



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



**Sudan University of Science & Technology**

**College of Graduate Studies**

*Faculty of computer science and information technology*

**A proposal for enhancement of Sudanese Public  
Universities websites using Page Ranking algorithm**

A thesis submitted in partial fulfillment of the requirement for degree of  
master in information technology

مقترح لتحسين مواقع الجامعات الحكومية السودانية باستخدام خوارزمية  
ترتيب الصفحات

**Prepared by:**

**Ibrahim Osman Noah Salih**

**Supervisor:**

**Dr. Mahdi Ibrahim**

**2021**

قال رسول الله صلى الله عليه وسلم  
من أتى بي بيوتاً فسلم عليهم  
سلموا له ولوالديه

# Acknowledgments

First and foremost, praises and thanks to the God, the Almighty, for His showers of blessings throughout my research work to complete the research successfully.

I would like to express my deep and sincere gratitude to my research supervisor, Dr. Hisham Mansour for giving me the opportunity to do research and providing invaluable guidance throughout this research.

I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future

Also I express my thanks to my sisters, brother, for their support. I would like to say thanks to my friends and research.

I express my special thanks to Sudan university of science and Technology for their genuine support throughout this research work.

## **Abstract**

Web Data Mining is an important area of Data Mining which deals with the extraction of interesting knowledge from the World Wide Web, Hyperlink analysis was important methodology used by famous search engine Google to rank the pages. The problem of this research is the low rank of Sudanese universities websites on the web especially the Sudanese public universities websites. The aim of this research is to analyze the Sudanese government universities to know the low Rank reason and suggest strategies to enhance the rank. In this research, scientific papers in this field were first reviewed, and then search engine optimization tools, university website links and PageRank algorithm were used. It was found that the reason for the low rank is poor design and insufficient internal and external links, so it suggested strategies to enhance the rank

## المستخلص

يعد تعدد بيانات الويب مجالاً مهماً للتنقيب في البيانات يتعامل مع استخراج المعرفة المثيرة للاهتمام من شبكة الويب العالمية ، تحليل الارتباط التشعبي منهجية مهمة يستخدمها محرك البحث الشهير جوجل لترتيب الصفحات. مشكلة هذا البحث هي تدني ترتيب مواقع الجامعات السودانية على شبكة الإنترنت خاصة مواقع الجامعات الحكومية السودانية. الهدف من هذا البحث هو تحليل مواقع الويب الخاصة بالجامعات الحكومية السودانية لمعرفة سبب التدني واقتراح الاستراتيجيات لتعزيز الرتبة. في هذا البحث أولاً تم الطلاع على الاوراق العلمية فى هذا المجال ومن ثم استخدمت أدوات تحسين محركات البحث وتحديد روابط موقع الويب الخاص بالجامعات وخوارزمية PageRank. تبين ان سبب الرتبة المنخفضة هو التصميم السيئ وعدم كفاية الروابط الداخلية والخارجية ، لذلك اقترحت استراتيجيات لتعزيز الرتبة.

## Abbreviation

KDD	Knowledge discover
SEO	Search Engine Optimization
PR	Page Rank
cPR	Credence <b>PageRank</b>
TLD	Top level domain
SUST	Sudan university of sciences & technology

Table of Contents	
<b>CHAPTER ONE</b> .....	<b>1</b>
<b>INTRODUCTION</b> .....	<b>1</b>
<b>CHAPTER ONE</b> .....	<b>2</b>
<b>INTRODUCTION</b> .....	<b>2</b>
1.1 Background .....	2
1.2 Problem statement .....	3
1.3 Objectives.....	3
1.4 Methodology.....	3
1.5 contribution of the research .....	3
<b>CHAPTER TWO</b> .....	<b>4</b>
<b>LITERATURE REVIEW AND RELATED WORK</b> .....	<b>4</b>
<b>CHAPTER TWO</b> .....	<b>5</b>
<b>LITERATURE REVIEW AND RELATED WORK</b> .....	<b>5</b>
2.1 Section One: Literature Review .....	5
2.1.1 Introduction .....	5
2.1.2 Introduction to web Mining.....	5
2.1.3 Web Mining Categories.....	5
2.1.3.1 Web content mining .....	6
2.1.3.2 Web structure mining .....	6
2.1.3.3 Web usage mining.....	6
2.1.4 Link analysis Algorithms.....	7
2.1.5 Page Rank.....	7
<b>2.1.6 PageRank Idea</b> .....	9
2.2 section two: related work .....	9
<b>2.2.1 Introduction</b> .....	9
2.2.2 Web Page Ranking using Link Attributes .....	9
2.2.3 Web Path Recommendations based on Page Ranking and Markov Models .....	10
2.2.4 The PageRank Citation Ranking: Bringing Order to the Web.....	10
2.2.5 Link Analysis Algorithms for Web Mining .....	10
2.2.6 Page Ranking Based on Number of Visits of Links of Web Page .....	11
2.2.7 Related Work Summary .....	11

<b>Chapter three</b> .....	<b>13</b>
<b>Methodology</b> .....	<b>13</b>
<b>CHAPTER THREE</b> .....	<b>14</b>
<b>METHODOLOGY</b> .....	<b>14</b>
3.1 Introduction .....	14
Figure (3) research methodology.....	14
3.2 How to measure PageRank? .....	15
3.4 PageRank example .....	17
3.5 PageRank analysis .....	18
3.6 SEO Tools.....	18
3.6.2 Link tools .....	18
3.6.3 Google’s Mobile-Friendly Test .....	18
<b>Chapter four</b> .....	<b>19</b>
<b>Implementation</b> .....	<b>19</b>
<b>CHAPTER FOUR</b> .....	<b>20</b>
<b>IMPLEMENTATION</b> .....	<b>20</b>
4.1 link analysis tools .....	20
4.2.1 Google PageRank Checker .....	20
4.2.2 seo spy glass .....	25
4.3.1 Google’s mobile friendly .....	34
➤ Mobile friendly test for Sudan university of science and technology .....	34
➤ Mobile friendly test for Khartoum university .....	37
<b>CHAPTER FIVE</b> .....	<b>40</b>
<b>DISCUSSIONS</b> .....	<b>40</b>
<b>CHAPTER FIVE</b> .....	<b>41</b>
<b>DISCUSSIONS</b> .....	<b>41</b>
5.1 Introduction .....	41
5.2 Improve your website’s user experience	41
5.3 Write great content optimized for SEO .....	41
5.4 Get more backlinks .....	41



5.5 Improve your page speed .....	41
5.6 Fix broken links .....	42
5.7 Optimize your images .....	42
5.8 Use H1 and H2 header tags.....	42
5.9 use Skyscraper SEO Technique .....	42
5.10 Metadata.....	42
5.11 Optimize robots.txt file .....	43
5.12 Add Sitemap files .....	43
5.13 Conclusion.....	43
5.14 Future work.....	43
<b>References</b> .....	<b>44</b>

## **CHAPTER ONE**

### **INTRODUCTION**

# CHAPTER ONE

## INTRODUCTION

### 1.1 Background

Web mining is an active research area in present scenario. Web Mining is defined as the application of data mining techniques on the World Wide Web to find hidden information. This hidden information i. e. knowledge could be contained in content of Web pages or in link structure of WWW or in Web server logs. Based upon the type of knowledge, Web mining is usually divided in three categories: Web content mining, Web structure mining and Web usage mining. An application of Web mining can be seen in the case of search engines. Most of the search engines are ranking their search results in response to users' queries to make their search navigation easier.[1]

The World Wide Web creates many new challenges for information retrieval. It is very large and heterogeneous. Current estimates are that there are over 150 million web pages with a doubling life of less than one year. More importantly, the web pages are extremely diverse, ranging from "What is Joe having for lunch today?" to journals about information retrieval. In addition to these major challenges, search engines on the Web must also contend with inexperienced users and pages engineered to manipulate search engine ranking functions.[2]

Currently, PageRank is not the only algorithm used by Google to order search results, but it is the first algorithm that was used by the company, and it is the best known.[3]

PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set[4].

## **1.2 Problem statement**

The problem of research can be divided into

1. The low rank of Sudanese universities websites on the web especially the Sudanese public universities websites.
2. Bad presence of the Sudanese public universities on the web.

## **1.3 Objectives**

3. Measuring the importance of website pages.
4. Help the Sudanese public universities to enhance their web site.
5. Improve the presence of the Sudanese public universities on the web.

## **1.4 Methodology**

In this research firstly Literature survey about PageRank then select PageRank algorithm then using SEO tools to and Specify the link of the university web site and then apply the algorithm.

## **1.5 contribution of the research**

The research contributes to developing the pages of Sudanese public universities websites through measuring the page rank and the use of SEO tools in analyzing the pages to know the reason of low Rank and tools for publishing back links of web site in order to enhance the Rank of web site and the presence of the website.

## **CHAPTER TWO**

### **LITERATURE REVIEW AND RELATED WORK**

## CHAPTER TWO

### LITERATURE REVIEW AND RELATED WORK

#### 2.1 Section One: Literature Review

##### 2.1.1 Introduction

This Chapter reviews the concept of web mining in general, and link analysis Algorithms and page rank, and some of related works.

##### 2.1.2 Introduction to web Mining

With the huge amount of information available online, the World Wide Web is a fertile area for data mining research. The Web mining research is at the cross road of research from several research communities, such as database, information retrieval, and within AI, especially the sub-areas of machine learning and natural language processing..[4]

Web mining is the Data Mining technique that automatically discovers or extracts the information from web documents. It is the extraction of interesting and potentially useful patterns and implicit information from artifacts or activity related to the Web.[5]

Web Data Mining is the application of data mining techniques to find interesting and potentially useful knowledge from web data. It is normally expected that either the hyperlink structure of the web or the web log data or both have been used in the mining process.[7]

##### 2.1.3 Web Mining Categories

this section we give the overview of each category. Web mining is categorized into three areas of interest based on which part of the Web to mine:[4]

- Web content mining
- Web structure mining
- Web usage mining.

