

# Simulator for Calculating PageRank by Javascript

H. Barboucha  
LABO MATSI, ESTO, B.P 473,  
University Mohammed I OUJDA, MAROC.

M. Nasri  
LABO MATSI, ESTO, B.P 473,  
University Mohammed I OUJDA, MAROC.

## ABSTRACT

The most famous search engines today, Google, Yahoo, Bing and Baidu utilize the popularity in their relevance criteria. In the case of web sites the number of pages is reduced, the determination of this index is more simple. In the case of the sites or the number of web pages is very important, the determination and the relevance of the page is difficult (if not impossible). To address this problem, we developed a program in JavaScript based matrices, which simulates the algorithm for calculating the relevance of web pages automatically.

## Keywords

Matrice, HTML5, CSS3, JavaScript, PageRank, SEO, Algorithmes, techniques.

## 1. INTRODUCTION

Search engines have developed methods for automatic sorting the results of research on the web. Their goal is to show in ten to twenty first answers the documents that best suits the subject. Google ranks pages search engine through the combination of several factors, the main door PageRank name. The PageRank algorithm calculates a popularity associated with each Web page. This is the index that is used to sort the result of a search keywords. The index is defined as: "The popularity of a page is all the greater that it has a large number of popular pages that link to it." So to find the index of a page, you must first know the index pages that link to it .. How to calculate this index? To answer this question, in a first part is an introduction about how Google's algorithm, a second part as a description of the matrix method to calculate the relevance of web pages, the third part is devoted to the development of a javascript program that automatically calculates the relevance of web pages. To conclude our work, we advocate recommendations for businesses.[1]

## 2. DEVELOPMENT OF GOOGLE PAGERANK ALGORITHM

PageRank is an approach used by Google to rank websites in the results of the search engine Google. The method was introduced by Google co-founder Larry Page, hence the name PageRank. This is indeed a way of calculating the relevance see the popularity of a page according to Google.

PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of the value of an individual page. In essence, Google interprets a link from page A to page B as a vote, by page A for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote.

Votes cast by pages that are themselves "important" weigh more heavily and help to make other "important" pages.[2]

The importance of any page comes largely relevance to other pages. The damping factor adjusts the derived value downward to limit the weight of votes.

X receives a page links emitted by pages  $y_1 \dots y_n$ , the parameter  $d$  is set between 0 and 1.

(We usually give to  $d$  0.85 of the value.)

A late calculate PageRank is based on the following formula:

$$f(x) = \frac{1-d}{N} + d \sum_N \frac{f(y_i)}{L(y_i)}$$

Where:  $N$  = number of pages.

$d$  = a damping factor.

$x, y_i$  pages of the website.

$L(y_i)$  = number of outgoing links  $y_i$  page

$f(y_i)$  =  $y_i$  PageRank

In addition to counting the number of hyperlinks to a page's PageRank method is based on the importance, while at the beginning of the method it took into account only the quantitative aspect ie the number of link, this aspect quickly showed its limits.

## 3. MATRIX METHOD :

PageRank is a mathematical property of the curve that describes a network of pages and their links. We can not really understand this concept, unless we are directly confronted with math involved. There are a number of ways to solve mathematically equivalent to the PageRank of a graph showing the connecting structure of a web page network. A practical approach is to use matrices.

When using arrays to compute PageRank we measure the centrality of eigenvectors of a matrix according to the modified network vectors. The calculation of the index of relevance of the iterative analytically becomes complex if not impossible to achieve if the number of pages is great, because there are a number of trick to solve this equation. The most practical approach is to use matrices. .[3][4]

### 3.1 Sample Matrix

Consider a diagram of the following web, consisting of 12 pages connected by links as shown in the following figure:

