

# Site Content Analyzer in Context of Keyword density and Key Phrase

Yogesh Yadav- Research Scholar-NIMS University Jaipur (Rajasthan)

Dr. P. K. Yadav- Govt. College Kanwali , Rewari (Haryana)

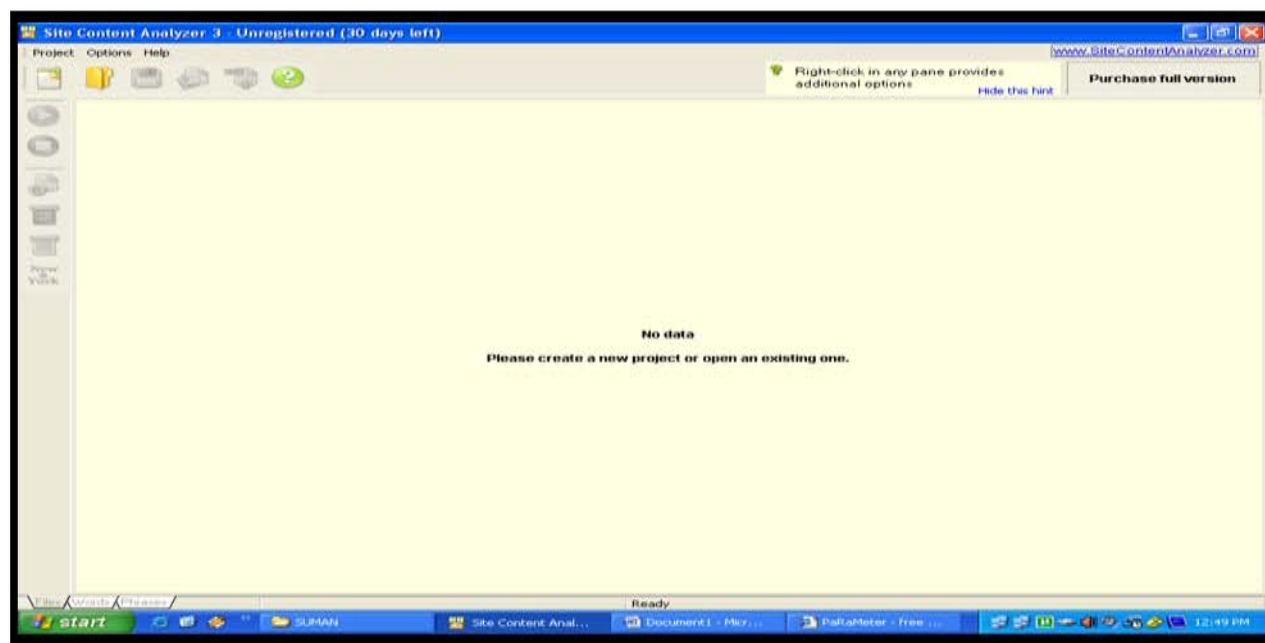
## Abstract

In this paper we describe about the simulation environment in which the research implementation work is done. Here we present the brief description of the Site content analyzer which we have used in web content analysis and keyword density.