### **2.1.3.1 Web content mining**

It deals with discovering useful information or knowledge from web page contents. Web Content Mining can be thought of the extending the work performed by basic search engines. Web content mining analyzes the content of Web resources. Information Retrieval is one of the research areas that provide a range of popular and effective, mostly statistical methods for Web content mining.[4]

### **2.1.3.2 Web structure mining**

It deals with discovering and modeling the link structure of web. Web information retrieval tools make use of only the text available on web pages but ignoring valuable information contained in web links. Web structure mining aims to generate structural summary about web sites and web pages. The main focus of web structure mining is on link information. Web bags may be used to discover visible web documents, luminous web documents (number of outgoing links) and luminous paths (set of inter-linked nodes) that most of the search engines fails to discover.[4]

### **2.1.3.3 Web usage mining**

It deals with understanding user behavior in interacting with the web or with a website. One of the aims is to obtain information that may assist web site reorganization or assist site adaptation to better suit the user. Web usage mining model is a kind of mining to server logs and its aim is getting useful users' access information in logs to make sites can perfect themselves with appropriate users' requirements, serve users better and get more economy benefits. Several surveys on Web usage mining exist in.[4]

Most data used for mining is collected from Web servers, clients, proxy servers, or server databases, all of them produce noisy data. Because Web mining is sensitive to noise, data cleaning methods are necessary.[4]

#### 2.1.4 Link analysis Algorithms

Web mining technique provides the additional information through hyperlinks where different documents are connected. We can view the web as a directed labeled graph whose nodes are the documents or pages and edges are the hyperlinks between them[8].

#### 2.1.5 Page Rank

The PageRank algorithm is the most popular link analysis algorithm, used broadly for assigning numerical weightings to web documents and used from web search engines in order to rank the retrieved results. The algorithm models the behavior of a random surfer, who either chooses an outgoing link from the page he's currently at, or "jumps" to a random page after a few clicks.[9]

The PageRank of a page is defined as the probability of the random surfer being at some particular time step  $k > K$  at this page. This probability is correlated with the importance of this page, as it is defined based on the number and the importance of the pages pointing to it. For sufficiently large  $K$  this probability is unique, as illustrated in what follows.[3]

Consider the web as a directed graph  $G$ , where the  $N$  nodes represent the web pages and the edges the links between them. The random walk on  $G$  induces a Markov Chain where the states are given by the nodes in  $G$ , and  $M$  is the stochastic transition matrix with  $m_{ij}$  describing the one-step transition from page  $j$  to page  $i$ . The adjacency function  $m_{ij}$  is 0 if there is no direct link from  $p_j$  to  $p_i$ , and normalized such that, for each  $j$ :

$$\sum_{i=1}^N m_{ij} = 1$$

(3).

PageRank is in essence the stationary probability distribution over pages induced by a random walk on the web. The convergence of PageRank is guaranteed only if  $M$  is irreducible and periodic. The periodicity of  $M$  is guaranteed in practice in the web context, whereas the irreducibility is satisfied by adding a dumping factor  $(1-\epsilon)$  to the rank propagation ( $\epsilon$  is a very small number, usually set to 0.15), in order to limit the



effect of rank sinks and guarantee convergence to a unique vector. PageRank can then be expressed as the unique solution to Equation:

$$PR = \epsilon \times M \times PR + (1 - \epsilon) p$$

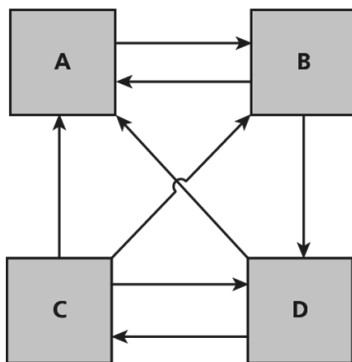
where  $p$  is a non-negative  $N$ -vector whose elements sum to 1. Usually  $m_{ij} = \frac{1}{\sum_{p_k \in Out(p_j)} |p_k|}$  and,  $p = \left[ \frac{1}{N} \right]_{N \times 1}$  i.e. the probability of

Randomly jumping to another page is uniform. By choosing, however,  $p$  to follow a non-uniform distribution, we essentially bias the resultant PageRank vector computation to favor certain pages.

Page Rank is a numeric value that represents how important a page is on the web. Page Rank is the Google's method of measuring a page's "importance." When all other factors such as Title tag and keywords are taken into account, Google uses Page Rank to adjust results so that more "important" pages move up in the results page of a user's search result display. Google Figs. Google calculates a page's importance from the votes cast for it. How important each vote is taken into account when a page's Page Rank is calculated. It matters because it is one of the factors that determine a page's ranking in the search results The order of ranking in Google works like this: Find all pages matching the keywords of the search. Adjust the results by Page Rank scores. The algorithm of Page Rank as follows:[3]

The original Page Rank algorithm is given in following equation

$$PR(P) = (1-d) + d(PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$



Where, PR (P)= PageRank of page P

PR (Ti) = PageRank of page Ti which link to page

C (Ti) =Number of outbound links on page TD = Damping factor which can be set between 0 and 1.

Figure 1.: Example of Backlinks

### 2.1.6 PageRank Idea

The idea behind PageRank is that good pages reference good pages. Hence, pages that are referenced by good pages have higher PageRank. Although there are several formulations of PageRank, we use the random surf metaphor. Suppose that you are a user surfing the Web in a random fashion, such that, if you are in a page, with certain probability at random to follow one of the links on the page where you are (removing self-links). Hence, the probability of being in page  $p$  is

$$PR(p) = \frac{q}{T} + (1 - q) \sum_i \frac{PR(r_i)}{L(r_i)}$$

Where  $T$  is the total number of pages,  $q$  is the probability of leaving page  $p$  (in the original work  $q = 0.15$  is suggested),  $r_i$  are the pages that point to page  $p$ , and  $L(r_i)$  is the number of links in page  $r_i$ . These values can then be used as page ranking, and can be computed by an iterative algorithm converging quite fast, as we are interested in the ranking order rather than the actual ranking values. The term  $q$  is called damping factor as decreases exponentially link spamming based in sequences of links that return to a page.

## 2.2 section two: related work

### 2.2.1 Introduction

This chapter discusses the literature concerned with Data Mining and Knowledge Discovery techniques used in PageRank (structure mining).

### 2.2.2 Web Page Ranking using Link Attributes

We present a variant of PageRank, WL-Rank, that considers different Web page attributes to give more weight to some links, improving the precision of the answers.[10]

### **2.2.3 Web Path Recommendations based on Page Ranking and Markov Models**

Markov models have been widely used for modeling users' navigational behavior in the Web graph, using the transitional probabilities between web pages, as recorded in the web logs. The recorded users' navigation is used to extract popular web paths and predict current users' next steps. Such purely usage-based probabilistic models, however, present certain shortcomings.

Since the prediction of users' navigational behavior is based solely on the usage data, structural properties of the Web graph are ignored. Thus important – in terms of PageRank authority score- paths may be underrated. In this paper we present a hybrid probabilistic predictive model extending the properties of Markov models by incorporating link analysis methods. More specifically, we propose the use of a PageRank-style algorithm for assigning prior probabilities to the web pages based on their importance in the web site's graph.[9]

### **2.2.4 The PageRank Citation Ranking: Bringing Order to the Web**

The importance of a Web page is an inherently subjective matter, which depends on the Readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them. We compare PageRank to an idealized random Web surfer. We show how to efficiently compute PageRank for large numbers of pages. And, we show how to apply PageRank to search and to user navigation[2]

### **2.2.5 Link Analysis Algorithms for Web Mining**

Web mining is the Data Mining technique that automatically discovers or extracts the information from web documents. Page Rank and Weighted Page Rank algorithms are used in Web Structure Mining to rank the relevant pages. In this paper we focused that by using Page Rank and Weighted Page Rank algorithms users may not get the required relevant documents easily, but in new algorithm Weighted Page Content Rank user can get relevant and important pages easily as it employs web structure mining and web content mining. The input parameters used in Page Ranker Backlinks,

Weighted Page Rank uses Backlinks and Forward Links as Input Parameter and Weighted Page Content Rank use Backlinks, Forward Link and Content as Input Parameters.[8]

### 2.2.6 Page Ranking Based on Number of Visits of Links of Web Page

Search engines generally return a large number of pages in response to user queries. To assist the users to navigate in the result list, ranking methods are applied on the search results. Most of the ranking algorithms proposed in the literature are either link or content oriented, which do not consider user usage trends. In this paper, a page ranking mechanism called Page Ranking based on Visits of Links(VOL) is being devised for search engines, which works on the basic ranking algorithm of Google i.e. PageRank and takes number of visits of inbound links of Web pages into account. This concept is very useful to display most valuable pages on the top of the result list on the basis of user browsing behavior, which reduces the search space to large scale. The paper also presents a method to find link-visit counts of Web pages and a comparison between VOL with the PageRank algorithm.[11]

### 2.2.7 Related Work Summary

#	Paper & author	Purpose of study	Algorithms used	Data	Results
1	Application of Webometrics Techniques for Measuring and Evaluating Visibility of University Library Websites in Sri Lanka  Application of Webometrics Techniques for Measuring and Evaluating Visibility of University Library Websites in Sri Lanka	Measuring and Evaluating Visibility of University Library	pageRank	Fifteen university library websites were selected for this study	Presence of University Libraries on the Web: Real Visibility and Impact The number of page count, inlink/external backlink count, rich files, scholar count

#	Paper & author	Purpose of study	Algorithms used	Data	Results
2	<p>Web Structure Mining: Exploring Hyperlinks and Algorithms for Information Retrieval</p> <p>Ravi Kumar P Department of ECEC, Curtin University of Technology, Sarawak Campus, Miri, Malaysia ravi2266@gmail.com</p>	<p>Hyperlink analysis Compare link analysis algorithm</p>	<p>PageRank, HITS (Hyperlink-Induced Topic Search) and other algorithms</p>	<p>Random links</p>	<p>Page rank algorithm is better</p>
3	<p>Link Analysis Algorithms For Web Mining</p> <p>Tamanna Bhatia Dept. of Computer Science, Desh Bhagat Engineering College, Mandi Gobindgarh, Punjab, India</p>	<p>give description about Weighted Page Content Rank (WPCR) based on web content mining and structure mining</p>	<p>Page Rank, Weighted Page Rank and Weighted Page Content Rank.</p>	<p>Back link</p>	<p>Page rank algorithm is better and its used in google search engine</p>
4	<p>Web Path Recommendations based on Page Ranking and Markov Models</p> <p>Magdalini Eirinaki, Michalis Vazirgiannis, Dimitris Kapogiannis Athens University of Economics and Business</p>	<p>present a hybrid probabilistic predictive model extending the properties of Markov models by incorporating link analysis methods</p>	<p>pageRank</p>	<p>Back link</p>	<p>results in more objective and representative predictions than the ones produced from the pure usage-based approaches</p>
5	<p>Web Page Ranking using Link Attributes</p> <p>Ricardo BaezaYates Emilio Davis Center for Web Research, CS Dept.</p>	<p>improving the precision of the answers.</p>	<p>PageRank, WLRank,</p>	<p>link</p>	<p>improving the precision of the answers.</p>

## **Chapter three**

### **Methodology**

## CHAPTER THREE

### METHODOLOGY

#### 3.1 Introduction

In this research firstly Literature survey about PageRank then select PageRank algorithm then using SEO tools to and Specify the link (select Sudan university as case study and Khartoum University and Neelain and Omdurman Islamic University for comparing) of the university web site and then apply the algorithm.

