Readme

"From Page Rank to Rank Brain" is an essay that attempts to "decloak" as well as "update" public knowledge about Google a.k.a. Alphabet's ranking algorithm. This text has then been altered through [3] "translation" processes.

Drawing on Constant's collection of scripts,[1] the first translation used "encryptionlinessha1.py" that "provides the ultimate reduction (although at the expense of human as well as machine legibility) by encrypting every line of your text as a 128-bit hash value. Each hash value can of course be reversed again if you try to match it with every single line of every single text existing."[2] The second translation uses a little python script called the "The Synonymizer" that corrupts your writing style by swapping out words in your text with randomized synonyms from WordNet.[3] With the third translation, the text was first read with the "text to speech" voice of "Alex" and saved as an audio file, then uploaded to "gentle", a robust yet lenient "forced aligner" built on Kaldi.[4] Forced aligners are computer programs that take media files and their transcripts and return extremely precise timing information for each word (and phoneme) in the media. How does it work? "As in all of these Machine Learning cases, you have to follow the data."[5]

Notes

[1] https://gitlab.constantvzw.org/machineresearch/reduction/tree/master/filters (Proximus NV ⇸ OVH SAS)
[2] In cryptography, SHA-1 (Secure Hash Algorithm 1) is a cryptographic hash function designed by the United States National Security Agency and is a U.S. Federal Information Processing Standard published by the United States NIST in 1993. SHA-1 produces a 160-bit (20-byte) hash value known as a message digest. A SHA-1 hash value is typically rendered as a hexadecimal number, 40 digits long.
[3] "Note:it may also corrupt the meaning of your text which replaces 'choice words' with synmyms." WordNet:
http://wordnet.princeton.edu/. (Proximus NV ⇸ Hurricane Electric, Inc. ⇸ Princeton University)
Thanks 2 Dave Young
[4] http://lowerquality.com/gentle/ (Proximus NV ⇸ Level 3 Communications, Inc. ⇸ Advania hf. ⇸ Thor Data Center ehf)
[5] In this case, it's the CALLHOME corpus, which is 120 unscripted 30-minute telephone conversations between native speakers of English in the 1990s.
https://catalog.ldc.upenn.edu/LDC97S42.
(Proximus NV ⇸ Cogent Communications ⇸ University of Pennsylvania) Thanks 2 Robert M. Ochshorn

---

FROM PAGE RANK TO RANKBRAIN

The Synonymizer

Page rank to heighten page concept page social station receive information technology basis in scientific citation index a form academic hierarchy hour angle now be transplant as a conceptual paradigm way detect information information be prioritize u.

Eponymous page membership algorithm was develop in be basically a popularity contest based on vote. Vitamin a connect do a node a high gear rank give birth more value a link arrive a node low rank. Scheme therefore arrogate two mark each information technology estimate value message information technology hub estimate value information technology connection to other page. Five hundred.

Be hush direction google search organize internet by crawl specify importance a web site establish on son it frequently other ride liaison to tons other measurement. Google search emphasis be to keep attention drug user to receive click on high effortlessly. However as gillespie point claim bring embody opaque vary divers criterion code algorithm constitute generally not evenly or.

Based on location search seeker equal through a set criterion. Two not only be creator content web page keep in arrest aside search merely traverse different or decide rank associate in nursing individual page. Largely through invert a whole engine industry receive develop approximately algorithm to figure out information technology recipe or signal. Past eighteen google take constantly pluck proprietary check about two hundred ingredient or indiana recipe. Three equal typically component be tied to such as word on a connect orient astatine a a page be on a secure waiter sol on.

Can besides be bind to a such as a searcher equal locate or search shop history. Discussion in extra domain registration duration outbound liaison quality cost some other exercise or. One major change in to core algorithm page crying be enable associate in nursing improvement indiana gather information or rather precisely sorting. Was associate in nursing update was implement in be think low enabling high choice result to heighten.

In april google launch update try to capture now devalue spam alternatively demote entire site as september update in very clock as part core algorithm. Five to component engine own suffer it part penguin lesser panda might be anoint filter boast pump plunge inch august was bombastic overhaul. Introduction a brand newly engine emphasis have shift to contextual less now about more about purpose behind it semantic capability be be at impale. Previously sealed cost astatine moment about other word in prison term meaning. Inside field connect search question web cost reflected along integrate voice trip inquiry. Hummingbird be new google locomotive belated surrogate part be then.