## Introduction

### 1. SITE CONTENT ANALYZER

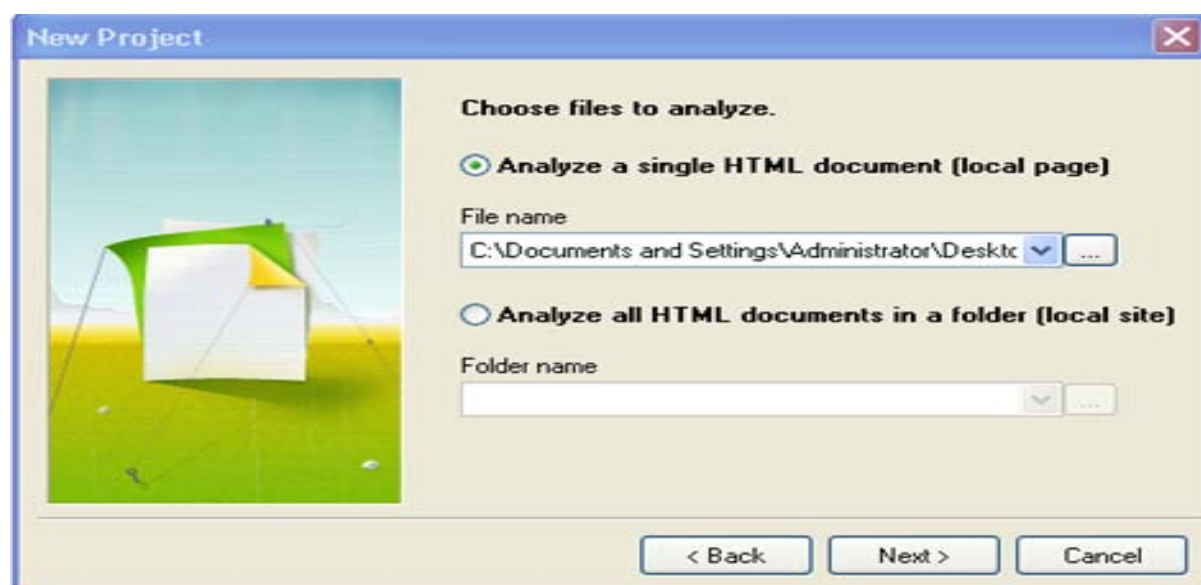
It is well-known (and is confirmed by the most professional SEO experts) that the web ranking of a website mostly depends on two main factors: the number and the quality of inbound links and the amount and the quality of website's content. While determining the quality of the content seems to be one of the most complex tasks in search engine optimization. The site content analyzer tool is the perfect tool which is used to measure the content quality. The Site content analyzer tool can listed such parameters as: keyword density[1], keyword weight[1,2], keyword distribution; discover the most relevant keyword and key phrases, find out the quality of in-site links, overview the whole site-wide picture and many more.



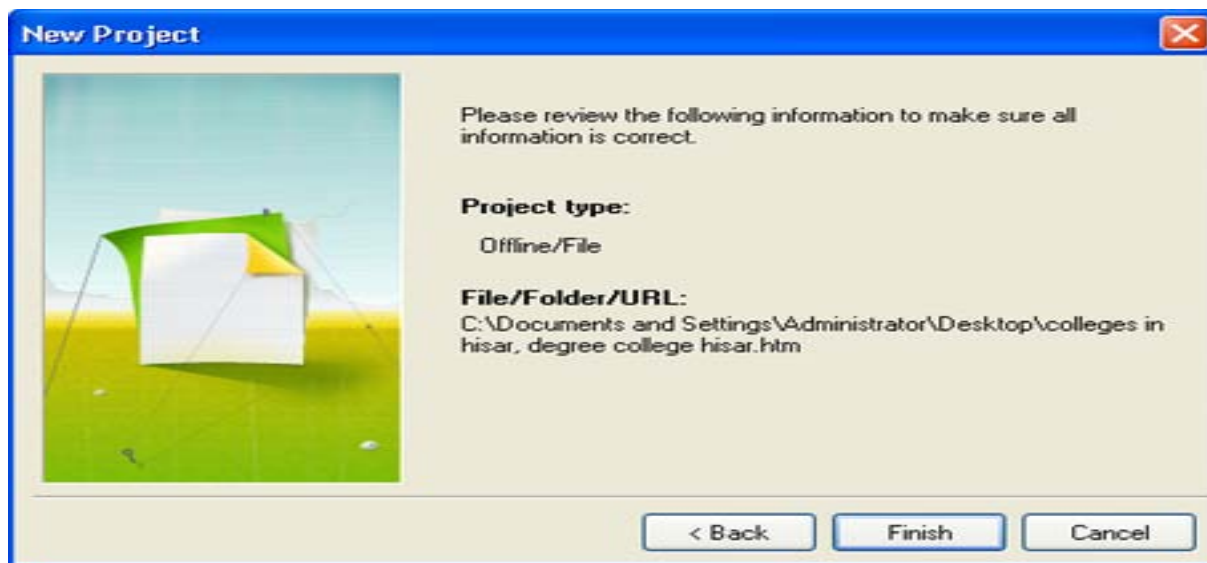
**Fig 1.1 The Site Content Analyzer Tool.**

### 1.1 Input for site content analyzer tool

The input for Site content analyzer tool is the Webpage URL. The Offline URL is stored in the system which is the input of the Site content analyzer tool.



**Fig1.2 The URL Address is input for the Site Content Analyzer tool .**



**Fig 1.3 Processing of Website.**

## **1.2 SITE CONTENT ANALYSER DESCRIPTION**

The Site content analyzer is used to measure the different quality terms.