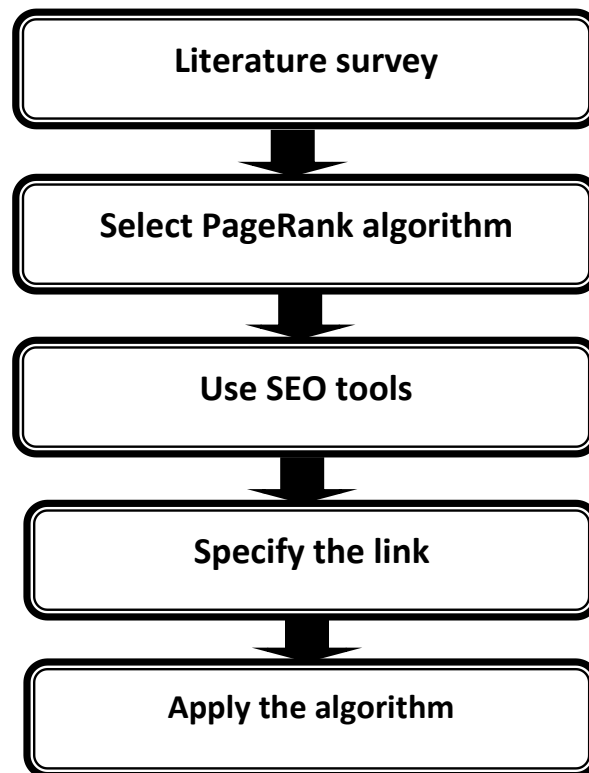


Figure (3) research methodology

### 3.2 How to measure PageRank?

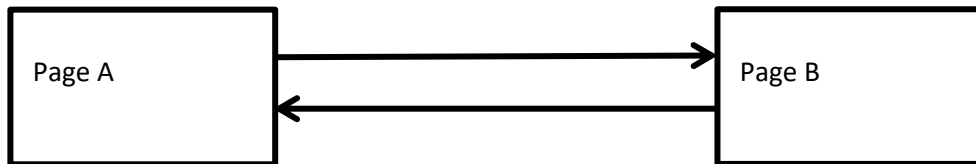
The PR of each page depends on the PR of the pages pointing to it. But we won't know what PR those pages have until the pages pointing to them have their PR calculated and so on... And when you consider that page links can form circles it seems impossible to do this calculation!

But actually it's not that bad. Remember this bit of the Google paper:

PageRank or  $PR(A)$  can be calculated using a simple iterative algorithm, and corresponds to the principal eigenvector of the normalized link matrix of the web.[12]

What that means to us is that we can just go ahead and calculate a page's PR without knowing the final value of the PR of the other pages. That seems strange but, basically, each time we run the calculation we're getting a closer estimate of the final value. So all we need to do is remember the each value we calculate and repeat the calculations lots of times until the numbers stop changing much.[12]

Let's take the simplest example network: two pages, each pointing to the other:



Figure(2) simple page Rank

Each page has one outgoing link (the outgoing count is 1, i.e.  $C(A) = 1$  and  $C(B) = 1$ ).

Guess 1

We don't know what their PR should be to begin with, so let's take a guess at 1.0 and do some calculations:

$$d = 0.85$$

$$PR(A) = 0.15 + d(PR(B)/1)$$

$$PR(B) = 0.15 + d(PR(A)/1)$$

i.e.

$$PR(A) = 0.15 + 0.85 * 1 = 1$$

$$PR(B) = 0.15 + 0.85 * 1 = 1$$



Guess 2

let's start the guess at 0 instead and re-calculate:

$$\begin{aligned} \text{PR(A)} &= 0.15 + 0.85 * 0 \\ &= 0.15 \end{aligned}$$

$$\begin{aligned} \text{PR(B)} &= 0.15 + 0.85 * 0.15 \text{ NB. we've already calculated a} \\ &= 0.2775 \text{ "next best guess" at PR(A) so we} \\ &\text{use it here} \end{aligned}$$

And again:

$$\begin{aligned} \text{PR(A)} &= 0.15 + 0.85 * 0.2775 \\ &= 0.385875 \end{aligned}$$

$$\begin{aligned} \text{PR(B)} &= 0.15 + 0.85 * 0.385875 \\ &= 0.47799375 \end{aligned}$$

And again

$$\begin{aligned} \text{PR(A)} &= 0.15 + 0.85 * 0.47799375 \\ &= 0.5562946875 \end{aligned}$$

$$\begin{aligned} \text{PR(B)} &= 0.15 + 0.85 * 0.5562946875 \\ &= 0.622850484375 \end{aligned}$$

and so on.

Guess 3

Well let's see. Let's start the guess at 40 each and do a few cycles:

$$\begin{aligned} \text{PR(A)} &= 40 \\ \text{PR(B)} &= 40 \end{aligned}$$

First calculation

$$\begin{aligned} \text{PR(A)} &= 0.15 + 0.85 * 40 \\ &= 34.25 \end{aligned}$$

$$\begin{aligned} \text{PR(B)} &= 0.15 + 0.85 * 34.25 \\ &= 29.1775 \end{aligned}$$

And again

$$\begin{aligned} \text{PR(A)} &= 0.15 + 0.85 * 29.1775 \end{aligned}$$

$$\begin{aligned}
 &= 24.950875 \\
 \text{PR(B)} &= 0.15 + 0.85 * 24.950875 \\
 &= 21.35824375
 \end{aligned}$$

those numbers are heading down alright! It sure looks the numbers will get to 1.0 and stop.

### 3.4 PageRank example

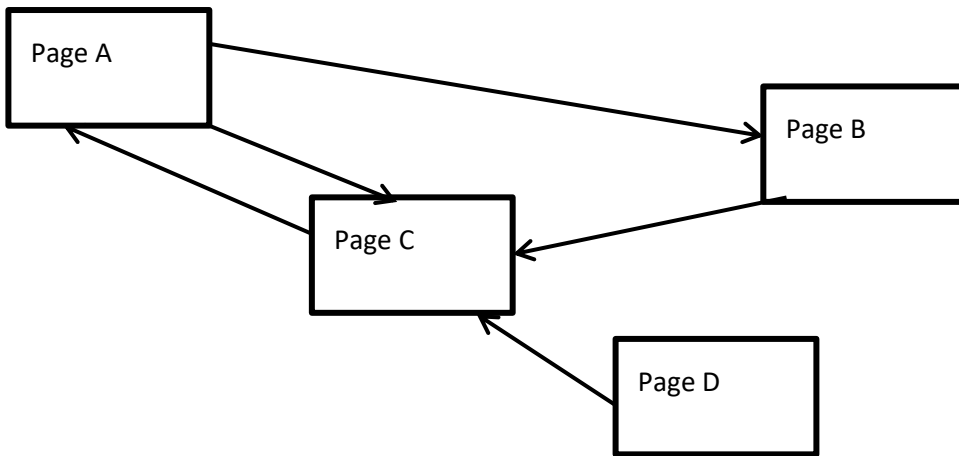


Figure ( 3 ) complicated PageRank

So the correct PR for the example is:

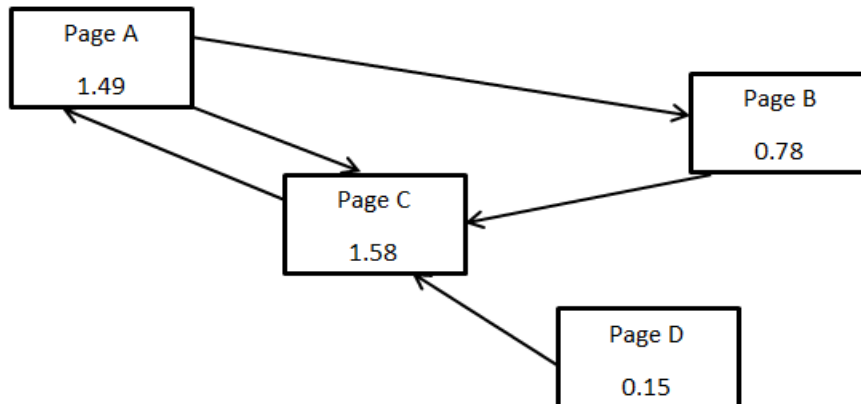


Figure (4) complicated PageRank

Average PR: 1.000

Look at Page D though - it has a PR of 0.15 even though no-one is voting for it (i.e. it has no incoming links)! Is this right?

The first part, or "term" to be technical, of the PR equation is doing this:

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

So, for Page D, no backlinks means the equation looks like this:

$$\begin{aligned} PR(A) &= (1-d) + d * (0) \\ &= 0.15 \end{aligned}$$

no matter what else is going on or how many times you do it.

**Observation:** every page has at least a PR of 0.15 to share out. But this may only be in theory - there are rumours that Google undergoes a post-spidering phase whereby any pages that have no incoming links at all are completely deleted from the index...

### 3.5 PageRank analysis

Page rank analysis is powerful suite of search engine optimization tools.

### 3.6 SEO Tools

#### 3.6.2 Link tools

- Page Rank checker
- SEO Spy Glass
- SEO Web Site Auditor

#### 3.6.3 Google's Mobile-Friendly Test

Get your site ready for Google's Mobile-first Indexing. You might have heard that Google recently made a big change to their algorithm. (This change is officially called "Mobile-First Indexing").

## **Chapter four**

### **Implementation**

## CHAPTER FOUR

### IMPLEMENTATION

#### 4.1 link analysis tools

Link analysis is a process of finding connections between different entities, it also provides information about how other variables or attributes can be used to characterize the type of link as well as its strength.