Launch about early it apparently citizenry cost searching even though may have not enroll exact. Exist rumor to be third most authoritative subsequently radio link contentedness guess manipulation a aside lend oneself synonym or stem. Complexity level question get rifle result indium associate in nursing improvement index network text file.

Question have besides change be now not entirely merely besides give voice sentence be deem question. Motivation to be translate to a certain to or to in order to be work analyze.

Nine reciprocal adaptability between drug user interface accept be control by former research. Therefore it be probable google arrogate complex question to group similar interest in arrange to. Ten learn be not always neutral.

Build aside use aside bias hang-up off on engineering.

Code can discriminate.

June constitute be implement every google search question industry speculate sum up content.

Heart murmur embody algorithm be or as it be mistake information technology smother.

According to google algorithm teach be feed historical batch search it make prediction.

Algorithm embody make remake in every case consumption every every change instrument.

Hertz be constantly repeat prediction be late adaptation adam hot.

Twelve there equal not computer herculean or fast or data hardening constitute besides humble to carry out type test.

Nowadays calculation be circulate over many enable footstep research to quicken.

Progress indiana technology help a configuration or come in concert different capability versatile through model parameter.

Finally or in case exist able to through repetition.

Be homo curator in wholly be a casing to be do work logic algorithm not only shape exploiter merely besides contribute user to internalize norm.

Wonder then be to extent be there human adaptation to algorithm in trickle or much dress algorithm affect homo learning merely besides discrimination merely besides agency toilet be catching.

FROM PAGE RANK TO RANKBRAIN

Alex

mhm i'm from a drunk to rent graeme p.r. a. eagles will be close to be pretty or do you want to be do one plus plus p.r. e. d. and see you know and the concept of page rank hasn't spaces into scientific antidrug citation index as sci most form of academic hierarchy that has now been grafted as of conceptual power done for the way we find information and how that information is prioritize for us the of you know much pay drank algorithm listed reluctant nineteen ninety eight and it's basically well you a popularity know popularity contest based on the votes a plane coming from and old with the high rank has more value than allin coming from you know with little rank the skiing there for assigned to the scores for each page it just already we're just amazed the value of the content of the page and it's hot value oh you and would we're just amazed to two value dollars movements links to other pages presently he were search they still they wait global search organizes the internet by crawling and the indexing witch determines the importance of the website based on the words it contains how often other sites link to it and dozens of other measures we've google search the emphasis is to keep the attention of the user been to have then click on the higher ratings effortlessly allow her as gillespie points out the exact words in their own pay them very very diverse users the criteria the code of algorithms or in the roads are generally obscure but not who you are from everyone based on user's histories location and search terms the surgeries personalized through instead of criteria not only are the creators of content of web pages get it checked by search engines but they're tracking of different factors are signals determine the ranking of an individual and she mostly verse engineering ball search engine optimization as c.e.o. industry has developed around gaming the algorithm just figure out it's recipe or signals you you say you know during the past eighteen years rule has constantly between their proprietary algorithm didn't dating around two hundred ingredients are signaled in the recipe saying those are typically factors that are tied to contend so just the words on a page teach the links pointing out of age a whether i'm pages on a server and so on and they can also be tied to use your soldiers whoever surgeries located where they're surgeon browsing history legs content he were density words and bowled duplicate content and domain don't mean registration to ration and our dog late quality or some other examples of factors or has one of the major changes in two thousand ten to the core algorithm of page rent 'cause the caffeine update witch enabled an improvement in the gathering of information or indexing actually instead of just sort sorting and the wasn't update that was implemented in two thousand eleven that down rank sites which are considered lower quality enabling higher quality pages terrorize enable two thousand twelve global are still and one of dig that attempts to couch ouch sites that now do you value go spam instead of being modine but just being the ranks of the entire site and as of september thirtieth two thousand sixteen bucks agent real time as part of the core algorithm analogous to the components of engine that has added part's replaced we're we're paying women and there might be your oil filtered and gas pump respectively the launch of humming birds in august two thousand thirteen was global smart just overall since two thousand one way the introduction of a brand new engine the emphasis has shifted to the gym to actual it's less now how about the key word and more about the intention behind it the semantic capabilities are what are at stake whereas previously certain key words for the focus at the moment it's about the other words in the sense and spend their meaning within this field of semantics search the rationale reality any linking surgeries and quit where does documents is reflected with their knowledge grass along with conversational church that incorporates voice activated inquiries they've humming bird is the new rule engine from two thousand thirteen the latest replacements part is that wrangler i am launched around early or eighteen teens thousand fifteen it ostensible they interprets what people are searching for even though they may have not entered the exact key words rank brain is rumored to be the third most important so you know after legs and content words and then first or use of the key word but by that's flying why i'm sending sentence them sports or them unless the complexity level of it queries has gone up resulting in an improvement of indexing web documents user's queries have also changed and er now not only key words but also multi words phrases and sentences they could be being wrong long tail inquiries things need to be translate into a respect for 'em and big it was too specific or uncommon into common in order to be processed and then realized this reciprocal adaptability between the years version interface has been verified by previous research should be they're there four for and this it's probably problem all that rule assigned seats complex queries two groups with similar interests in order to collaborative results or a machine learning algorithms are not always neutral they're built by humans and use by humans and there are biases rub off on the technology code can discriminate as a two thousand sixteen rank brain is being implemented for every google search aquarium b. s. radio industry speculate summarizing the pages content to them or is that the algorithm isn't dowding are are learning as it were affirmative from people's mistakes and the surroundings according to a movie algorithm learns off line being fed historical bashed search just bring which it makes for their actions and they all the rhythm for me it and we made in every incidence of their use because every click everywhere he changes the tour incrementally vicious cycle it's constantly repeat it under the predictions are correct the latest versions of rag brain go live previously there or not computers powerful or fast enough for the data sets we're too small to carry out this type of testing nowadays book the connotation augmentation is distributed over many machines enabling the pace of the research to click on this progress in technology facilitate some consolation are coming together of different capabilities from various sources through models and parameters eventually the subject object or learn or or didn't in this case the algorithm is able to predict threw reputation whereas the human generator in all of us they were so case to be made to the working logic survey's algorithms not only shady user practices but also lead users to internalize if they're norms and priorities the harsh invent as to what extent is they're human interaction jew algorithms in this culture a generation process out much drought a vacuum and learning and whether or not only discrimination but also and she can be contagious except to do you um mind

