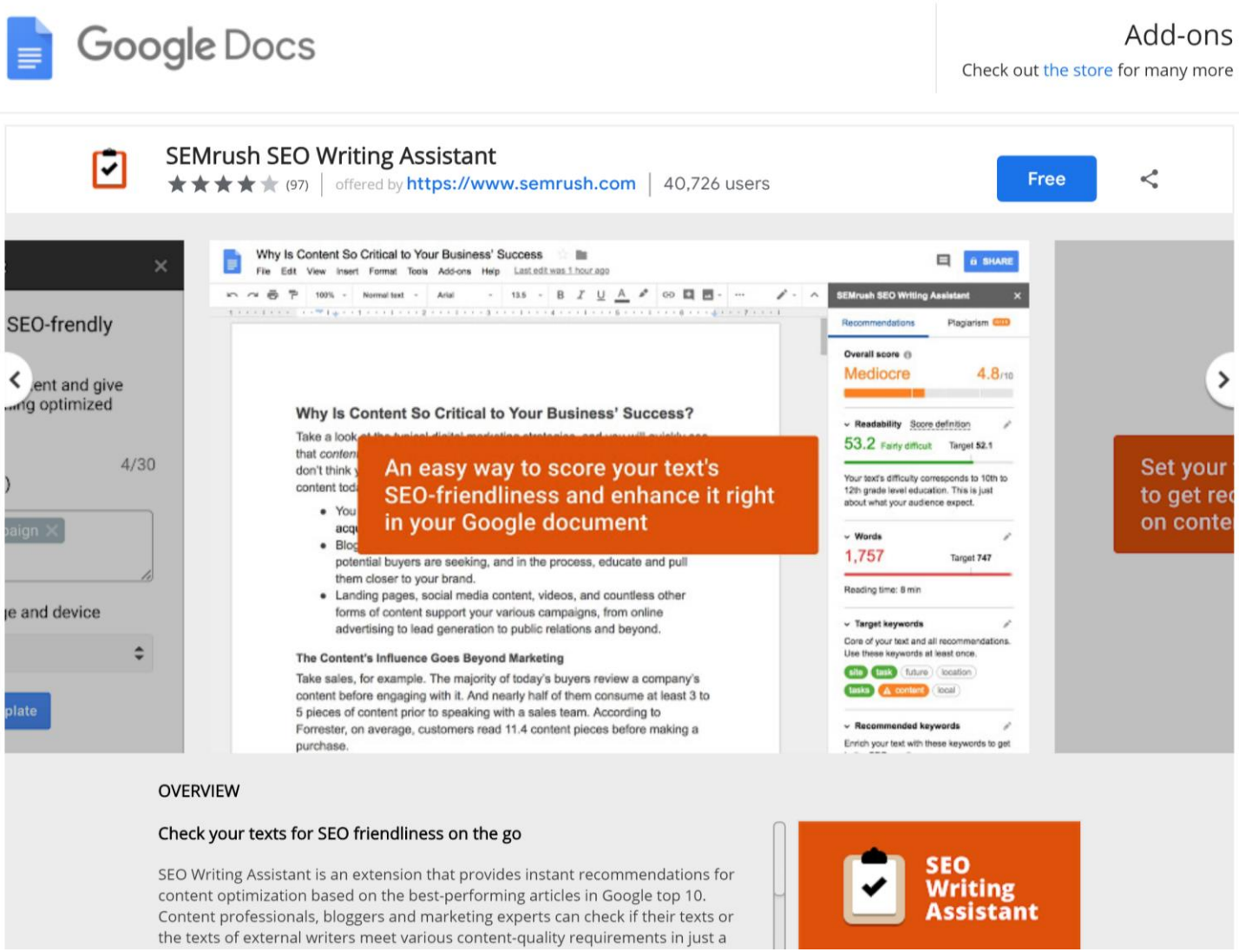
The topic explorer helps you identify semantically clustered keyword opps.



You can generate a new page around the core topics.

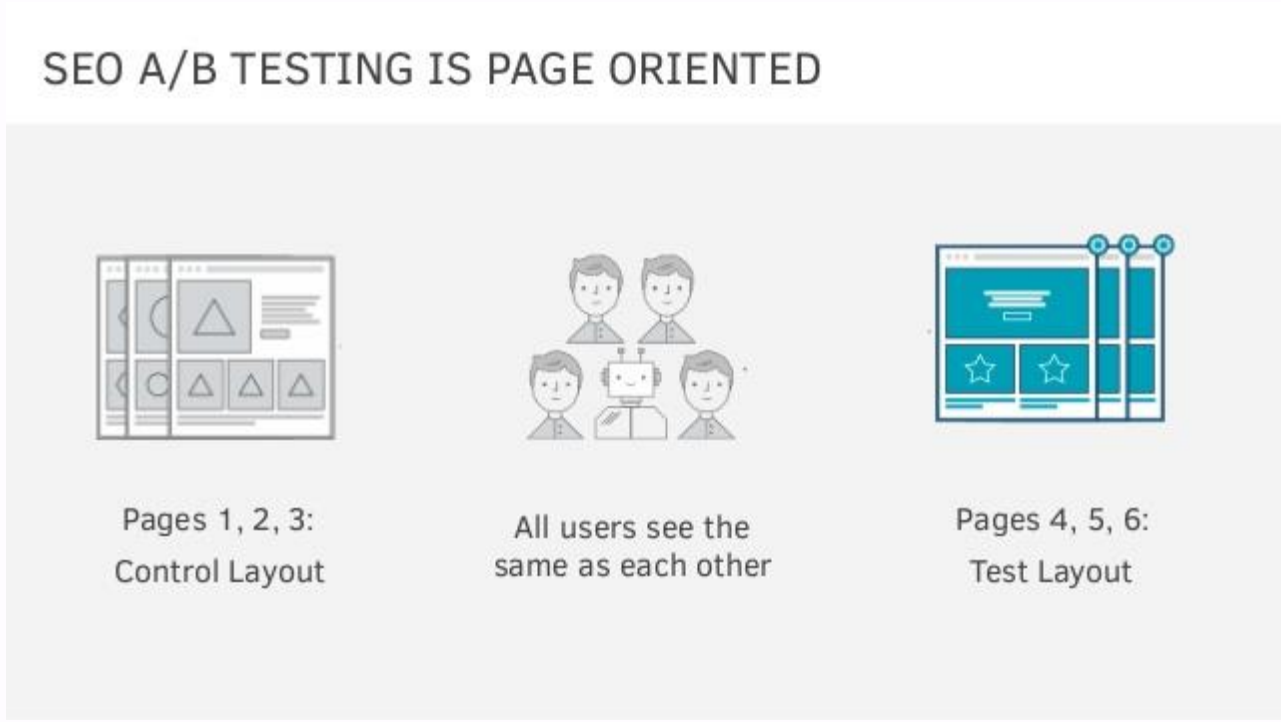


It abstracts a lot of the topic modeling and gives you detailed direction as you write.