#### 4.2 There are several link analysis tools:

##### 4.2.1 Google PageRank Checker

Free online tool to check rank of specific page and provide information like :

- Google page rank
- domain authority
- global rank
- links
- and more

➤ apply on Sudan University web site

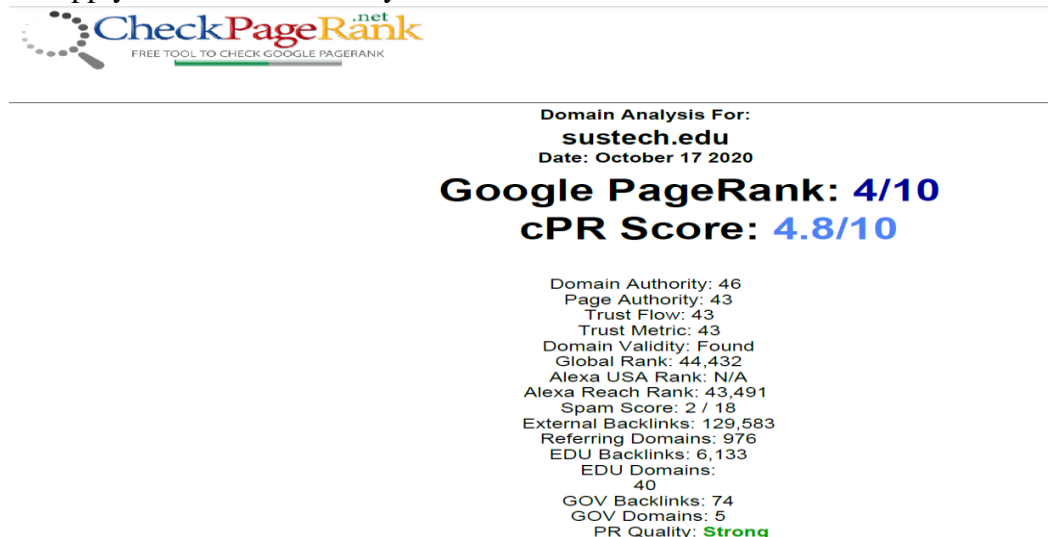


Figure (5) check PageRank from Sudan University

➤ apply on Omdurman Islamic University web site



Domain Analysis For:

**oiu.edu.sd**

Date: October 17 2020

**Google PageRank: 3/10**

**cPR Score: 3.5/10**

Domain Authority: 32  
Page Authority: 32  
Trust Flow: 15  
Trust Metric: 15  
Domain Validity: Found  
Global Rank: 97,712  
Alexa USA Rank: N/A  
Alexa Reach Rank: 133,173  
Spam Score: 0 / 18  
External Backlinks: 5,681  
Referring Domains: 174  
EDU Backlinks: 4,739  
EDU Domains:  
26  
GOV Backlinks: 59  
GOV Domains: 4  
PR Quality: **Moderate**

Figure (6) check PageRank from  
Omdurman Islamic university

➤ apply on neelain University web site



Domain Analysis For:

**neelain.edu.sd**

Date: October 17 2020

**Google PageRank: 3/10**

**cPR Score: 3.6/10**

Domain Authority: 28

Page Authority: 39

Trust Flow: 16

Trust Metric: 16

Domain Validity: Found

Global Rank: 214,338

Alexa USA Rank:

N/A

Alexa Reach Rank: 226,465

Spam Score: 1 / 18

External Backlinks: 10,010

Referring Domains: 422

EDU Backlinks: 5,288

EDU

Domains: 35

GOV Backlinks: 21

GOV Domains: 4

PR Quality: **Moderate**

Figure (7) check PageRank from Neelain  
University

➤ apply on Khartoum University web site



Domain Analysis For:

**uofk.edu**

Date: October 17 2020

**Google PageRank: 5/10**

**cPR Score: 5.2/10**

Domain Authority: 51

Page Authority: 45

Trust Flow: 47

Trust Metric: 47

Domain Validity: Found

Global Rank: 46,620

Alexa USA Rank: N/A

Alexa Reach Rank: 58,139

Spam Score: 1 / 18

External Backlinks: 154,089

Referring Domains: 1,783

EDU Backlinks: 126,131

EDU

Domains: 108

GOV Backlinks: 217

GOV Domains: 12

PR Quality: **Very Strong**

Figure (8) check PageRank from khartoum University





## Domain Comparison

Domain Comparison				
Parameter	sustech.edu	uofk.edu	oiu.edu.sd	neelain.edu.sd
Domain Strength	4.71	5.39	2.40	3.04
Domain InLink Rank	45	59	33	38
Domain Alexa rank	35,117	54,565	330,709	181,477
Domain Age	19y 10m	20y 1m	N/A	N/A
Total Backlinks	37,398	258,977	1,725	6,724
Total Linking Domains	984	2,299	154	359
IP Addresses	641	1,450	126	176
C-Blocks	574	1,272	122	166
Dofollow backlinks	34,785	254,829	1,686	6,164
Domains linking from homepage	39	103	11	14
Top TLD	34.2% com	40.0% com	35.7% com	20.9% com
Top Country	33.8% USA	38.8% USA	46.1% USA	32.3% USA
Text links	99.3%	94.5%	72.5%	14.4%
Anchors	1,333	3,639	133	159
Anchor text diversity	3.6%	1.4%	7.7%	2.4%
Linked Pages	1,484	2,964	47	117

Figure (9) PageRank comparison

when we compare four Sudanese public universities for example we observe Khartoum university has high page rank. in order to the high number of backlink and linking domain and domain linking from home page and total linking domain.

#### 4.2.2 seo spy glass

seo tools give you all the information you need to become the seo consultant for your website. And we use it to:

- Find your competitors' inbound links.
- Locate the Google PageRank for each link.
- Count the number of links by PageRank 0 to 10.
- Calculate the value of each link.
- See if the link is from a homepage.
- Determine if the link is from a blog or a forum.
- Discover which links bring the most traffic to your competitors.
- Locate the age of each linking website.
- Check your competition in more than 400 search engines

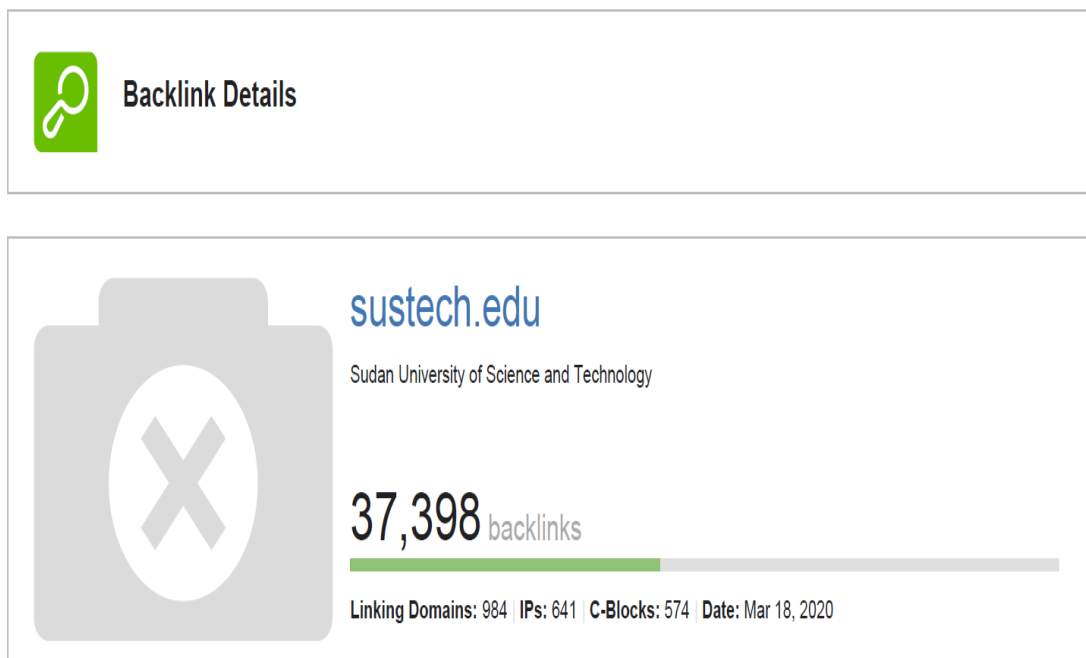


Figure (10) Sudan university back link details

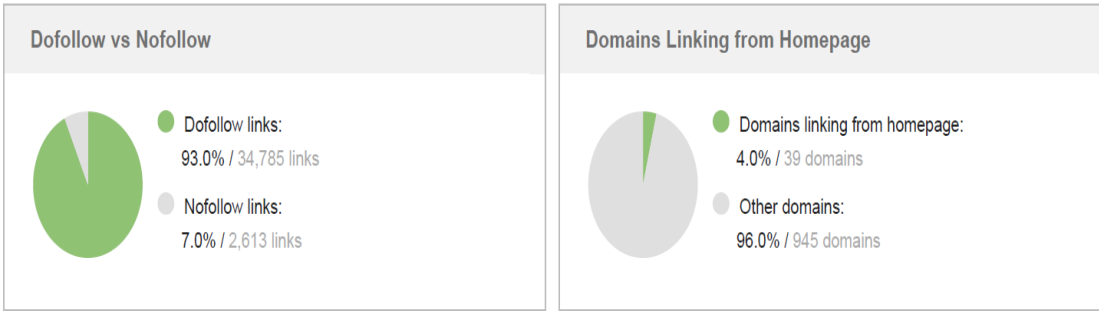


Figure (11) Sudan university back link details

“No flow” tag is there to ask search engines not to “flow” the link – so there is no supposed search engine optimization benefit. And “do flow” is the opposite

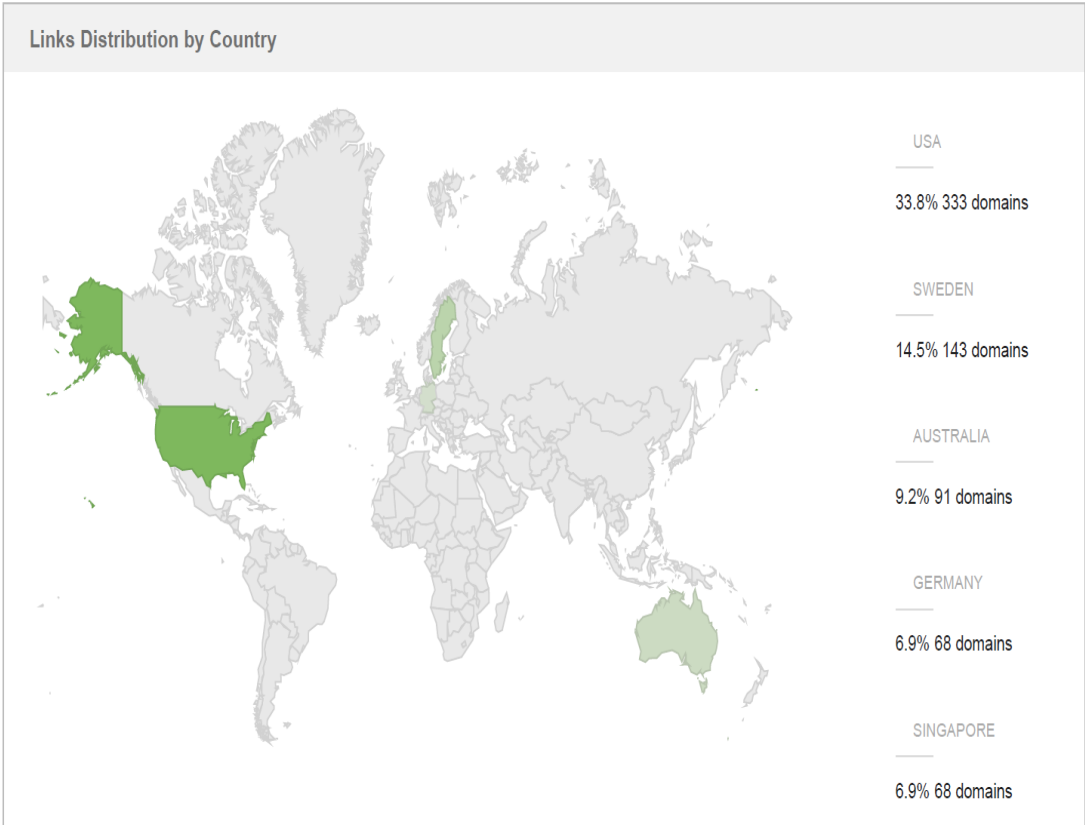


Figure (12) Sudan university back link details

Most of the countries where the backlink is published.