- ❖ Keyword mode: - It defines the no. of keywords in counted in different tags .
- ❖ Key phrase mode:-
- ❖ Keyword density and keyword weight
- ❖ Keyword cloud
- ❖ Link mode

### **1.2.1 Keywords mode**

Keyword is a single word in the text of a web page. This is a term search engines operate with. Every word on a page is in fact a keyword; it alters the overall rank of a page in search engine's index. The way it does that depends on its weight, distribution and density.

The screenshot shows the Site Content Analyzer 3 interface. The main window displays a table of keyword analysis results. The table has columns for Word, Total, Title, Head, Anch, Alt text, Link, Bold, Italic, Imag, MetaK, MetaD, Com, and Body. The 'Word' column lists various keywords, and the 'Total' column shows their frequency and weight. The 'Title', 'Head', 'Anch', 'Alt text', 'Link', 'Bold', 'Italic', 'Imag', 'MetaK', 'MetaD', 'Com', and 'Body' columns show the distribution of each keyword across these HTML tags. The 'Word' column is sorted by total frequency, with 'hisar' having the highest frequency (16).

Word	Total	Title	Head	Anch	Alt text	Link	Bold	Italic	Imag	MetaK	MetaD	Com	Body
hisar	16 [1.1... 2 [1.42...]	-	-	2 [1.42...]	-	-	2 [1.42...]	-	-	-	8 [5.67...]	-	4 [2.84...]
college	15 [1.0... 1 [0.71...]	-	-	5 [3.55...]	-	-	5 [3.55...]	-	-	-	6 [4.26...]	-	3 [2.13...]
degree	5 [3.55... 1 [0.71...]	-	-	1 [0.71...]	-	-	1 [0.71...]	-	-	-	2 [1.42...]	-	1 [0.71...]
colleges	4 [2.84... 1 [0.71...]	-	-	1 [0.71...]	-	-	1 [0.71...]	-	-	-	1 [0.71...]	-	1 [0.71...]
jat	4 [2.84...]	-	-	3 [2.13...]	-	-	3 [2.13...]	-	-	-	-	-	1 [0.71...]
crm	3 [2.13...]	-	-	2 [1.42...]	-	-	2 [1.42...]	-	-	-	-	-	1 [0.71...]
region	3 [2.13...]	-	-	-	-	-	-	-	-	-	-	-	3 [2.13...]
academic	2 [1.42...]	-	-	1 [0.71...]	-	-	-	-	-	-	-	-	1 [0.71...]
read	2 [1.42...]	-	-	2 [1.42...]	-	-	-	-	-	-	-	-	-
haryana	2 [1.42...]	-	-	1 [0.71...]	-	-	1 [0.71...]	-	-	-	-	-	1 [0.71...]
placement	2 [1.42...]	-	-	2 [1.42...]	-	-	-	-	-	-	-	-	-
best	1 [0.71...]	-	-	1 [0.71...]	-	-	1 [0.71...]	-	-	-	-	-	-
memorial	1 [0.71...]	-	-	1 [0.71...]	-	-	1 [0.71...]	-	-	-	-	-	-
inaugurated	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
governor	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
april	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
contact	1 [0.71...]	-	-	1 [0.71...]	-	-	-	-	-	-	-	-	-
career	1 [0.71...]	-	-	1 [0.71...]	-	-	-	-	-	-	-	-	-
welcome	1 [0.71...]	-	-	1 [0.71...]	-	-	-	-	-	-	-	-	-
ram	1 [0.71...]	-	-	1 [0.71...]	-	-	1 [0.71...]	-	-	-	-	-	-
chajju	1 [0.71...]	-	-	1 [0.71...]	-	-	1 [0.71...]	-	-	-	-	-	-
shri	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
led	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
elite	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
justice	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
decided	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
lamba	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
city	1 [0.71...]	-	-	1 [0.71...]	-	-	1 [0.71...]	-	-	-	-	-	-
chakravarti	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
educationally	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
rural	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
backward	1 [0.71...]	-	-	-	-	-	-	-	-	-	-	-	1 [0.71...]
has	1 [0.71...]	-	-	-	-	-	-	-	-	-	1 [0.71...]	-	-

Fig 1.4 The output defines the keyword mode.