Google Docs
Add-ons
Check out the store for many more
SEMrush SEO Writing Assistant
(97) | offered by https://www.semrush.com | 40,726 users
Free
An easy way to score your text's SEO-friendliness and enhance it right in your Google document
Why Is Content So Critical to Your Business' Success?
OVERVIEW
Check your texts for SEO friendliness on the go
SEO Writing Assistant is an extension that provides instant recommendations for content optimization based on the best-performing articles in Google top 10. Content professionals, bloggers and marketing experts can check if their texts or the texts of external writers meet various content-quality requirements in just a
SEO Writing Assistant

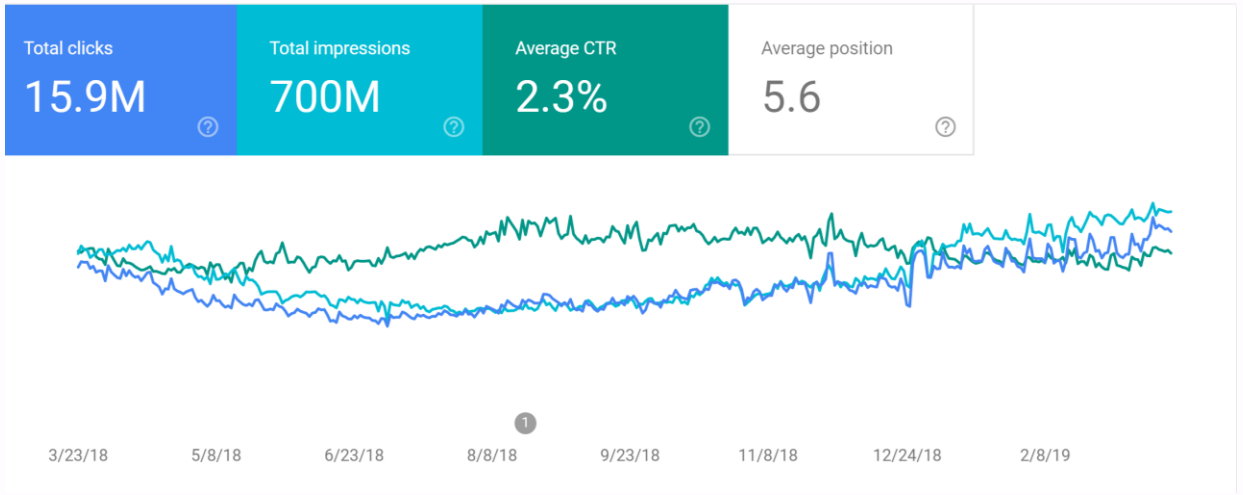
A/B Test Metadata

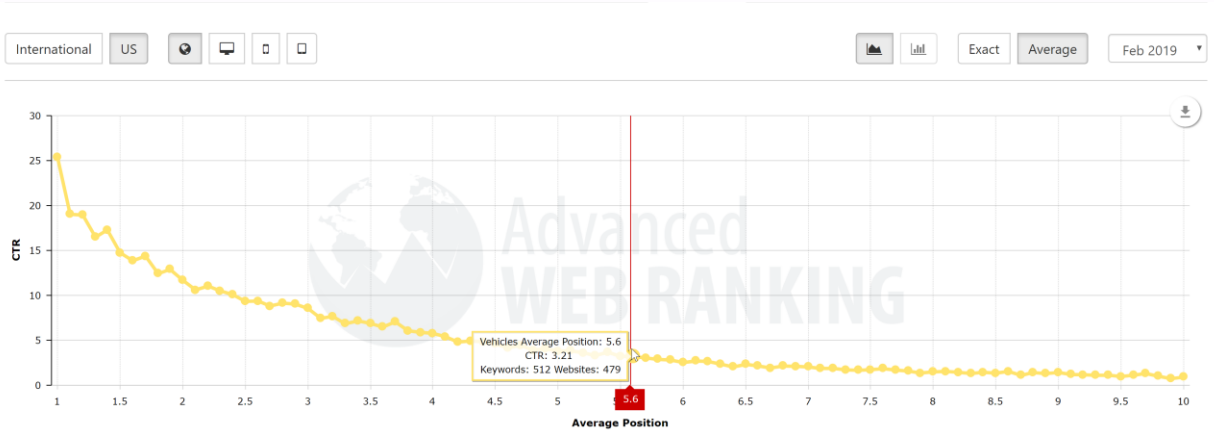
A/B test your page titles and meta descriptions by creating different buckets of pages, making changes, and reviewing their CTR performance.