---

FROM PAGE RANK TO RANKBRAIN

encryptionlinessha1.py

a1e1c6be7dda0e1637f704febcc46fc0d7c11bd8
ba8ab5a0280b953aa97435ff8946cbcbb2755a27
18435c50442c4752e18ce421cf5b6d7637b14db0
ba8ab5a0280b953aa97435ff8946cbcbb2755a27
7e788fb56bd7b3218df2d92660726f809a3ff4bf
0504b125ca32b43609178b54615d0e977133aa45
e165e4d16cfd2cf0a8ae8840e1ddbd2321822 9fb
767de1876198f54c5d89bf6fd400438597fe962c
ba8ab5a0280b953aa97435ff8946cbcbb2755a27
089035a38e79247542f9f947c526b46dca76aa8f
ba8ab5a0280b953aa97435ff8946cbcbb2755a27
2f13fd3f48e6332dd81c38752e23bf0d291f a222
ba8ab5a0280b953aa97435ff8946cbcbb2755a27
302c19c13e1da293bae24441a322fd191b5ea7f2
c2724665d03ca72a4368ebea907285d8c4c87f10
c59ef8e41e404f221913dce8312bbadd86fb7722
93b17cbe3d16a4ace1839c1aca76d6aac8217f2a
b46eb56e7d5d55999cc988914d3d49aec67ca85
ba8ab5a0280b953aa97435ff8946cbcbb2755a27
d73a4a559bof0598ebf272f a94331339f051c4
ba8ab5a0280b953aa97435ff8946cbcbb2755a27
8fa99a8df73b92db3ca267d5f2efe46055cba13
ba8ab5a0280b953aa97435ff8946cbcbb2755a27
c6b1f0f6edobab101f7bb77b2afbc35a5a43013c
2ed26a3a9b069e3ad443d066fc6823a73524c9e4
docbfc8101b6f8c71a755a2f2556ca5396f8e1c3
2a1b3d3187e202cdb85d9c9af2946d82cd1048bd