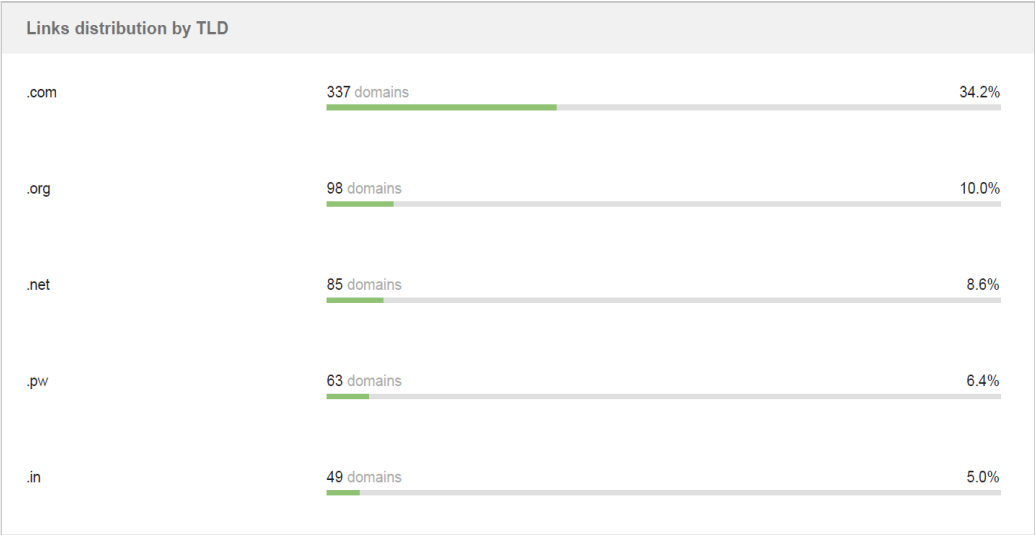


Figure (13) Sudan university back link details

TLD is stand for top-level-domain and it's the last segment of the domain name. so its explain the number of back link in each LTD

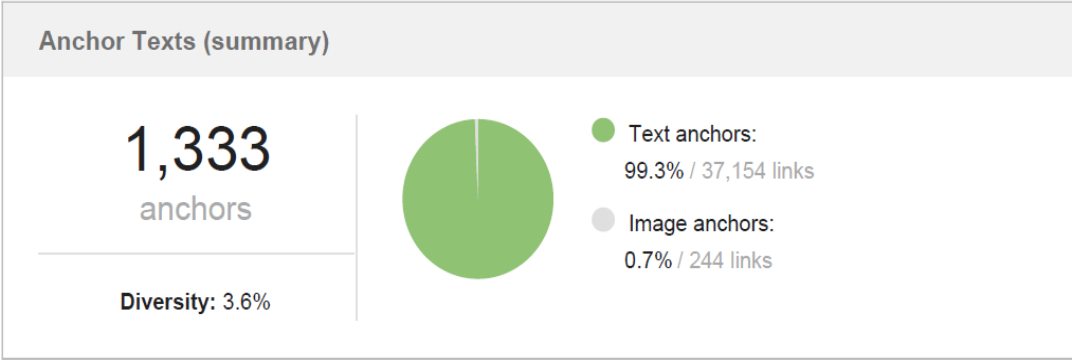


Figure (14) Sudan university back link details

The figure explain the number of anchor text and its type . is it image or text , and we notice that the majority is text.

Top Anchors & Image ALT Texts			
جامعة السودان للعلوم و التكنولوجيا	35.9% / 13,412 pages	sustech.edu	15.1% / 167 domains
sudan university of science and technology	32.0% / 11,960 pages	267324. sustech.edu	13.7% / 152 domains
sudan university for science & technology	7.5% / 2,800 pages	sudan university of science and technology	8.9% / 99 domains
جامعة السودان	4.7% / 1,750 pages	http://www.sustech.edu	5.0% / 56 domains
sudan university of science & technology	2.1% / 787 pages	جامعة السودان للعلوم و التكنولوجيا	3.6% / 40 domains
جامعة السودان للعلوم و التكنولوجيا	1.8% / 661 pages	sudan university of science & technology	3.2% / 36 domains
sustech.edu	1.4% / 531 pages	http://www.sustech.edu/	3.0% / 33 domains
university of sudan for science and technology	0.8% / 314 pages	www.sustech.edu	2.4% / 27 domains
http://repository.sustech.edu/bitstr...=1&isallowed=y	0.8% / 313 pages	جامعة السودان للعلوم و التكنولوجيا	1.9% / 21 domains
267324. sustech.edu	0.5% / 205 pages	جامعة السودان	1.4% / 16 domains
présentation	0.4% / 154 pages	http://mit.sustech.edu/ocwweb/jsp/emailp...i no.us.org	1.2% / 13 domains
http://www.sustech.edu/	0.4% / 153 pages	<no text>	1.2% / 13 domains
www.sustech.edu	0.4% / 153 pages	sudan university for science and technology	1.1% / 12 domains
redirect to sustech.edu	0.4% / 148 pages	http://mit.sustech.edu/ocwweb/jsp/emailpag e.jsp?title=free%20online%20mit%20course%20materials%20%7c%20literature%20%7c%20mit%20opencourseware&referer=htt ps://onlinecasino.us.org	1.1% / 12 domains
http://www.sustech.edu	0.4% / 147 pages	*the structural performance of tower cranes	1.0% / 11 domains

Figure (15) Sudan university back link details

ALT text also known as “alt attributes” . its HTML code to describe the appearance and function of an image on a page. This figure illustrates anchor and image ALT text in backlink.

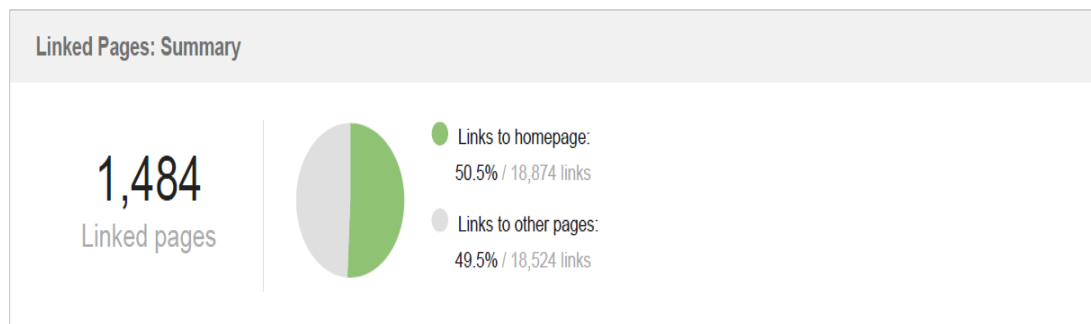


Figure (16) Sudan university back link details

In this figure we observe that is 50.5% links to home page and 49.5 links to other page

Top linked Pages			
/[homepage]	48.1%	18,003 pages	
/sudannewar/	36.9%	13,812 pages	
/[homepage]	2.2%	841 pages	
/sudannewar/index.php	0.8%	302 pages	
/ar	0.8%	299 pages	
/sudannewar	0.6%	240 pages	
/index.php/College_of_Forestry_and_Range_Science/home	0.4%	133 pages	
/OcwWeb/jsp/emailpage.jsp?title=Free Online MIT Course Materials   Literature   MIT OpenCourseWare&referer=https://onlinecasino.us.org	0.3%	130 pages	
/index.php/College_of_Agricultural_studies/home	0.3%	108 pages	
/OcwWeb/jsp/emailpage.jsp?title=Free Online MIT Course Materials   Literature   MIT OpenCourseWare&referer=https://onlinecasino888.us.org	0.3%	105 pages	
/sudannewar/about_sust.php	0.2%	59 pages	
/OcwWeb/jsp/emailpage.jsp?title=Free Online MIT Course Materials   Literature   MIT OpenCourseWare&referer=https://onlinecasinoxplay.us.org	0.1%	52 pages	
/OcwWeb/jsp/emailpage.jsp?title=Free Online MIT Course Materials   Literature   MIT OpenCourseWare&referer=https://onlinecasino24.us.org	0.1%	49 pages	
/OcwWeb/jsp/emailpage.jsp?title=Free Online MIT Course Materials   Literature   MIT OpenCourseWare&referer=https://playonlinecasino.us.org	0.1%	42 pages	
/handle/123456789/1149/browse?value=Supervisor,- +Hamid+Soliman+AbdAlla&type=author	0.1%	1 domains	
المستط/22907/123456789/handle/bitstream/pdf?sequence=2&isAllowed=y...	0.1%	1 domains	
/amirmohamed/en/contact	0.1%	1 domains	
/staff_publications/20100606055314807.pdf	0.1%	1 domains	
التر/18473/123456789/handle/bitstream/الخدمات الإلكترونية/pdf?sequence=1&isAllowed=y....	0.1%	1 domains	
/browse?type=author&value=karama,+Aisha+Musa+Mohamed	0.1%	1 domains	
/index.php/JECS/article/view/188	0.1%	1 domains	
p.البحث/22723/123456789/handle/bitstream/df?sequence=3&isAllowed=y	0.1%	1 domains	
/handle/123456789/1198/browse?value=Supervisor,-Yahia+Ahmed+Raja&type=author	0.1%	1 domains	
/handle/123456789/13246/browse?value=Zakaria+Ali+Ahmed+Osman&type=author	0.1%	1 domains	
/handle/123456789/1292/browse?value=Aradeb,+Mohamed+Ilyas+Mohamed&type=author	0.1%	1 domains	
/handle/123456789/18624	0.1%	1 domains	
/handle/123456789/2907	0.1%	1 domains	
/handle/123456789/2909	0.1%	1 domains	
/handle/123456789/9770	0.1%	1 domains	

Figure (117) Sudan university back link details

this figure is illustrates top of linked pages in Sudan university of science and technology web site.