Analysis: Compare your CTR

AWR publishes CTR stats by vertical https://advancedwebranking.com/ctrstudy





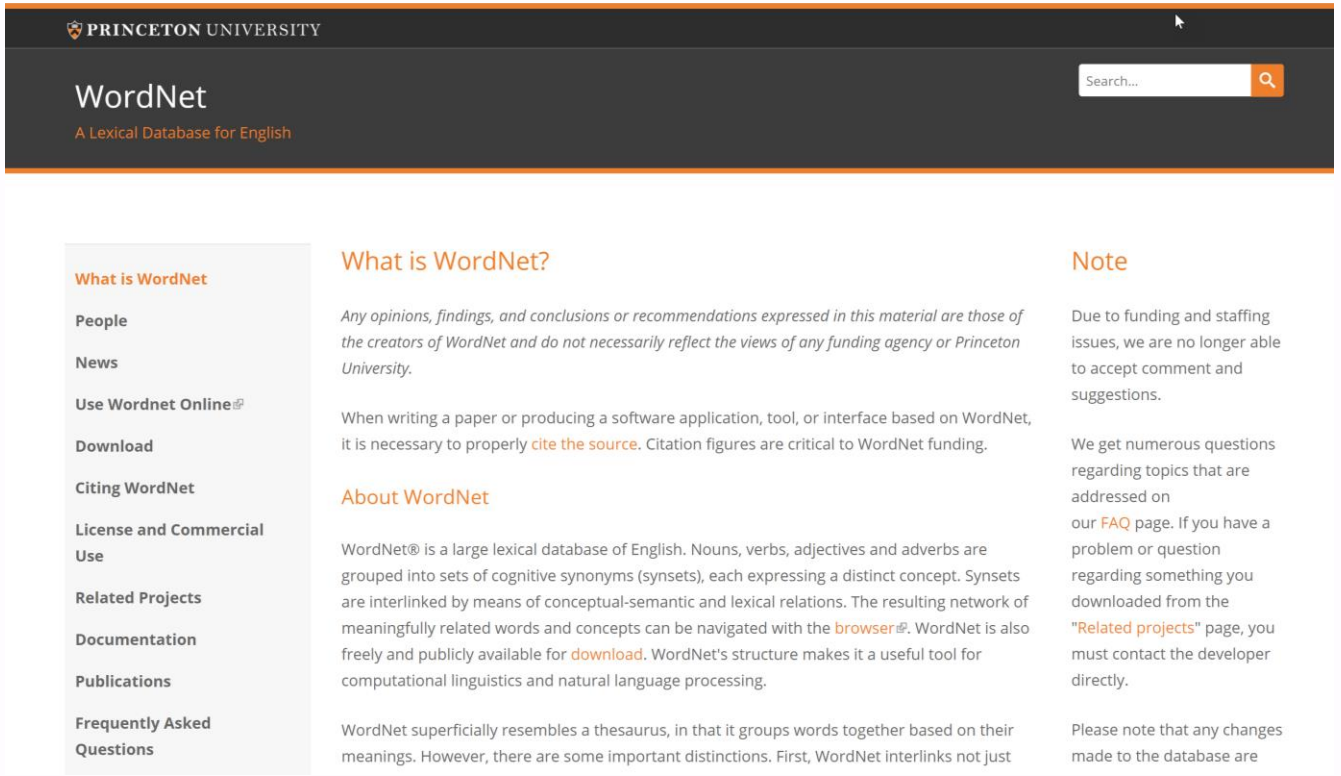
Hypothesis: Identify Specifically Where You Underperform

Develop your hypotheses based keywords that underperform based on the expected CTR.

Query	Clicks	Impressions	CTR	Position	Expected CTR	Difference
[redacted]	41793	2052877	2.04	1.7	14.37	-12.33
[redacted]	3275	47827	6.85	1.2	18.93	-12.08
[redacted]	5309	215846	2.46	1.7	14.37	-11.91
[redacted]	3064	109718	2.79	1.7	14.37	-11.58
[redacted]	39293	2383742	1.65	1.9	12.89	-11.24
[redacted]	4280	298092	1.44	1.8	12.41	-10.97
[redacted]	8745	315591	2.77	1.9	12.89	-10.12
[redacted]	4359	48040	9.07	1.1	19.03	-9.96
[redacted]	4560	567108	0.80	2.1	10.61	-9.81
[redacted]	3327	253411	1.31	2.2	11.03	-9.72
[redacted]	4566	299271	1.53	2.2	11.03	-9.50
[redacted]	3531	318808	1.11	2.1	10.61	-9.50
[redacted]	5382	324103	1.66	2.2	11.03	-9.37
[redacted]	3829	70760	5.41	1.5	14.75	-9.34
[redacted]	9124	710214	1.28	2.1	10.61	-9.33
[redacted]	3088	130651	2.36	2.0	11.69	-9.33
[redacted]	8232	347780	2.37	2.0	11.69	-9.32
[redacted]	23113	1778926	1.30	2.1	10.61	-9.31
[redacted]	27552	1086293	2.54	2.0	11.69	-9.15
[redacted]	9183	605786	1.52	2.3	10.51	-8.99
[redacted]	8886	436335	2.04	2.2	11.03	-8.99
[redacted]	3844	232721	1.65	2.1	10.61	-8.96
[redacted]	10808	443893	2.43	2.2	11.03	-8.60
[redacted]	6997	900369	0.78	2.5	9.33	-8.55
[redacted]	4841	566764	0.85	2.5	9.33	-8.48
[redacted]	4927	305649	1.61	2.4	10.08	-8.47
[redacted]	3635	226266	1.61	2.4	10.08	-8.47

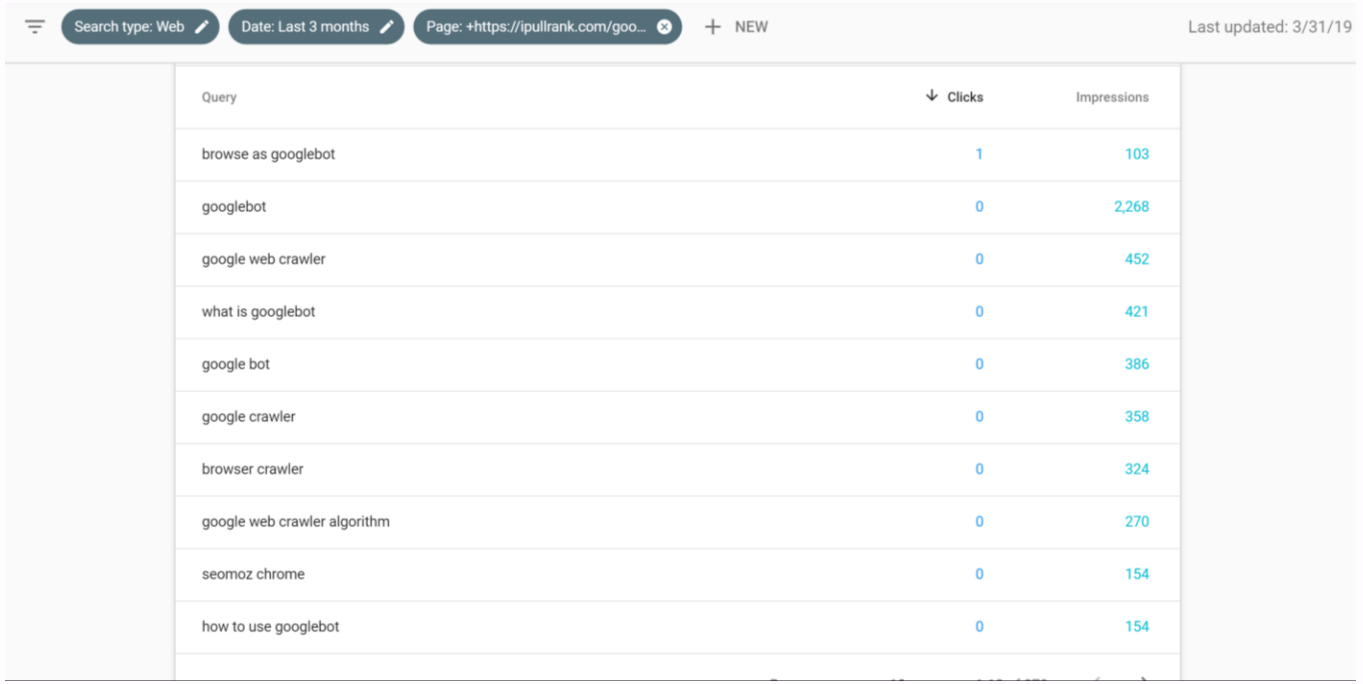
Use Wordnet to Scale Keyword Variant Discovery

https://wordnet.princeton.edu/



Automate Inclusion of Emerging Keywords from GSC in Rankings

Monitor your queries on a page-level week over week to identify new emerging keyword opportunities and automatically add them to your rank tracking.



