



FOR STUDENTS : ALL THE INGREDIENTS OF A GOOD ESSAY

☰ Menu



Essay: Review and Summary of Cornell's "PageRank" Algorithm

January 7, 2019 by admin

Essay details:

- **Subject area(s):** Information technology essays
- **Number of words:** 1251
- **Price:** Free download
- **File format:** PDF

Overall rating: **0** out of **5** based on 0 reviews.

Text preview of this essay:

The full version of this essay has 1251 words and is available to download in PDF format above.

Everyday, billions of entries are typed into search engines such as Google, Yahoo, Bing, etc. At an incredulous speed, thousands of suggested websites and journal articles are generated that are most closely related to the keywords entered by the user. However, deciphering an algorithm that would take certain keywords and output the "best" suited websites for the user's interest proved to be no easy task. After experimentation with certain methods, employees at the Google company settled on an algorithm that utilized the skills and ideas of linear algebra in order to generate websites from keywords.

Until the early 1990's, search engines used text based ranking systems in order to decide which pages were most relevant to a given query. These text based ranking systems would take the user-entered keyword and count the frequency with which that word appeared in websites across the Internet database. The pages that contained the keyword most often were labeled as the most relevant, and they would appear at the top of the "suggested" websites list. The pages that had less frequency of the word would appear towards the end of the list, and the pattern continued this way. However, there were a number of problems that came with this approach. For example, say I wanted to learn more about Purdue University, and I typed the word "Purdue" into my search engine. Ideally, a source listing vast and comprehensive information regarding the university

should be the prioritized web page. Yet, what if I had hypothetically created a website with solely the word "Purdue" listed thousands of times, accompanied by no other information? This would likely be my top page suggestion, although it contains minimal to no actual material about the school. The underlying problem with the text based ranking systems is that it does not correctly measure the relevance of the web page to the user's interests. One page may contain a much greater deal of vital information than another and still yet mention the actual keyword name less.

Researchers harped earnestly over the topic, eager to find an algorithm that would prove more efficient than the current text based ranking system. In 1998, Stanford graduate students Larry Page and Sergey Brin proposed the "PageRank" algorithm, an alternate method of effectively matching websites to keywords. The PageRank algorithm was based off of the assumption that a website's relevance is directly correlated to the quality and the amount of the pages that link to it. Essentially, more important websites are likely to receive more links from other websites. For example, if we create a certain web page a and include a hyperlink to web page b, than we consider b as important and relevant to our topic. If there are lots of pages linking to b, it seems to be the common consensus that b will store meaningful information and will likely be a top suggestion. We can iteratively assign a rank to each web page based off of the ranks of the pages pointing to it. In understanding this algorithm and its application to linear algebra, it helps to picture the world wide web as a directed graph network, with the web pages as nodes and the edges as links between them. Let us conduct an example of the "PageRank" algorithm by imagining a miniscule Internet consisting of just four websites connected to each other through hyperlinks on each page. For simplicity, we can call them "Page 1", "Page 2", "Page 3", and "Page 4". They are each referenced in the way described by the picture below.

We can translate this picture of interconnected websites into a directed graph with each website as a node. Edges are created using the hyperlinks listed on each web page. For example, if I have a link to page 1 on my current page 3, I will create an edge stemming from page 3 and directed towards page 1. After all of these relationships have been correctly identified and constructed, the directed graph should look like the picture below, with numbers 1-4 corresponding to pages 1-4.

The final step in creating this directed graph is to assign edge weights to each edge. Edge weights can be described as "the likelihood of an edge forming", or in our case, "the importance of the edge relative to the node it's directed from". We will operate under the condition that when on a certain webpage, you must choose a certain hyperlink to navigate to and to each with equal probability. For example, because my web page #1 has links to pages 2,3, and 4, the probability of going from 1 to each of the other web pages in a uniform $1/3$ rd. Because node 2 only has links to pages 3 and 4, the edge weights assigned to each respective edge is $1/2$. In other words, node 2 transfers half of its importance to node 4 and the other half to node 3. Filling in these edge weights will yield a graph corresponding to the figure below.

Once more, this graph can be translated into a transition matrix between the nodes. This transition matrix is an alternate way of describing the relationships between nodes and will make it convenient when calculations are needed to be executed. Notice how there are all values of 0 across the diagonal. This is because it is impossible for a web page to link to itself, corresponding to an edge weight of 0. All other entries in the matrix are listed as they were in the directed graph. The nodes listed horizontally at the top are "from" nodes and the vertical nodes listed to the left are the "to" nodes (i.e., the edge weight from node 3 to node 1 is 1, due to page 1 being the sole hyperlink on page 3).

Now that the transition matrix has been calculated for the directed graph, we can begin calculating the overall "importance" of each of the web pages. Intuitively, the overall importance of each web page is the importance of the pages it is directed from, all summed. In other words, web page 3 directs all of its importance to web page 1, while page 4 only delivers half of its importance. Therefore, the importance of web page 1 can be calculated by summing the importance of webpage 3 and one half of the importance of web page 4. Similarly, the importance of web page 2 can be calculated by dividing the importance of page 1 by 3. By continuing this application throughout the matrix, we can come up with four individual equations that

relate the "importance" of the four web pages.

$$X_1 = X_3 + \frac{1}{2} * X_4$$

$$X_2 = \frac{1}{3} * X_1$$

$$X_3 = \frac{1}{3} * X_1 + \frac{1}{2} * X_2 + \frac{1}{2} * X_4$$

$$X_4 = \frac{1}{3} * X_1 + \frac{1}{2} * X_2$$

where X_1, X_2, X_3, X_4 denote the importance of the individual web pages.