#### 1.2.1.1 Keyword mode description

In the description of keyword mode the different calculation can be done by Site content analyzer. The keyword mode considering the keywords which are important. In the keyword mode important keyword has on-page optimization techniques values. Each keyword has explained in the different tags .

- Title
- Head
- Anchor
- Alt text
- Link
- Bold
- Italic
- Image
- Meta keywords
- Meta description
- Comment
- Body

In keyword mode the important keyword is counted how much occur in the above tags which are used in the website pages coding.

#### 1.2.2 Key phrases mode

In key phrases mode the calculation of key phrases could be done. The important key phrases which are included in the website coding is showing. The weight, count and density of key phrases in a website are calculated in key phrases mode.

### 1.2.2.1 Key phrase

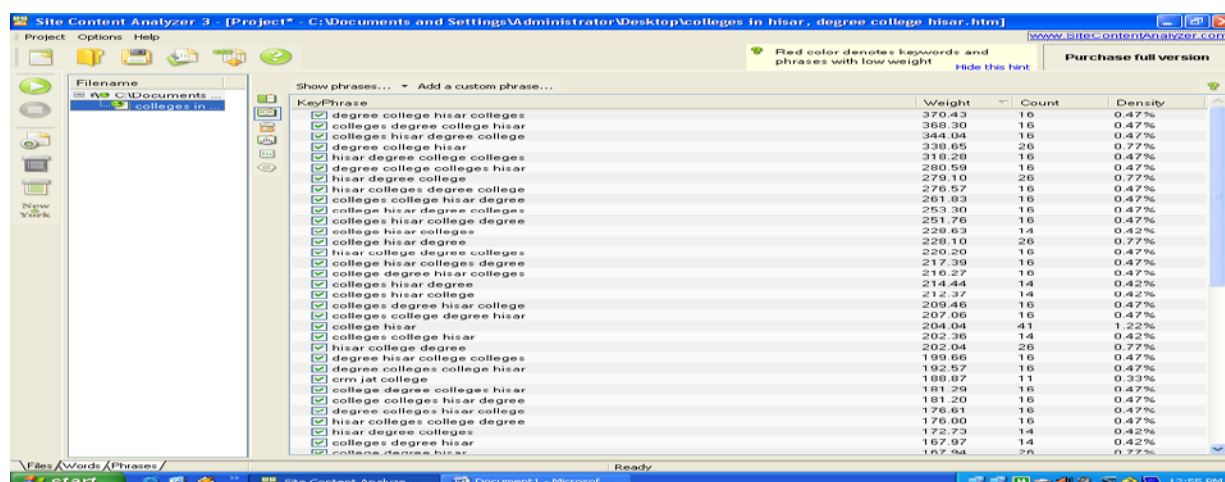
Key phrase is a phrase of two or more keywords. Site Content Analyzer has got an ability to analyze not only single keywords, but also a phrase which is the unique ability and is very convenient. Searchers rarely put a single-word query in the Google search box. Most searched queries are always 2- or 3-word phrases. Site Content Analyzer treats several keywords as a phrase if they are located no more than 10 keywords far away from each other. For instance, in the above sentence: "Site Content" is a phrase, while "Site located" is not.

### 1.2.2.2 Key phrase density

Key phrase density is a relative value that represents how many times Site Content Analyzer meets a phrase on the given page. For instance, if a phrase meets 4 times and there are 100 key phrases in total, the density of a phrase is  $4 / 100 * 100\% = 4\%$

### 1.2.2.3 Key phrase weight

Key phrase weight is value that shows the importance of a phrase. The weight is summarized from the weights of all keywords that build the key phrase with respect to their relative position and distance. This means in particular, that a key phrase built of two words that are located far away from each other has lesser weight than the one built of two words standing next to each other in the text of a web page.