Query	Clicks	Impressions
browse as googlebot	1	103
googlebot	0	2,268
google web crawler	0	452
what is googlebot	0	421
google bot	0	386
google crawler	0	358
browser crawler	0	324
google web crawler algorithm	0	270
seomoz chrome	0	154
how to use googlebot	0	154

grepw(.*)rds

API Documentation Plans and Pricing Blog Support

Support

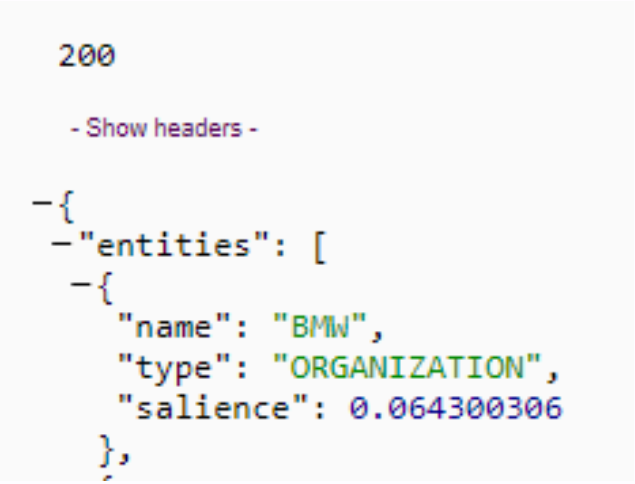
11 New APIs for GrepWords Customers!

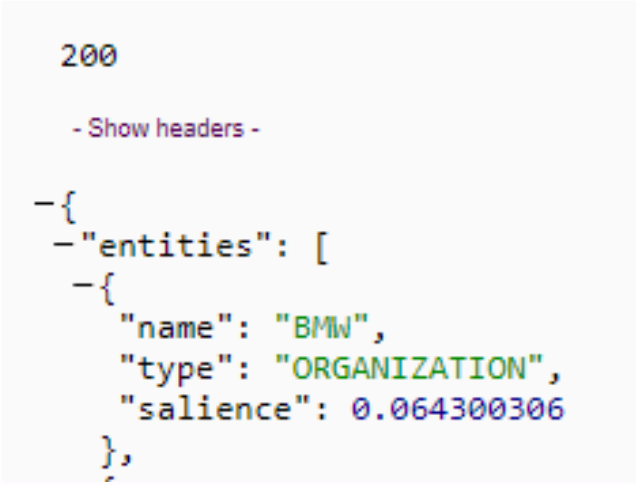
We have spent the last 2 weeks now rolling out a feature every single day. If you need a quick recap, we rolled out... Local Intent Keywords Commercial Intent Keywords Video Intent Keywords Suggestions for Amazon Suggestions for Ebay...

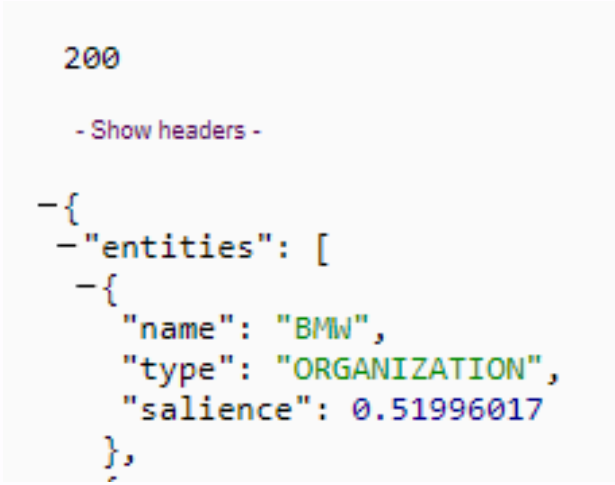
Review Competitors' Usage of Entities

When Google is processing content, they extract and score entities. Compare your page's entity salience and ordering against your competitors to identify what you need to feature.

Do the same as you do your keyword research.







Use Entities for Intelligent Internal Link Building

You can use the same data on your own site to determine how to automate internal link building.

Google Cloud Why Google Solutions Products Pricing Getting started Docs Support Console

AI & Machine Learning Products Contact sales

Cloud Natural Language

Derive insights from unstructured text using Google machine learning.