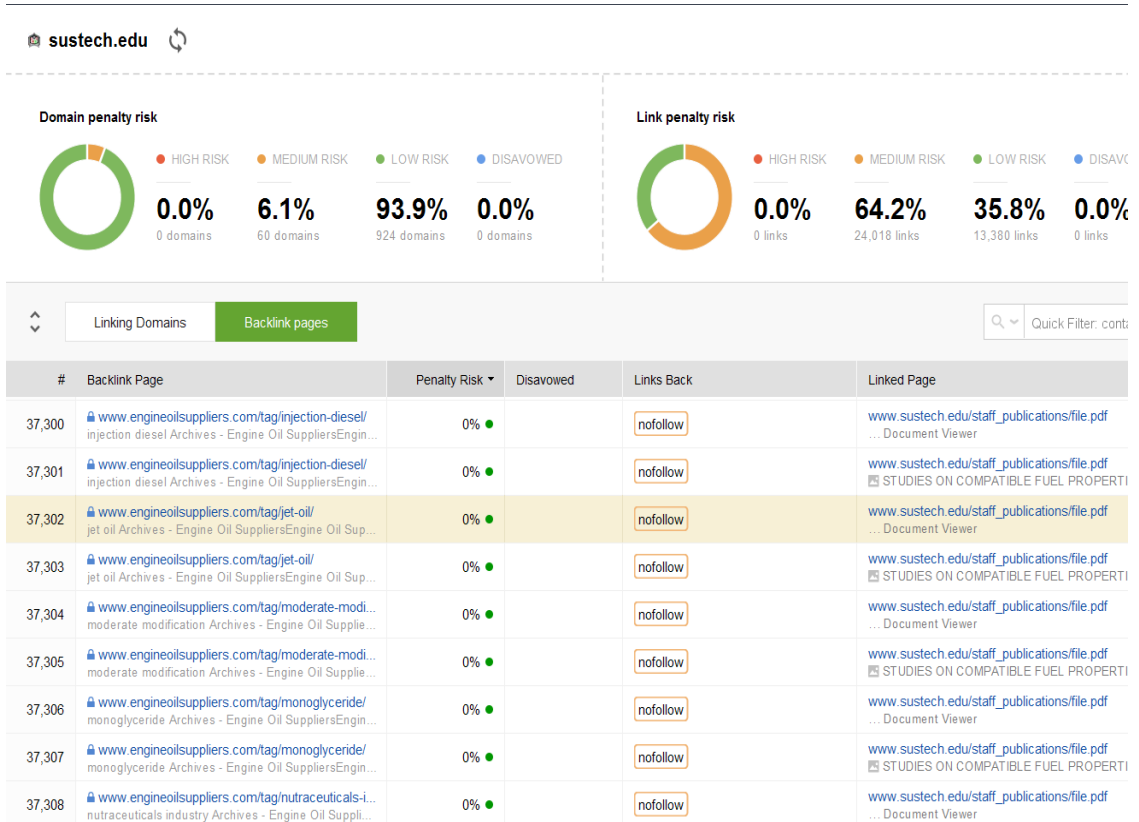


Figure (17) Sudan university back link details

Risk penalty metric let you detect harmful links in your backlink profile and evaluate exactly how dangerous each of your backlink can be for your websites SEO.

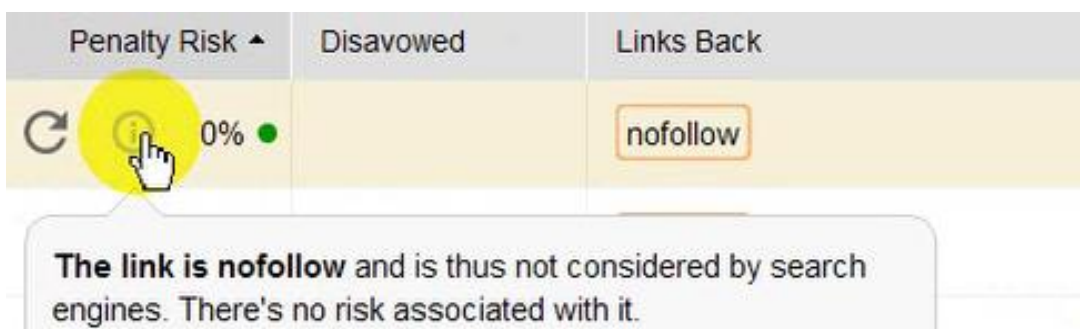


Figure (18) Sudan university back link details

Example of penalty risk Applying in Sudan university of science and technology and the risk is no flow. For the above link the value by default will be 0%, as nofollow links are not considered by search engines and thus may cause no harm.

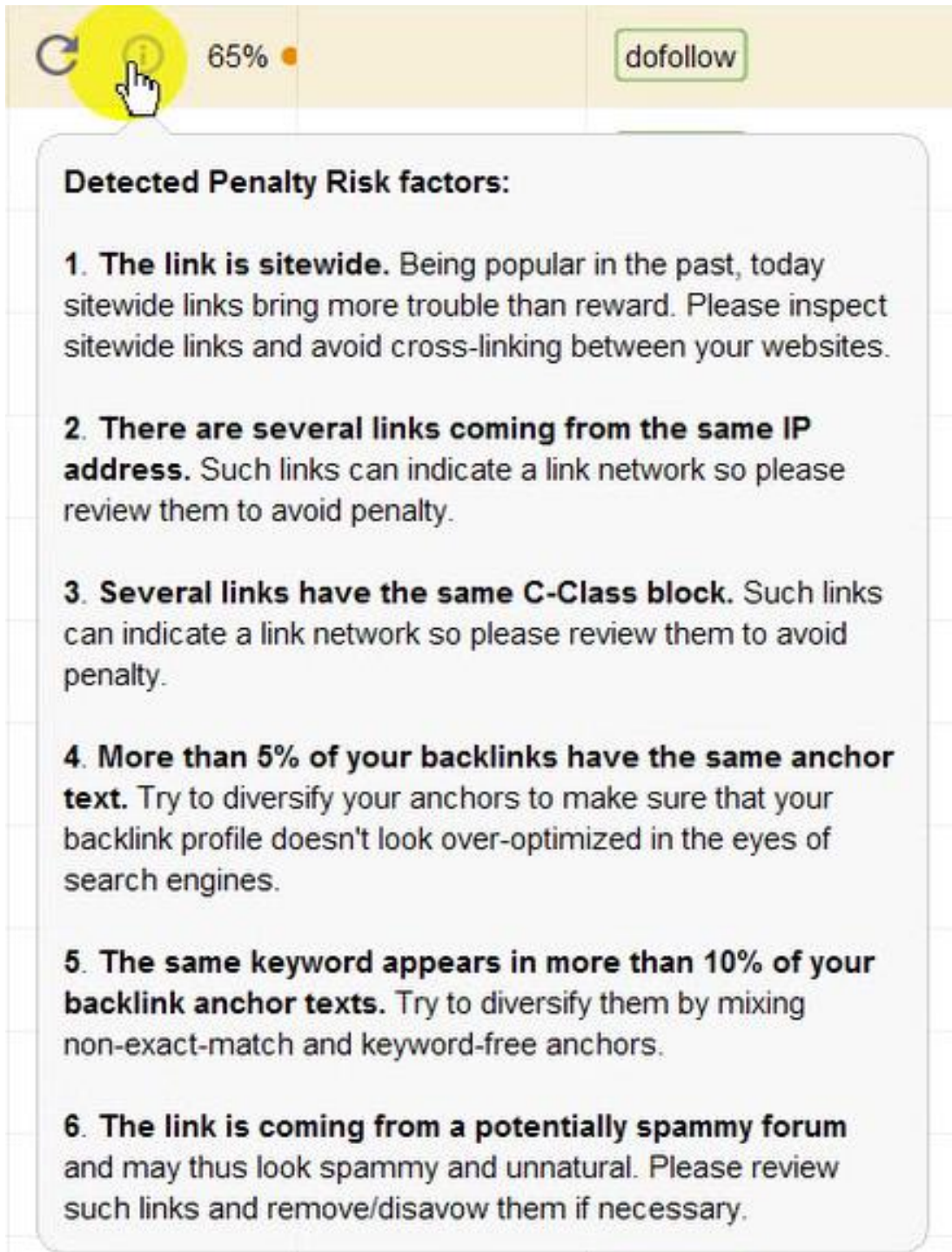


Figure (19) Sudan university back link details

Medium risk in sudan university website and the figure explain the reason of the penalty risk.