KeyPhrase	Weight	Count	Density
degree college hisar colleges	370.43	16	0.47%
colleges degree college hisar	368.30	16	0.47%
colleges hisar degree college	344.04	16	0.47%
degree college hisar	338.65	26	0.77%
hisar degree college colleges	318.28	16	0.47%
degree college colleges hisar	280.59	16	0.47%
hisar degree college	279.10	26	0.77%
hisar colleges degree college	276.57	16	0.47%
colleges college hisar degree	261.83	16	0.47%
college hisar degree colleges	253.30	16	0.47%
colleges hisar college degree	251.76	16	0.47%
college hisar colleges	228.63	14	0.42%
college hisar degree	228.10	26	0.77%
hisar college degree colleges	220.20	16	0.47%
college hisar colleges degree	217.38	16	0.47%
college degree hisar colleges	216.27	10	0.47%
colleges hisar degree	214.44	14	0.42%
colleges hisar college	212.37	14	0.42%
colleges degree hisar college	209.46	16	0.47%
colleges college degree hisar	207.06	16	0.47%
college hisar	204.04	41	1.22%
colleges college hisar	202.36	14	0.42%
hisar college degree	202.04	26	0.77%
degree hisar college colleges	199.66	16	0.47%
degree colleges college hisar	192.57	16	0.47%
crm jat college	188.67	11	0.33%
college degree colleges hisar	181.29	16	0.47%
college colleges hisar degree	181.20	16	0.47%
degree colleges hisar college	176.61	16	0.47%
hisar colleges college degree	176.00	16	0.47%
hisar degree colleges	172.73	14	0.42%
colleges degree hisar	167.97	14	0.42%
colleges degree hisar	167.94	26	0.77%

Fig 1.5 This output defines the key phrases mode

## 1.2.3 Keyword density and keyword weight mode[1]

In the keyword density and keyword weight mode the calculation of important keywords are often. The two terms can be calculated.

### ***1.2.3.1 Keyword density***

Keyword density is relative value that reflects the frequency of a keyword on a page. Unlike keyword count, which is absolute value, keyword density is relative to the entire number of keywords on a page. This allows to user to compare keywords on a pages with different amount of text on them.

Keyword density is calculated as follows:  $\text{Density} = \text{Count} / \text{Total Count} * 100\%$

For instance, if a keyword meets 5 times and there are 200 keywords in total on a page, its keyword density is:  $5 / 200 * 100\% = 2.5\%$

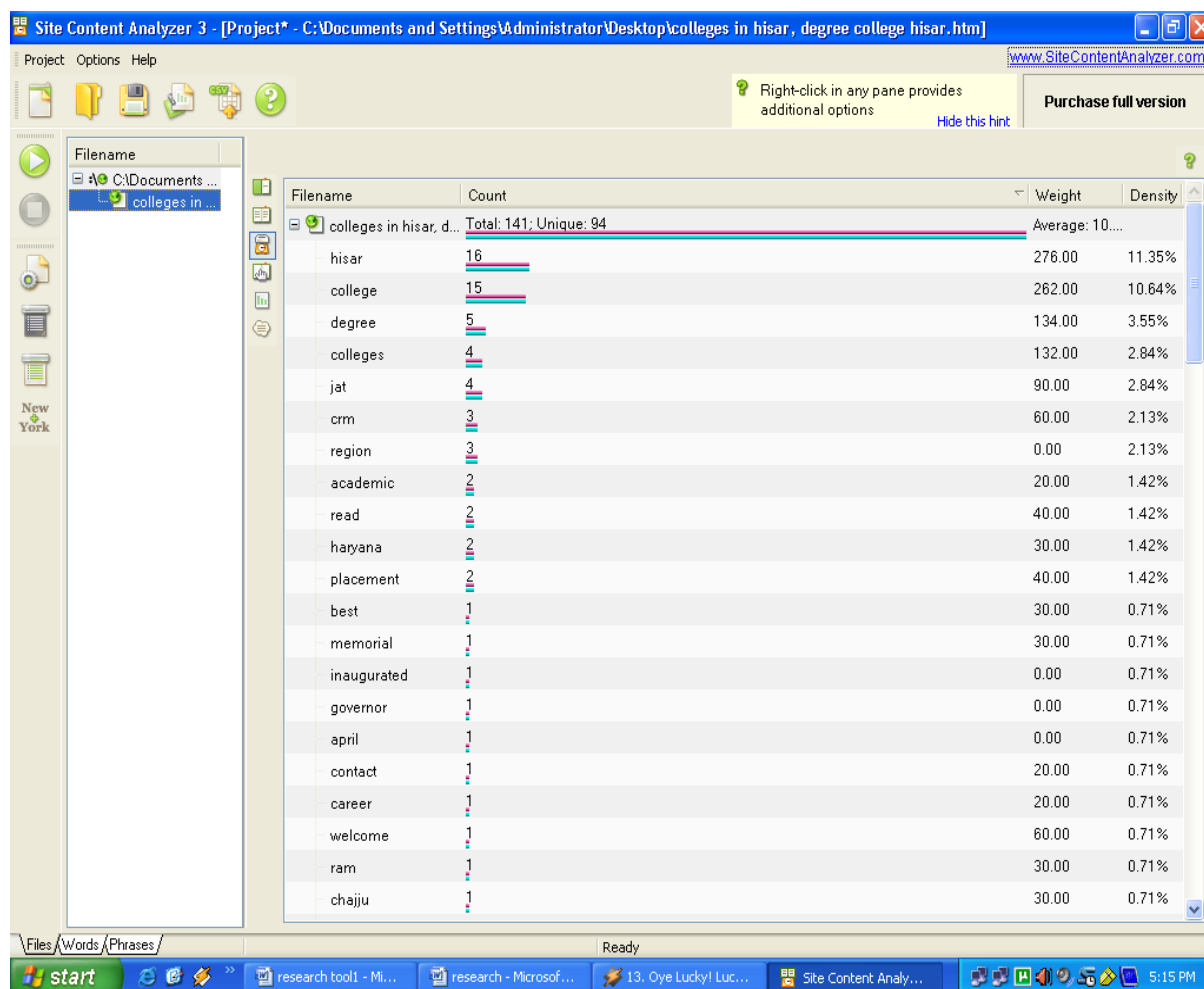
Keyword density is very important in SEO, because it tells search engines of the frequency of this or that keyword.