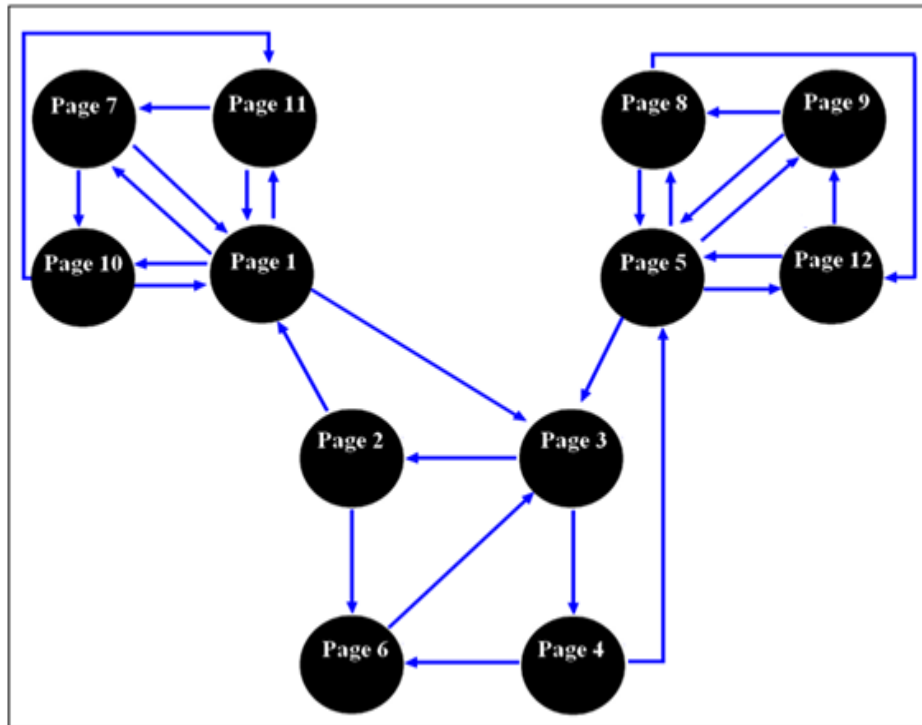


Fig 1 : webpages Network and their connections to each other

Our graph is a representation of a web page network and how they are connected to each other.

The inter-page links diagram above can be represented in the matrix model. When considering a square matrix of size equal to the number of pages: 12x12 describing our web pages Network presented by the following table:

Pages	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	1	0	0	0	1	0	0	1	1	0
2	1	0	0	0	0	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0	0	0
4	0	0	0	0	0	1	0	0	0	0	0	0
5	0	0	1	0	0	0	0	1	1	0	0	1
6	0	0	1	0	0	0	0	0	0	0	0	0
7	1	0	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	1	0	0	0	0	0	0	1
9	0	0	0	0	1	0	0	1	0	0	0	0
10	1	0	0	0	0	0	0	0	0	0	1	0
11	1	0	0	0	0	0	1	0	0	0	0	0
12	0	0	0	0	1	0	0	0	1	0	0	0

Matrix 12 pages

It is clear that it is difficult (if not impossible) to calculate the ranking hand for a large number of pages, so to make things

easier and use information held for this study, has programmed a simple JavaScript pagerank simulator based on matrices, which simulates the PageRank algorithm and for establishing the automatically calculated.

The matrix A is then:

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Along With:

$$A_{ij} = 1 \quad \text{If the page } i \text{ sends a link to } j$$

$$A_{ij} = 0 \quad \text{If the page does not emit link to } j$$

This grid is called the adjacency matrix of a graph illustrating a web page network.

According to the data of the table above now we will translate it into a computer programming language. In this example we use JavaScript. It has a number of classes that enable us to form a matrix and a vector in a number of dimensions. [5]

In programming terms, a similar matrix array of arrays.

### 3.2 Choosing the JavaScript language

JavaScript is a scripting programming language mainly used in interactive web pages but also for servers. It is an object oriented language prototype, that is to say that the basic language and key interfaces are provided by objects that are instances of classes, but which are each equipped with constructors to create their properties, including a prototype

property that allows to create custom objects heirs. In addition, functions are first-class objects.

## 4. SIMULATOR PAGERINK JAVASCRIPT

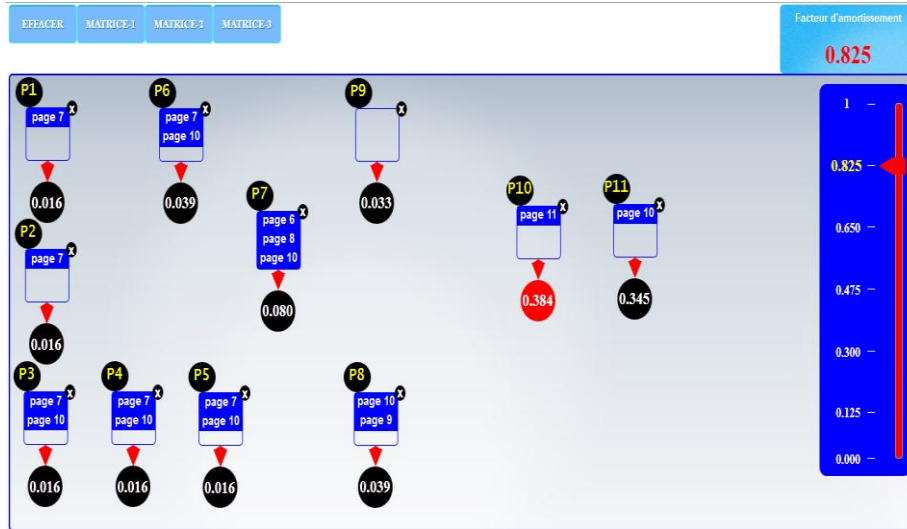


Fig 2 : The graph above describes a webpages network. Each page lists its links to other pages