Example of penalty risk on Sudan university domain and its percent :

SUST domain	Risk
<a href="http://www2.iium.edu.my/kahts/research-publications/publications-0">http://www2.iium.edu.my/kahts/research-publications/publications-0</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kahts/research-publications/publications-2">http://www2.iium.edu.my/kahts/research-publications/publications-2</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kahts/research-publications/publications">http://www2.iium.edu.my/kahts/research-publications/publications</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kahts/resource-centre/events">http://www2.iium.edu.my/kahts/resource-centre/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kahts/resource-centre/news">http://www2.iium.edu.my/kahts/resource-centre/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kahts/resource-centre/venue-booking-system">http://www2.iium.edu.my/kahts/resource-centre/venue-booking-system</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/11057/events">http://www2.iium.edu.my/kcdi/11057/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/11057/news">http://www2.iium.edu.my/kcdi/11057/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/113/events">http://www2.iium.edu.my/kcdi/113/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/113/news">http://www2.iium.edu.my/kcdi/113/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/128/events">http://www2.iium.edu.my/kcdi/128/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/128/news">http://www2.iium.edu.my/kcdi/128/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/15108/events">http://www2.iium.edu.my/kcdi/15108/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/15108/news">http://www2.iium.edu.my/kcdi/15108/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/15391/events">http://www2.iium.edu.my/kcdi/15391/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/15391/news">http://www2.iium.edu.my/kcdi/15391/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/15701/events">http://www2.iium.edu.my/kcdi/15701/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/15701/news">http://www2.iium.edu.my/kcdi/15701/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/165/events">http://www2.iium.edu.my/kcdi/165/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/165/news">http://www2.iium.edu.my/kcdi/165/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/183/events">http://www2.iium.edu.my/kcdi/183/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/183/news">http://www2.iium.edu.my/kcdi/183/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/1915/events">http://www2.iium.edu.my/kcdi/1915/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/1915/news">http://www2.iium.edu.my/kcdi/1915/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/1920/events">http://www2.iium.edu.my/kcdi/1920/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/1920/news">http://www2.iium.edu.my/kcdi/1920/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/19690/events">http://www2.iium.edu.my/kcdi/19690/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/19690/news">http://www2.iium.edu.my/kcdi/19690/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/20138/events">http://www2.iium.edu.my/kcdi/20138/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/20138/news">http://www2.iium.edu.my/kcdi/20138/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/20276/events">http://www2.iium.edu.my/kcdi/20276/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/20276/news">http://www2.iium.edu.my/kcdi/20276/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/2125/events">http://www2.iium.edu.my/kcdi/2125/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/2125/news">http://www2.iium.edu.my/kcdi/2125/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/2140/events">http://www2.iium.edu.my/kcdi/2140/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/2140/news">http://www2.iium.edu.my/kcdi/2140/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/21584/events">http://www2.iium.edu.my/kcdi/21584/events</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/21584/news">http://www2.iium.edu.my/kcdi/21584/news</a> International Islamic University Malaysia	48%
<a href="http://www2.iium.edu.my/kcdi/227/events">http://www2.iium.edu.my/kcdi/227/events</a> International Islamic University Malaysia	48%

Figure (20) Sudan university back link details - risk

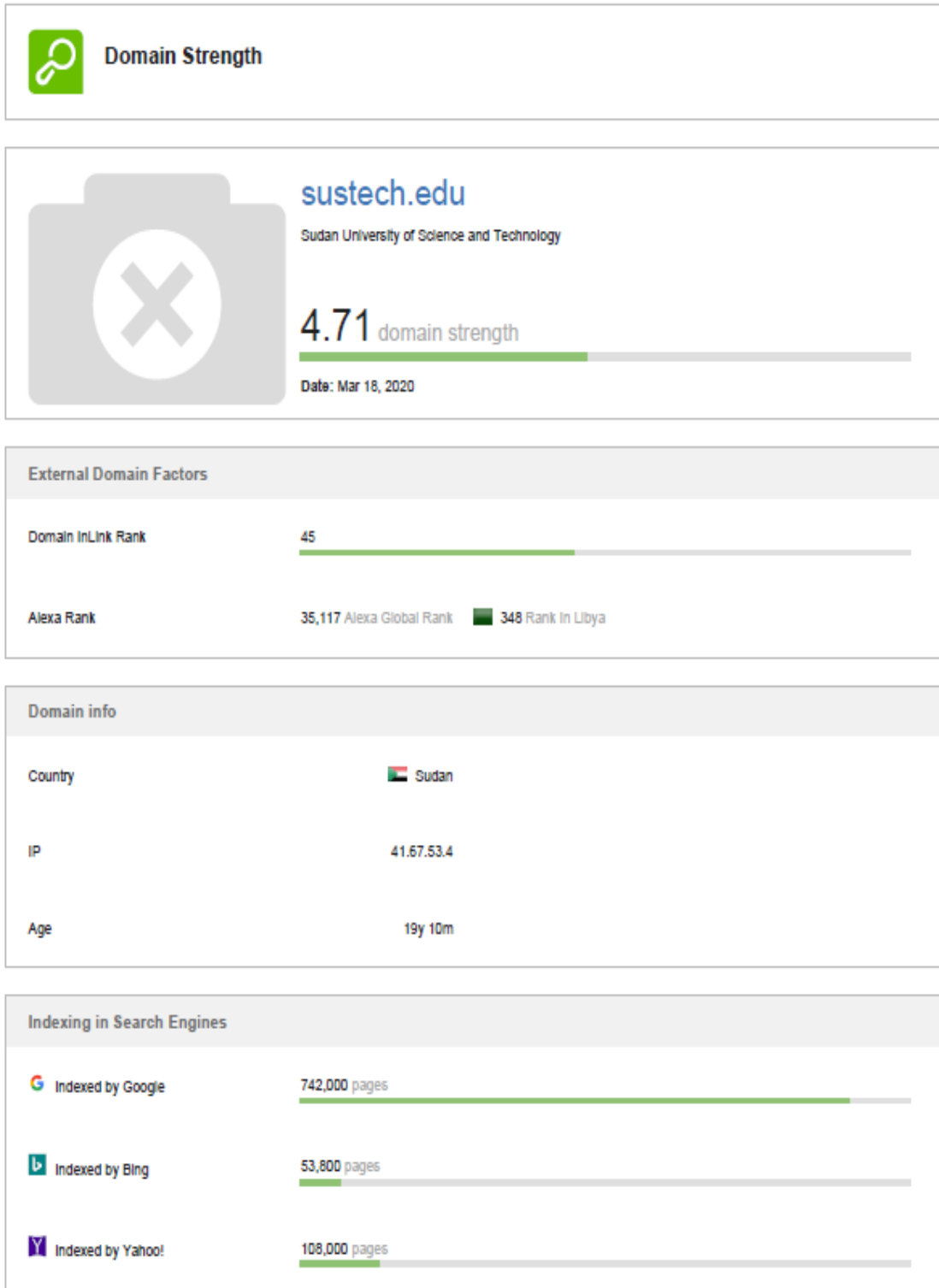


Figure (21) Sudan university back link details

The figure explain some details of Sudan university domain.

### 4.3.1 Google's mobile friendly

Google recently made a big change to their algorithm. This change is officially called (“Mobile-First Indexing”). The bottom line is this: If your site isn't optimized for mobile devices, you're in big trouble.

#### ➤ Mobile friendly test for Sudan university of science and technology

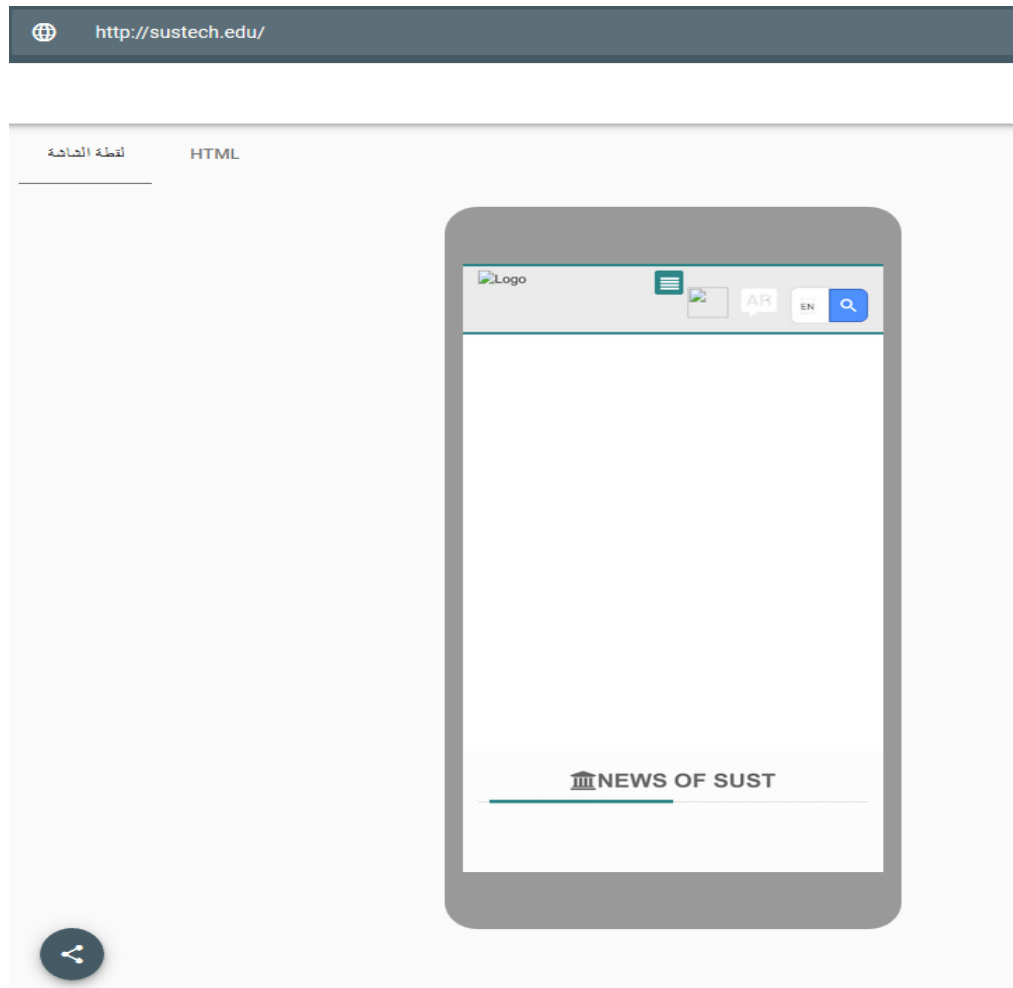




Figure (22) Mobile friendly test for Sudan university

مشاكل تحميل الصفحة عرض التفاصيل 



تم اختبارها في: 2020/10/31 - 1:32 م

## الصفحة مناسبة للجوّال.

هذه الصفحة سهلة الاستخدام على جهاز جوّال

تم اختبارها في: 2020/10/31 - 1:32 م


## تم تحميل أجزاء من الصفحة

تعدّ تحميل بعض موارد الصفحة، وقد يؤثر هذا في تقييم محرك البحث Google لصفحتك وتعرّفه على محتواها. يمكنك حل مشاكل توفّر أي موارد يمكن أن تؤثر في تعرّف محرك البحث Google على محتوى صفحتك.

 مزيد من المعلومات

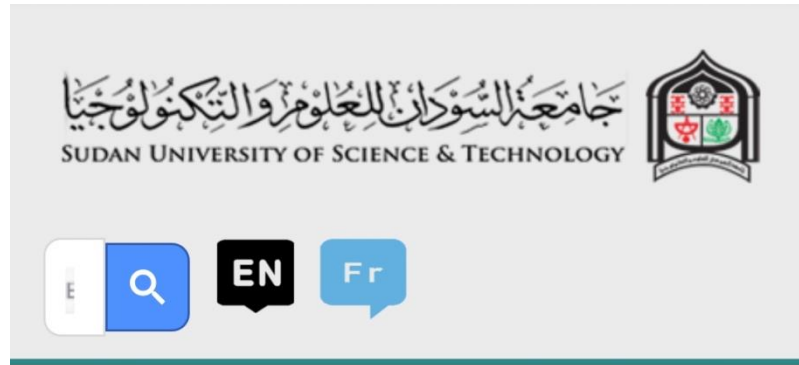
التفاصيل

وكيل المستخدم
Googlebot للهاتف الذكي

تعدّ تحميل 26 موردًا للصفحة. 