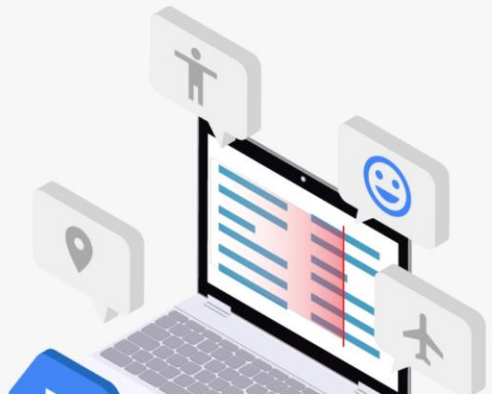
GO TO CONSOLE

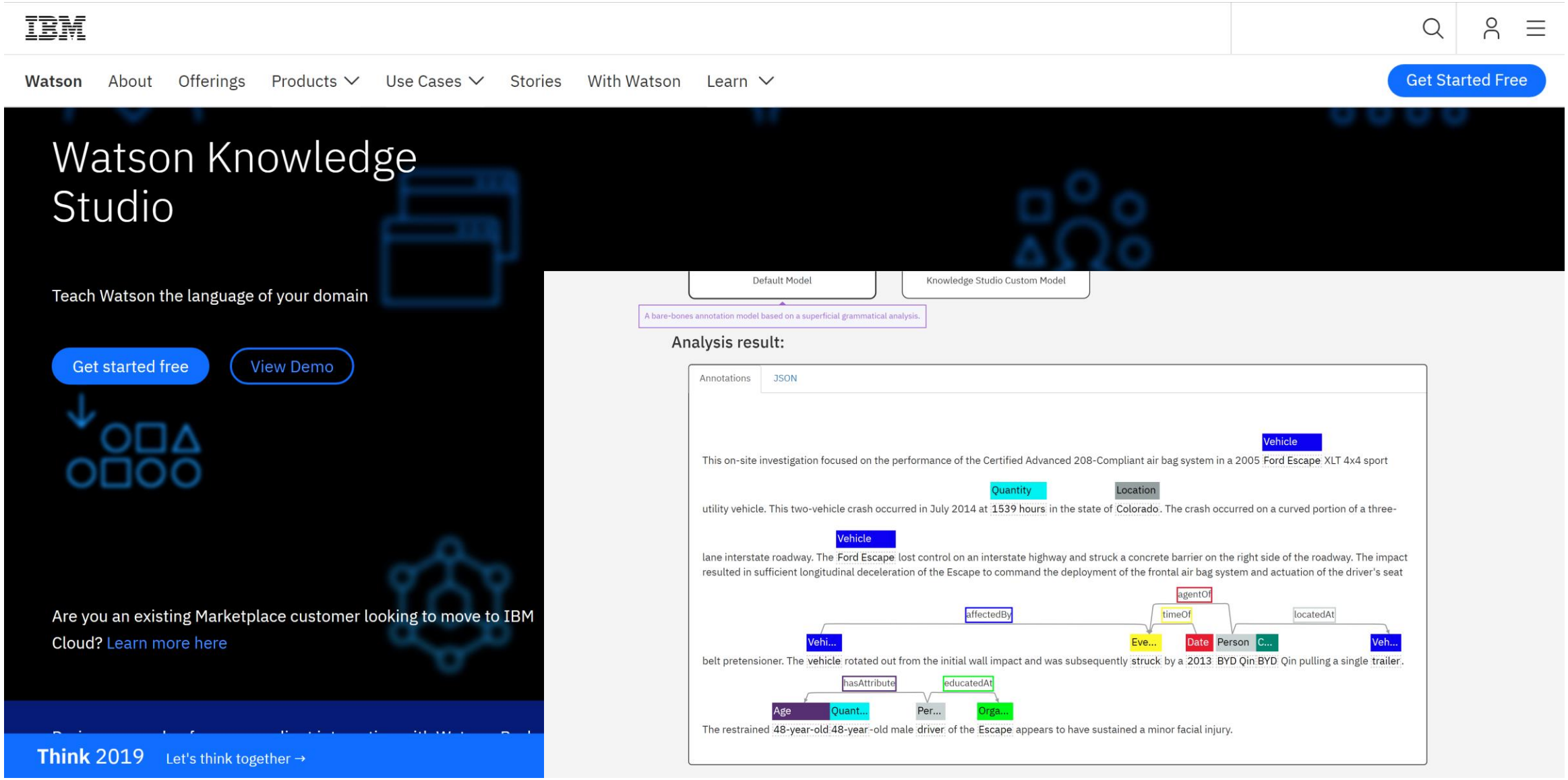
View documentation for this product.

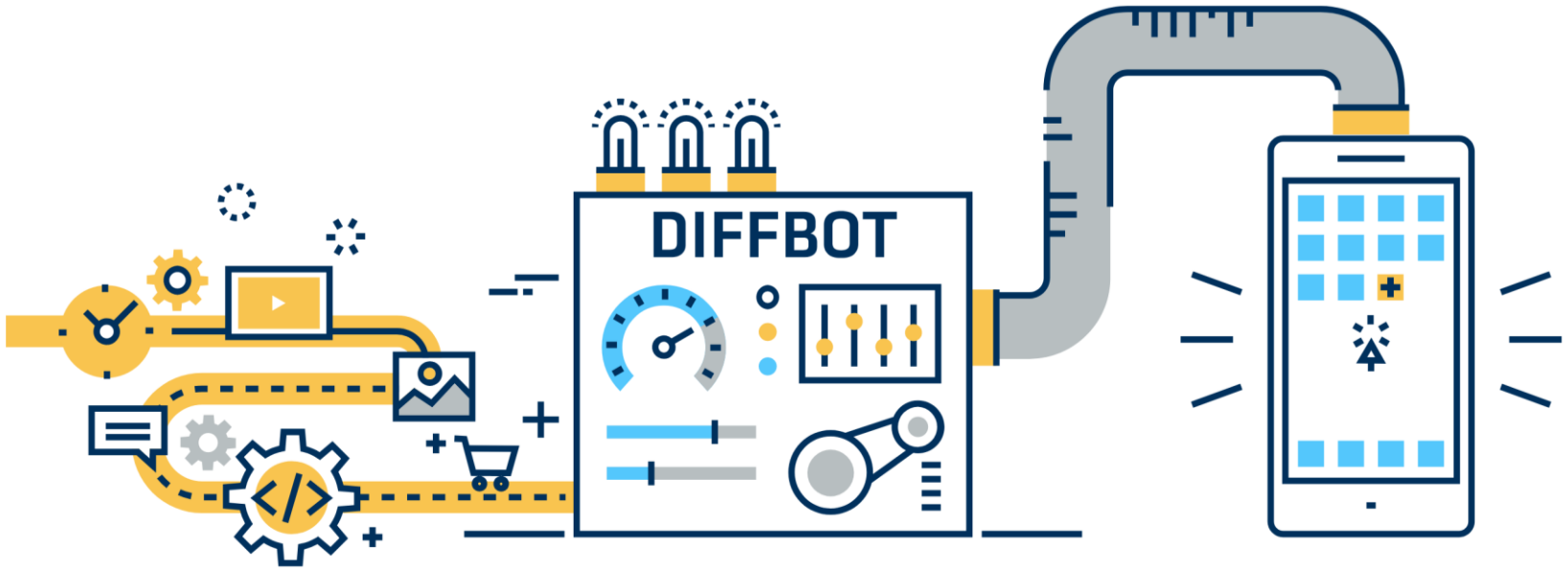
Powerful text analysis

Google Cloud Natural Language reveals the structure and meaning of text both through powerful pretrained machine learning models in an easy to use REST API and through custom models that are easy to build with AutoML Natural Language BETA. Learn more about Cloud AutoML.