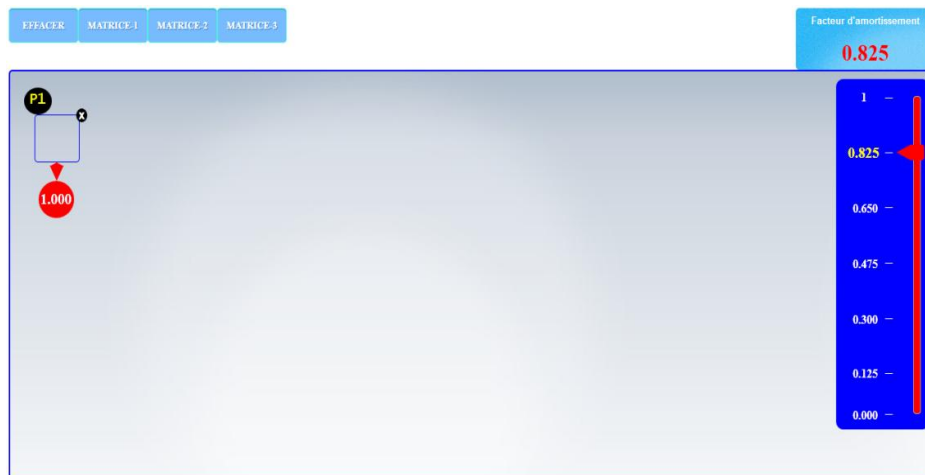
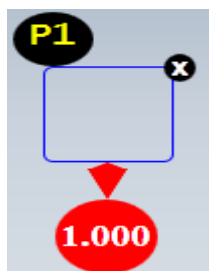
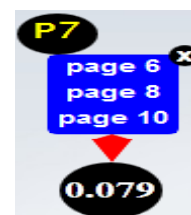


Fig3: Interface simulator

To add a page, click on the "page space" (The above graph). This creates a page represented by the icon of the page you see below:



Page without links



Page with links to pages 6, 8 and 10

## 5. STOCHASTIC MATRIX

To calibrate the probabilities are called upon damping coefficient. When using a damping factor of 0.85 we obtain the resulting matrix:

PageRank can also be regarded as a representation of how likely it is that a person randomly clicking on the link or on a particular page, we need to convert the matrix to represent the likelihood of representation.

For example, since Page 5 has 3 links on it, there is one in three chance that we will randomly click on one of these links.

Type Matrix is known as stochastic matrix, and in this case we will seek row stochastic matrix. In other words, we want each line to add up to 1. We will return represents the top of the network link to the form that takes into account the probability that a random user would navigate to a specific page.[6]

$$\begin{pmatrix} 0.01 & 0.01 & 0.26 & 0.01 & 0.01 & 0.01 & 0.26 & 0.01 & 0.01 & 0.26 & 0.26 & 0.01 \\ 0.51 & 0.01 & 0.01 & 0.01 & 0.01 & 0.51 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.51 & 0.01 & 0.51 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.01 & 0.51 & 0.51 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.26 & 0.01 & 0.01 & 0.01 & 0.01 & 0.26 & 0.26 & 0.01 & 0.01 & 0.26 \\ 0.01 & 0.01 & 1.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.51 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.51 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.51 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.51 \\ 0.01 & 0.01 & 0.01 & 0.01 & 0.51 & 0.01 & 0.01 & 0.51 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.51 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.51 & 0.01 & 0.01 \\ 0.51 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.51 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.01 & 0.51 & 0.01 & 0.01 & 0.01 & 0.51 & 0.01 & 0.01 & 0.01 \end{pmatrix}$$

Is assigned to each page's relevance index  $P_i$   $PR(P_i)$ . This produces a linear system of equations twelve to twelve unknowns  $PR(P_1)$ ,  $PR(P_2)$ , ...,  $PR(P_{12})$ , which is written in the matrix form.[8].[9]

$$S = K + a * A * S$$

Along with:

- A : Matrix twelve rows and twelve columns,
- S : The vector solution that includes the unknown  $PR(P_i)$
- K : Vector twelve lines such as:

$$\begin{pmatrix} \frac{1-a}{N} \\ \frac{1-a}{N} \\ \frac{1-a}{N} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1-a}{N} \\ \frac{1-a}{N} \end{pmatrix}$$

- a: damping factor
  - N: Number of inbound link
- The linear system is then:

$$S = \begin{pmatrix} \frac{1-a}{N} \\ \frac{1-a}{N} \\ \frac{1-a}{N} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1-a}{N} \\ \frac{1-a}{N} \end{pmatrix} + a * \begin{pmatrix} L(P_1, P_1) & L(P_1, P_2) \cdot & L(P_1, P_{12}) \\ L(P_2, P_1) & L(P_2, P_2) \cdot & L(P_2, P_{12}) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ L(P_{12}, P_1) & L(P_{12}, P_2) \cdot & L(P_{12}, P_{12}) \end{pmatrix} * \begin{pmatrix} PR(P_1) \\ PR(P_2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ PR(P_{12}) \end{pmatrix}$$