المورد	التنوع	الحالة
<a href="https://googleads.g.doubleclick.net/pagead/id">https://googleads.g.doubleclick.net/pagead/id</a>	XHR	تم حظر Googlebot بواسطة robots.txt
<a href="http://sustech.edu/img/fr3.png">http://sustech.edu/img/fr3.png</a>	صورة	خطأ آخر
<a href="http://sustech.edu/css/bootstrap.min.css">http://sustech.edu/css/bootstrap.min.css</a>	ورقة أنماط	خطأ آخر
<a href="http://sustech.edu/css/fonts/fonts.css">http://sustech.edu/css/fonts/fonts.css</a>	ورقة أنماط	خطأ آخر
<a href="http://sustech.edu/images/News/20190723123629739.png">http://sustech.edu/images/News/20190723123629739.png</a>	صورة	خطأ آخر
<a href="http://sustech.edu/img/Conferences-hover.jpg">http://sustech.edu/img/Conferences-hover.jpg</a>	صورة	خطأ آخر
<a href="http://sustech.edu/img/Conferences.jpg">http://sustech.edu/img/Conferences.jpg</a>	صورة	خطأ آخر

Figure (23) Mobile friendly test for Sudan university



## أخبار الجامعة



Figure (24) Mobile friendly test for Sudan university

The page is suitable for viewing on mobile but with some problems loading some of the page resources like pictures and style sheet pages

➤ **Mobile friendly test for Khartoum university**



Figure (25) Mobile friendly test for khartoum university



تم اختبارها في: 2020/10/31 - 12:57 م

### الصفحة غير مناسبة للجوّال

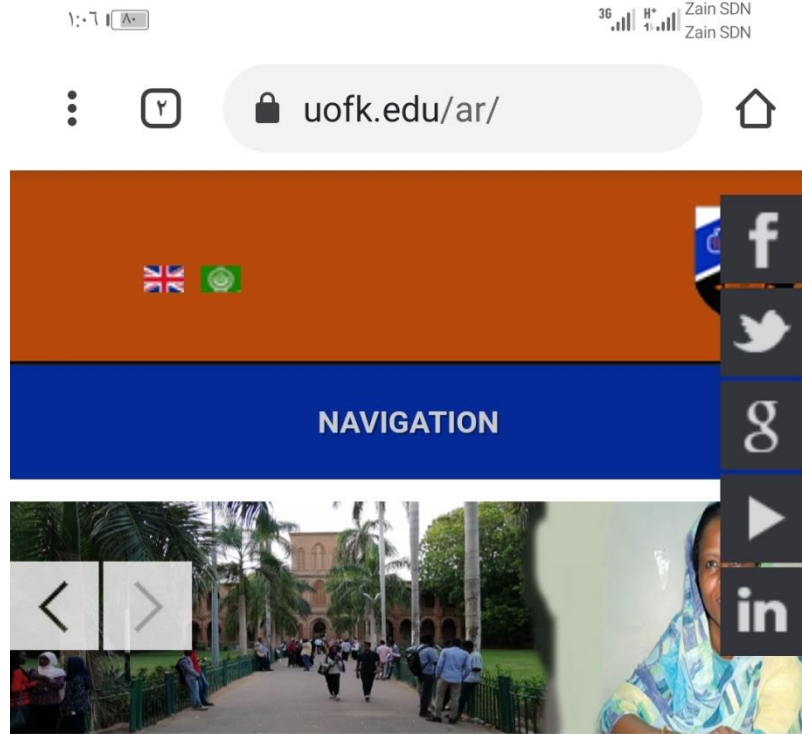
يمكن أن تكون هذه الصفحة صعبة الاستخدام على جهاز جوّال

إصلاح صفحتك في JOOMLA! 🛠

إصلاح المشاكل التالية البالغ عددها 2

- العناصر القابلة للنقر قريباً جداً من بعضها ❌
- النص صغير جداً ولا يمكن قراءته ❌

Figure (26) Mobile friendly test for khartoum university



نظام ادارة التعلم



ت الالكترونية

Figure (26) Mobile friendly test for khartoum university

The page is not suitable for display on mobile, with some problems loading some page resources and not suitable for display



## **CHAPTER FIVE**

### **DISCUSSIONS**

## **CHAPTER FIVE**

### **DISCUSSIONS**

#### **5.1 Introduction**

Through these results, it is clear to us that Sudanese universities have a very low ranking, so these strategies must be applied to the websites of Sudanese universities. There are several strategies for improving the page rank of web site

#### **5.2 Improve your website's user experience**

According to a study on ranking factors by **SEMrush**, the top four ranking factors are website visits, time on site, pages per session, and bounce rate. All four of those are directly related to the experience your website is providing your users. It's pretty simple: If your website is easy and enjoyable to use and offers valuable information, you'll get more visitors, who will stay longer on your site and visit more pages, and that will improve your search rank

#### **5.3 Write great content optimized for SEO**

Having high-quality content on your website is one of the best ways to increase traffic and improve your search rank. In fact, in a recent poll of marketing professionals, 57% said on-page content development was the most effective SEO tactic

#### **5.4 Get more backlinks**

Backlinks—links from other sites to yours—are one of the most heavily weighted Google ranking factors. Incoming links to your website content from high-authority domains show your site's authority, bring in traffic, and improve your search ranking

#### **5.5 Improve your page speed**

Google began using mobile page speed as a ranking factor in July, and while it said the "speed update" will only affect a small percentage of sites, it's still a good idea

to make sure your site is as fast as possible—both to improve your search rank and provide a good user experience.

## **5.6 Fix broken links**

Having quality links on your website is one of the best ways to boost traffic and establish the authority of your site, but links that bring up a 404 error hurt both your user experience and your search rank. Use a tool like Broken Link Check or Dead Link Checker to find any broken links.

## **5.7 Optimize your images**

Optimizing your images is crucial to boosting the speed of your site and making it easier for Google to identify the images on your web pages. Make sure all your website images are compressed, have descriptive filenames, and include alt text.

## **5.8 Use H1 and H2 header tags**

Not only do headers make your content more readable and easier to understand for humans and search engines alike, there is a strong correlation between Google search rank and the use of header tags in the body of your content. Header tags are also a great way to show the structure of your content and emphasize your key points

## **5.9 use Skyscraper SEO Technique**

The Skyscraper Technique is an SEO research and content creation framework that aims to make your site the tallest “skyscraper” on the internet. Like a skyscraper, you become the one everyone is talking about

## **5.10 Metadata**

When designing your website, each page contains a space between the <head> tags to insert metadata, or information about the contents of your page. If you have a CMS site originally produced by the UMC web team will have pre-populated this data for you. However, it is important for you to review and update Metadata as your site changes over time.

### **5.11 Optimize robots.txt file**

Sometime we found that there are lots of pages of website which are indexing and they are not useful such as blog tag pages, hello word, sample pages. We should block these pages to reduce the reducing crawler time wasting.

### **5.12 Add Sitemap files**

Most of the time we found that we don't optimize the sitemap file properly, Sitemap is really helpful for the crawler to know about your web pages, It would be good if we add XML and HTML both the sitemap for the website

### **5.13 Conclusion**

The result of this research is to contribute to the development of the Sudanese public universities' websites by using web exploration techniques to find out the shortcomings and problems to solve them. It became clear that Sudanese universities have a low ranking, so it is necessary to use techniques to improve the appearance of the site in search engines (SEO) to increase the ranking of the site's pages in the search engines.

### **5.14 Future work**

There are many search engine optimization techniques that depend on the page ranking algorithm, which is the first algorithm used by Google company. Google has now updated the algorithm and added other parameters and called it (hummingbird).

## References

- [1] (Duhan, Sharma and Bhatia, 2009) Duhan, N., Sharma, A. K. and Bhatia, K. K. (2009) 'Page ranking algorithms: A survey', *2009 IEEE International Advance Computing Conference, IACC 2009*, (March), pp. 1530–1537. doi: 10.1109/IADCC.2009.4809246.
- [2] Kumar, S. (2012) 'Analyzing the Facebook workload', *Proceedings - 2012 IEEE International Symposium on Workload Characterization, IISWC 2012*, pp. 111–112. doi: 10.1109/IISWC.2012.6402911.
- [3] Volunteers (2020) *wikipedia, web*. Available at: <https://en.wikipedia.org/wiki/PageRank>.
- [4] Kosala, R. and Blockeel, H. (2000) 'Web Mining Research: A Survey', 2(1). Available at: <http://arxiv.org/abs/cs/0011033>.
- [5] Baeza-Yates, R. and Boldi, P. (2010) 'Web structure mining', *Studies in Computational Intelligence*, 311(4), pp. 113–142. doi: 10.1007/978-3-642-14461-5\_5.
- [6] Singh, B. and Kumar Singh, H. (2010) 'Web Data Mining research: A survey', 2010 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2010, pp. 661–670. doi: 10.1109/ICCIC.2010.5705856.
- [7] Rastogi, A. et al. (2012) 'Web Mining: a Comparative Study', *International Journal Of Computational Engineering Research*, 2(2), pp. 2250–3005.
- [8] Bhatia, T. (2011) 'Link analysis algorithms for web mining', *International Journal of Computer Science and Telecommunications Vol. 2*, 4333(2), pp. 243–246.
- [9] Eirinaki, M., Vazirgiannis, M. and Kapogiannis, D. (2005) 'Web path recommendations based on page ranking and Markov models', *Proceedings of the International Workshop on Web Information and Data Management WIDM*, pp. 2–9. doi: 10.1145/1097047.1097050.

[10] Baeza-Yates, R. and Davis, E. (2004) ‘Web Page Ranking using Link Attributes Categories and Subject Descriptors’, Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, pp. 328–329.

[11] Kumar, G., Duhan, N. and Sharma, A. K. (2011) ‘Page ranking based on number of visits of links of Web page’, 2011 2nd International Conference on Computer and Communication Technology, ICCCT-2011, (September), pp. 11–14. doi: 10.1109/ICCCT.2011.6075206.

[12] Ianrogers (2020) ianrogers, uk. Available at: <http://ianrogers.uk/google-page-rank/>.