### ***1.2.3.2 Keyword weight***

Keyword weight is one of the most important keyword parameters in SEO. It shows the importance of a keyword to search engines according to the tags it is enclosed within. For instance, if a user has a page about cooking then put the word "cooking" in the title of the page and perhaps mention it somewhere in the headings of articles. So, a search engine indexing then the page, will notice "cooking" in title and headings and assigns it high weight signifying the high chance that the page is about cooking. In Site Content Analyzer can assign each tag with some weight value. For instance, by default the <title> tag is assigned with 100 while the <strong> tag gets 10. If a keyword occurs several times on a page, its weight is summarized. For example, if a keyword in title, two times in bold text and once in anchor text it gets by default:  $100 \text{ (from title)} + 2 * 10 \text{ (from bold text)} + 20 \text{ (anchor text)} = 140$ .

### ***1.2.3.3 Keyword distribution***

While keyword density allows evaluating the average "popularity" of a keyword on a page, a keyword distribution allows seeing how this popularity is distributed across the volume of a page or website. Keywords in the beginning of a page usually are more important to search engines, than the ones somewhere in the middle.



**Fig 6.6 The output is keyword density and keyword weight mode**

### 1.2.4 Keyword cloud mode

The keyword cloud defines the appearances of important keywords. The keyword cloud is a visual depiction of keywords used on a website; keywords having higher density are depicted in large fonts. The main keywords of a website appear in large fonts.