Solving this system of equations is equivalent to asking for the solutions of the equations below.

From our knowledge of linear algebra, calculating the eigenvectors corresponding to a certain eigenvalue will give us the solution to the system of equations. In calculating the eigenvectors corresponding to an eigenvalue of 1, we know they are in the form of.. (below)

Because eigenvectors are just scalar multiples of one another, we can essentially choose one to be our PageRank vector. Let us choose our PageRank vector as that with which the sum of every entry is equal to 1. Essentially, we are just "normalizing" the matrix. Each page has a different "weight" or "importance" ; we want to standardize this with respect to a maximal value of 1. Adding up each of our entries gives us the number 31. We then divide each entry by that number to give us our normalized vector as desired.

About admin

View all posts by admin

About this essay:

This essay was submitted to us by a student in order to help you with your studies.

If you use part of this page in your own work, you need to provide a citation, as follows:

Essay Sauce, Review and Summary of Cornell's "PageRank" Algorithm. Available from:
<<http://www.essaysauce.com/information-technology-essays/review-and-summary-of-cornells-pagerank-algorithm/>> [Accessed 03-07-19].

Don't like this essay? Find another:

Review this essay:

Please note that the above text is only a preview of this essay. The full essay has 1251 words and can be downloaded free in PDF format, using the link above.

Name *	<input type="text"/>

Email	<input type="text"/>
Rating *	☆☆☆☆☆
Comments (optional)	<input type="text"/>
	<input type="submit" value="Submit"/>

Latest reviews:

[www-essaysauce-com-information-technology-essays-review-and-summary-of-cornells-pagerank-algorithm--.pdf](#)

■ Information technology essays

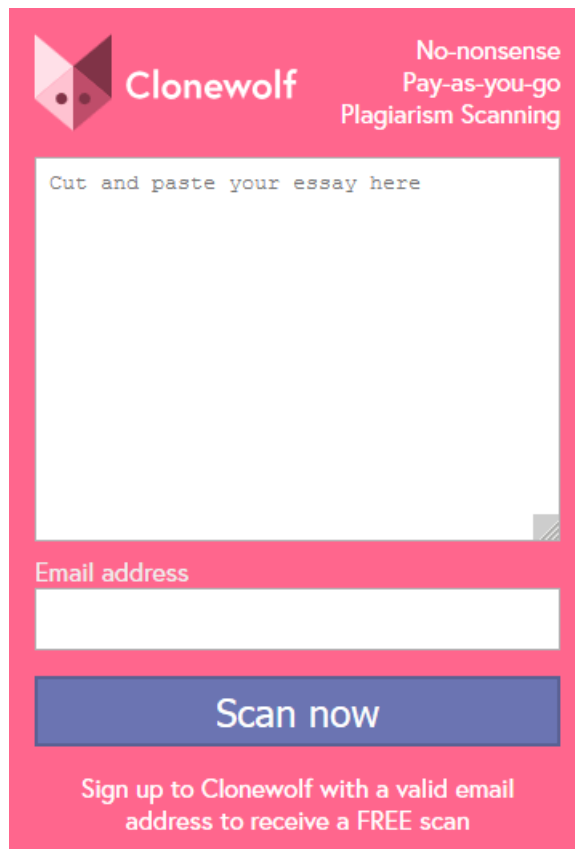
< Ellen DeGeneres

> Henrietta Lacks

Search for student essays:

About EssaySauce, the student essay site:

EssaySauce.com is a free resource for students, providing thousands of example essays to help them complete their college and university coursework. Students can use our free essays as examples to write their own.



The image shows a pink-themed interface for Clonewolf Plagiarism Scanning. At the top left is the Clonewolf logo, a stylized wolf head in shades of pink and purple. To its right, the text reads "Clonewolf" in a bold, sans-serif font, followed by "No-nonsense Pay-as-you-go Plagiarism Scanning" in a smaller font. Below the logo and text is a large white text area with the placeholder "Cut and paste your essay here". Underneath this is a white input field labeled "Email address". A prominent blue button with the text "Scan now" is centered below the email field. At the bottom of the interface, a pink banner contains the text: "Sign up to Clonewolf with a valid email address to receive a FREE scan".

Latest student essays:

Force to temperature

Germination and seedling emergence

Gross primary production, community respiration...

Leaf area index estimation from allometric ...

Photosynthesis Lab

SILICON

Fossils

Seaweed

METHODS OF BUSINESS RESEARCH

The Relationship Among Enactment, Organizing, and Sensemaking

Student essay categories:

Business essays

Economics essays

Education essays

Engineering essays
Finance essays
Health
Health essays
History essays
Information technology essays
Law essays
Literature essays
Marketing essays
Music Essays
Photography and arts
Photography and arts essays
Politics
Politics essays
Psychology essays
Religious studies and Theology essays
Science essays
Sociology essays
Zoology essays

Average review:

Overall rating: **0** out of **5** based on 0 reviews.

Q: Is EssaySauce.com free?

Yes! EssaySauce.com is a completely free resource for students. You can view our **terms of use** here.

Why use Essay Sauce?

The brightest students know that the best way to learn is by example! EssaySauce.com has thousands of great essay examples for students to use as inspiration when writing their own essays.

Is Essay Sauce completely free?

Yes! EssaySauce.com is a completely free resource for students. You can view our **terms of use** here.

Info:

[About](#)

[Content policy](#)

[Essay removal request](#)

[Privacy](#)

[Terms of use](#)