You can use Cloud Natural Language to extract information about people, places, events, and much more mentioned in text documents, news articles, or blog posts. You can use it to understand sentiment about your product on social media or parse intent from customer conversations happening in a call center or a messaging app. You can analyze text uploaded in your request or integrate with your document storage on Google Cloud




IBM
Watson
About
Offerings
Products
Use Cases
Stories
With Watson
Learn
Get Started Free
Watson Knowledge Studio
Teach Watson the language of your domain
Get started free
View Demo
Are you an existing Marketplace customer looking to move to IBM Cloud? Learn more here
Think 2019
Let's think together →
Default Model
Knowledge Studio Custom Model
A bare-bones annotation model based on a superficial grammatical analysis.
Analysis result:
Annotations
JSON
Vehicle
This on-site investigation focused on the performance of the Certified Advanced 208-Compliant air bag system in a 2005 Ford Escape XLT 4x4 sport
Quantity
Location
utility vehicle. This two-vehicle crash occurred in July 2014 at 1539 hours in the state of Colorado. The crash occurred on a curved portion of a three-
Vehicle
lane interstate roadway. The Ford Escape lost control on an interstate highway and struck a concrete barrier on the right side of the roadway. The impact resulted in sufficient longitudinal deceleration of the Escape to command the deployment of the frontal air bag system and actuation of the driver's seat
agentOf
affectedBy
timeOf
locatedAt
Vehi...
Eve...
Date
Person
C...
Veh...
belt pretensioner. The vehicle rotated out from the initial wall impact and was subsequently struck by a 2013 BYD Qin BYD Qin pulling a single trailer.
hasAttribute
educatedAt
Age
Quant...
Per...
Orga...
The restrained 48-year-old 48-year-old male driver of the Escape appears to have sustained a minor facial injury.



Knowledge Graph

Query the web for rich connected entities

Learn More

AI:X

Extract structured data from any URL with AI Extractors

Learn More

Crawlbot

Scale up your extraction to 10,000s of domains with Crawlbot

Learn More

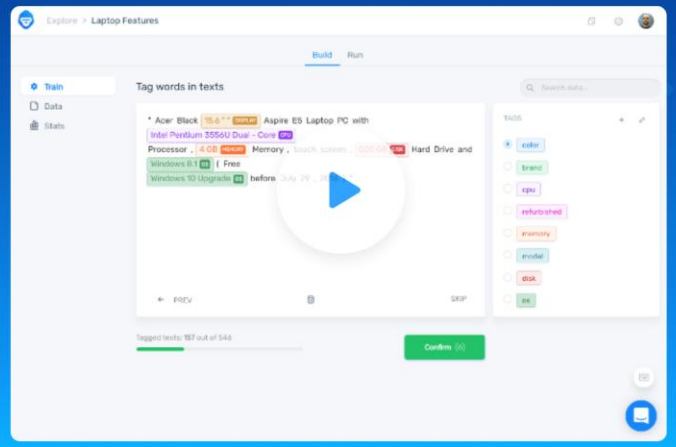
MonkeyLearn

SOLUTIONS CUSTOMERS PRODUCT PRICING RESOURCES Log in

Build High Accuracy Text Analysis

From topic classification and sentiment analysis to entity extraction. Train and integrate custom machine learning models in a matter of days, not months.