Where  $L$  is the function of the contiguity such that:

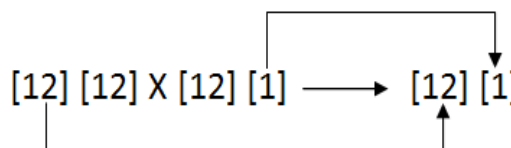
$L(P_i, P_j)$  is 0 if the page does not link  $P_j$   $P_i$ ,  
Normalized such that, for each  $j \sum_{i=1}^{12} L(P_i, P_j) = 1$

In our example the matrix A is defined as:

In this study we selected three different structures, simplified for the sake of simplification of the calculations.

$$A = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \\ S_8 \\ S_9 \\ S_{10} \\ S_{11} \\ S_{12} \end{pmatrix}$$

The matrix of size  $[12] [12]$  and the vector S are high  $[12] [1]$ . The result is therefore size vector  $[12] [1]$ :



The resolution of this matrix system in a direct way is almost impossible given the large number of unknowns and equations. For this reason we have to appeal to frequent iterative methods for solving the system.

Above JavaScript function created based on matrices, which simulates the PageRank algorithm and to establish the automatically.[10][11]

### 1- Function javascript

```

calculated.Matrix.prototype.row_stochastic =
function(damping_factor) {

    var row_length = this.elements[0].length;
    var d = (1 - damping_factor) /
row_length;

    var row_total = [];

    for (var x = 0; x < row_length; x++) {
        row_total.push(0);
        for (var y = 0; y < row_length; y++) {
            row_total[x] +=
this.elements[x][y];
        }
        var a1 = this.elements.clone();

        for (var x = 0; x < row_length; x++) {
            for (var y = 0; y < row_length; y++) {
                if (row_total[x] > 0) {
                    a1[x][y] =
a1[x][y]/row_total[x] + d;
                }
                else {

```

```
        a1[x][y] = (1/row_length) +  
d;  
    }  
}  
  
    return $M(a1);  
  
}
```

## 6. CONCLUSION

The simulator that creates javascript based matrices plays a key role in calculating the importance of web pages and automatically establish their classification: it is clear that it is difficult (if not impossible) to calculate the ranking by hand to a large number of pages,

Optimize the architecture of the links of a site for the PageRank is choosing the pages that PageRank should be the most important. This is usually the home page but in the case of a site with many pages, the headings inputs pages can be involved.

Our recommendation is to spend time creating rich content for your visitors (for a business visitor is a potential customer!), Because it is the added value of your current site.

## 7. REFERENCES

- [1] Thierry Chopin, Jean-François Jamet, Emmanuel Monnet Le e-commerce en France et en Europe la Francepeut-elle devenir un leader Mars 2009.
- [2] D.R.W. Holton, I. Nafea, M. Younas, I. Awan .A class-based scheme for E-commerce web servers: Formal specification and performance evaluation.

- [3] Journal of Network and Computer Applications, Volume 32, Issue 2, March 2009, Pages 455-460.
- [4] Ushio Sumita, Jun Yoshii. Enhancement of e-commerce via mobile accesses to the Internet Electronic Commerce Research and Applications, Volume 9, Issue 3, May–June 2010, Pages 217-227.
- [5] Faisal Nabi. Secure business application logic for e-commerce systems Original Research Article Computers & Security, Volume 24, Issue 3, May 2005, Pages 208-217.
- [6] Manoli Albert, Jordi Cabot, Cristina Gómez, Vicente Pelechano. Generating operation specifications from UML class diagrams: A model transformation approach Data & Knowledge Engineering, Volume 70, Issue 4, April 2011, Pages 365-389.
- [7] Francisco García-Sánchez, Rafael Valencia-García, Rodrigo Martínez-Béjar. An integrated approach for developing e-commerce applications Expert Systems with Applications, Volume 28, Issue 2, February 2005, Pages 223-235.
- [8] Rémi BACHELET, “cours de web analytics, chapitre 4 du module du SEO”
- [9] Ecole central de lile, 17/02/2011, sous licence creative commons.
- [10] <http://fr.wikipedia.org>.
- [11] Web analytics : Mesurer le succès et maximiser les profits de votre site web par Nicolas Malo, Jacques Warren, Marc Lolivier: ISBN: 2212544812 Eyrolles edition: 1ere 1/11/2009
- [12] Le web analytics par Jacques Warren, Marc Lolivier Nicolas Malo: ISBN: 2212544812 Eyrolles edition: 2009