SIGN UP FREE → Request a Demo



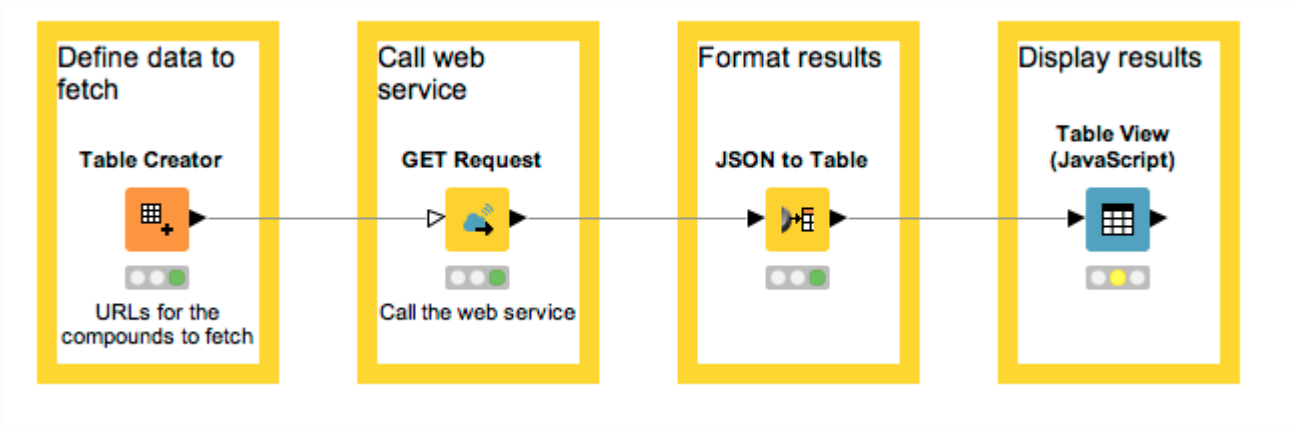
Customized text analysis

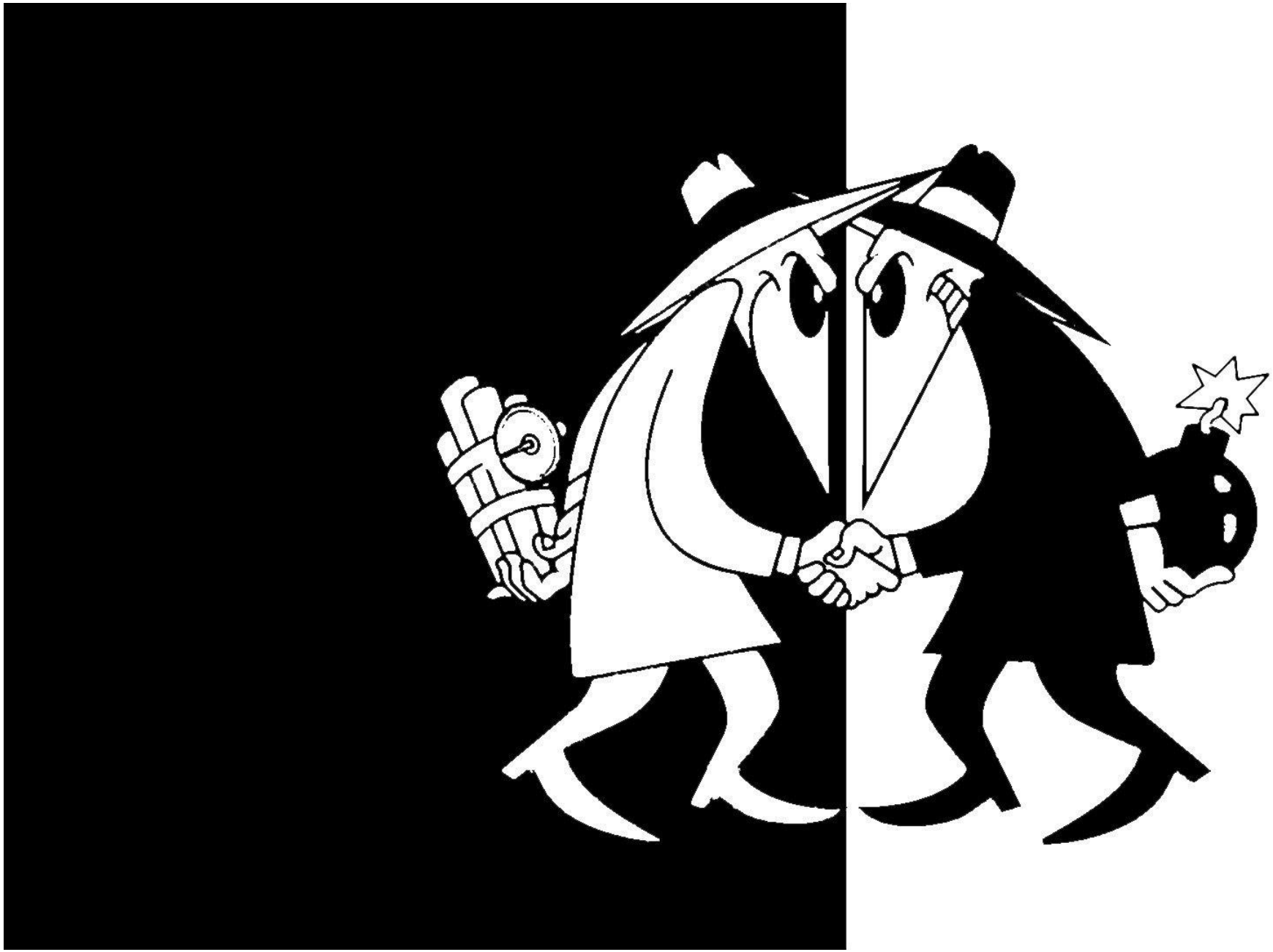
Obtain high accuracy results by training your custom

Uncomfortable with APIs?

If APIs sound daunting to you, here's a Knime Analytics workflow that will make it easier for you to pull data from an API and format it into a table.

https://www.knime.com/blog/a-restful-way-to-find-and-retrieve-data

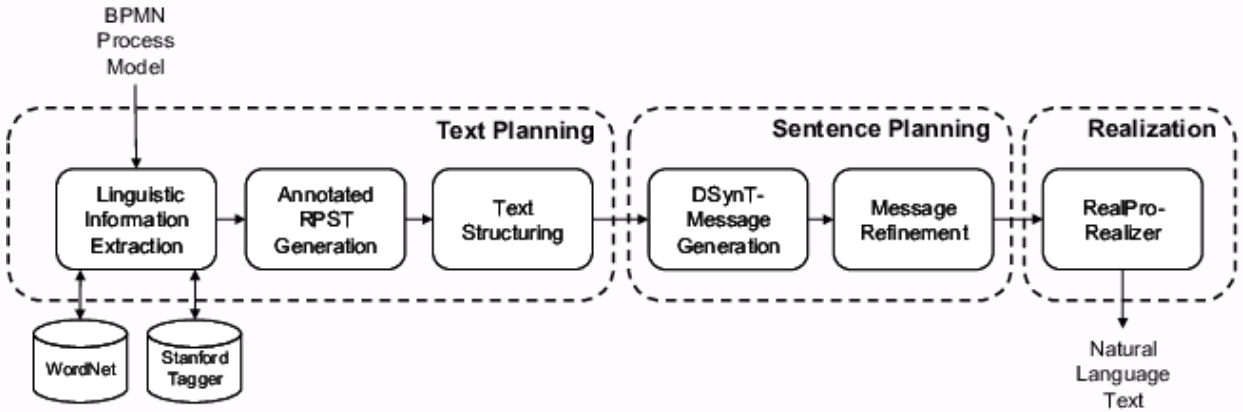


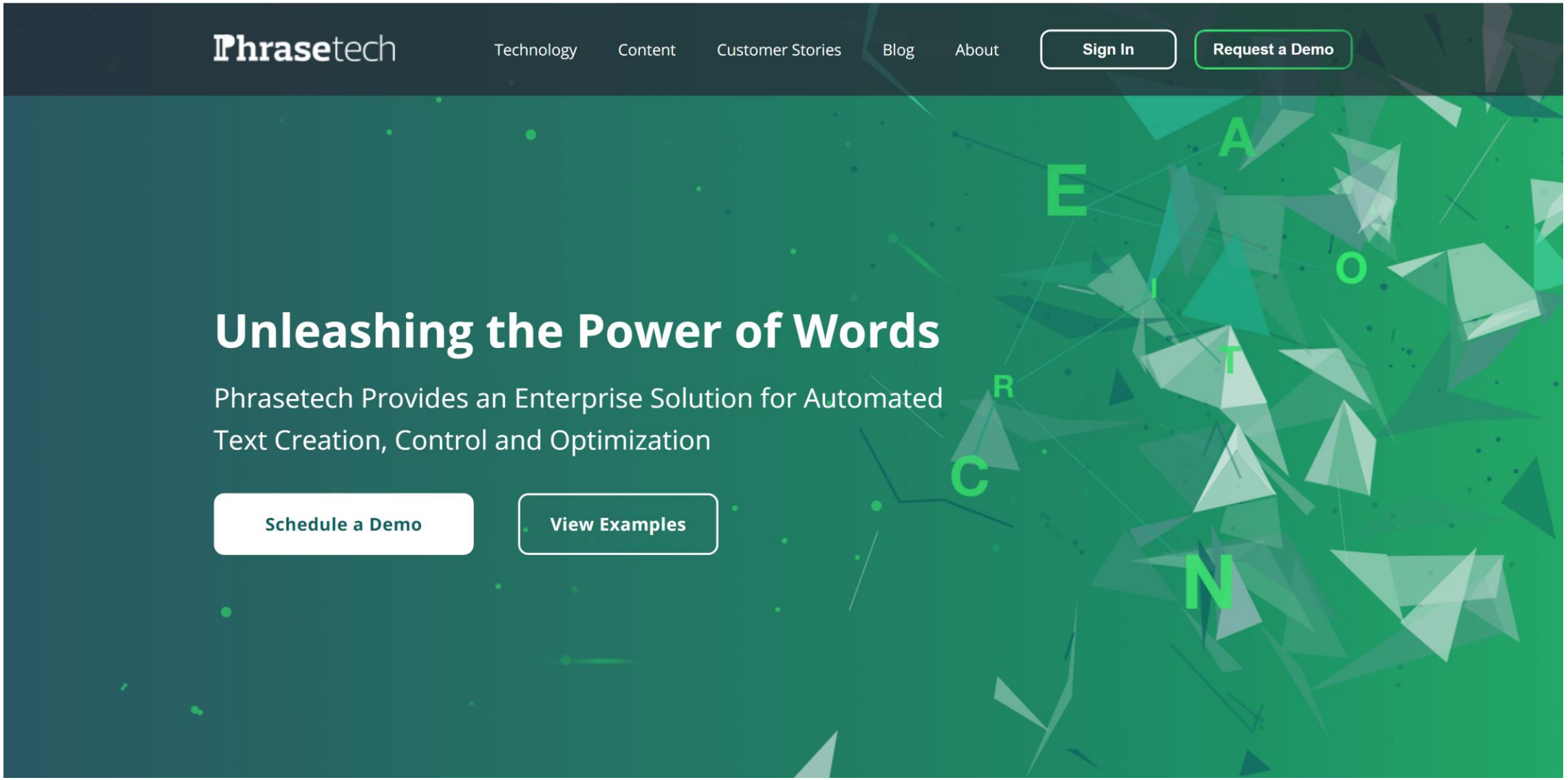


Natural Language Generation is the Next Big Thing

I suspect that in the near future, a lot of this will be available to automate.

There are already solutions that are more than just sophisticated content spinners.




Phrasetech
Technology
Content
Customer Stories
Blog
About
Sign In
Request a Demo
Unleashing the Power of Words
Phrasetech Provides an Enterprise Solution for Automated Text Creation, Control and Optimization
Schedule a Demo
View Examples

Wrapping Up

Bringing it Home!

How to do effective Technical Content Optimization

A. Determine Search Engine Expectations

B. Optimize your Content

C. Test All the Things

D. Win the SERPs

I'm Mike King.

(@IPULLRANK)

We are iPullRank.

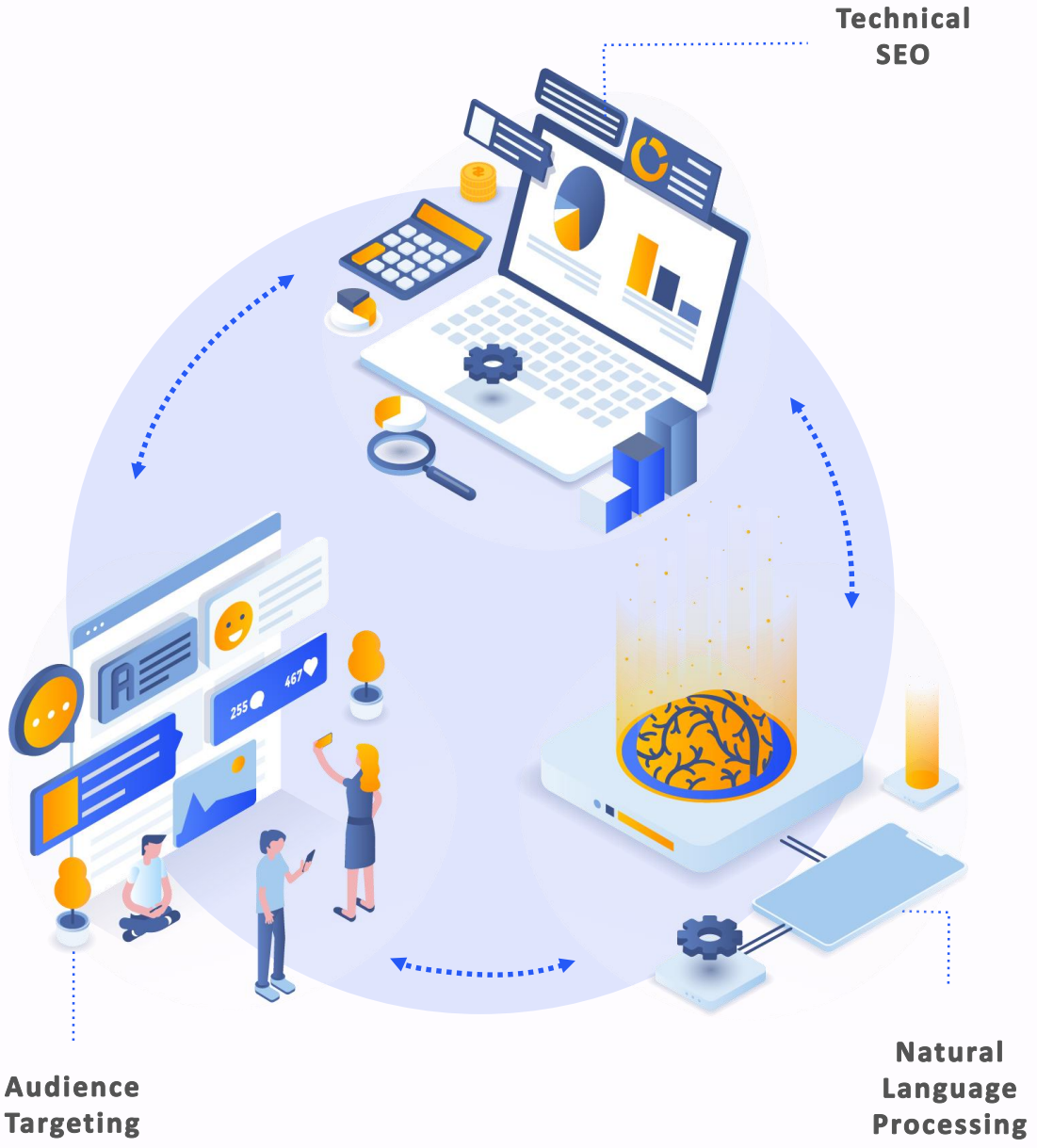
(@IPULLRANKAGENCY)

iPULLRANK

This is Our CEO.

(And, I'm #ZORASDAD)

At iPullRank, we're incorporating Natural Language Processing into our workflows to better understand the expectations of Google.



Thank You | Q&A

Mike King
Managing Director

Twitter: @iPullRank
Email: mike@ipullrank.com

iPULLRANK