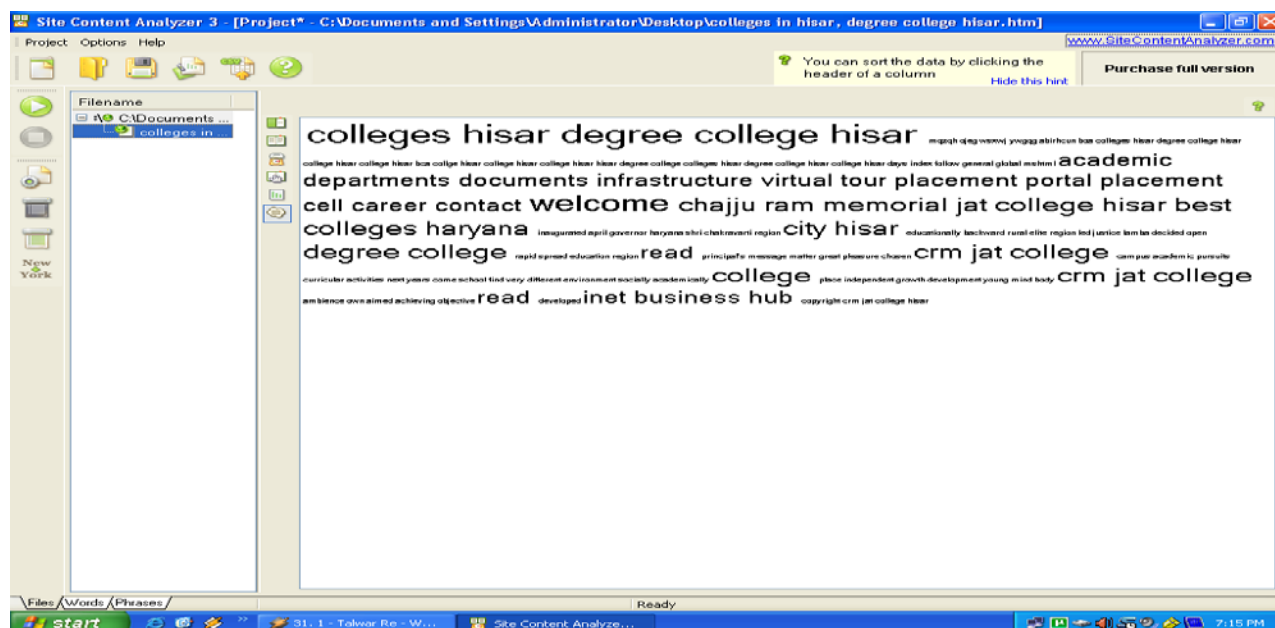


Fig 6.7 The output is of keyword cloud.

### 1.2.5 Linked Mode

In linked mode the list of linked web pages can be displayed. The linked web pages is the interlinking web pages in a website. The websites pages are linked that list are to be shown in linked mode.

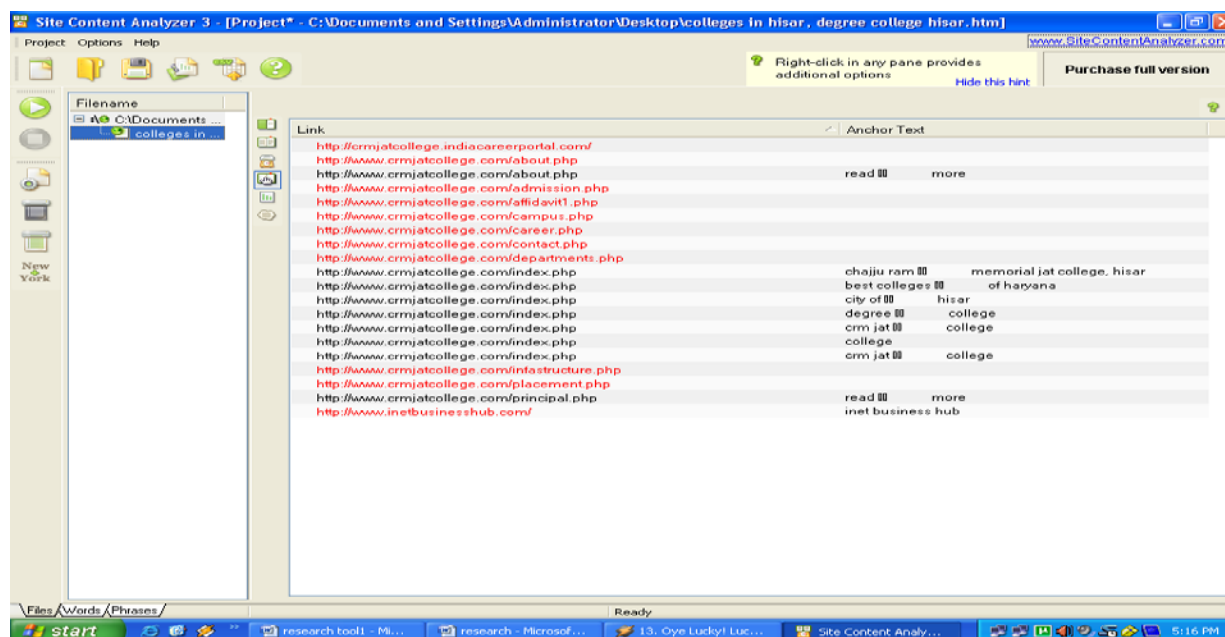


Fig 6.8 The output defines the number of linking pages in a website.

## 2 Report Generation

The Site content analyzer tool can be generated a report of all contents. In this report all about the webpages contents. The different calculation of factors that can be included in the website is defined in this report.

The report includes the different information detail

### 2.1 File information

- File summary
- Keyword density report
- Key phrases report
- Keyword weight report
- Links report

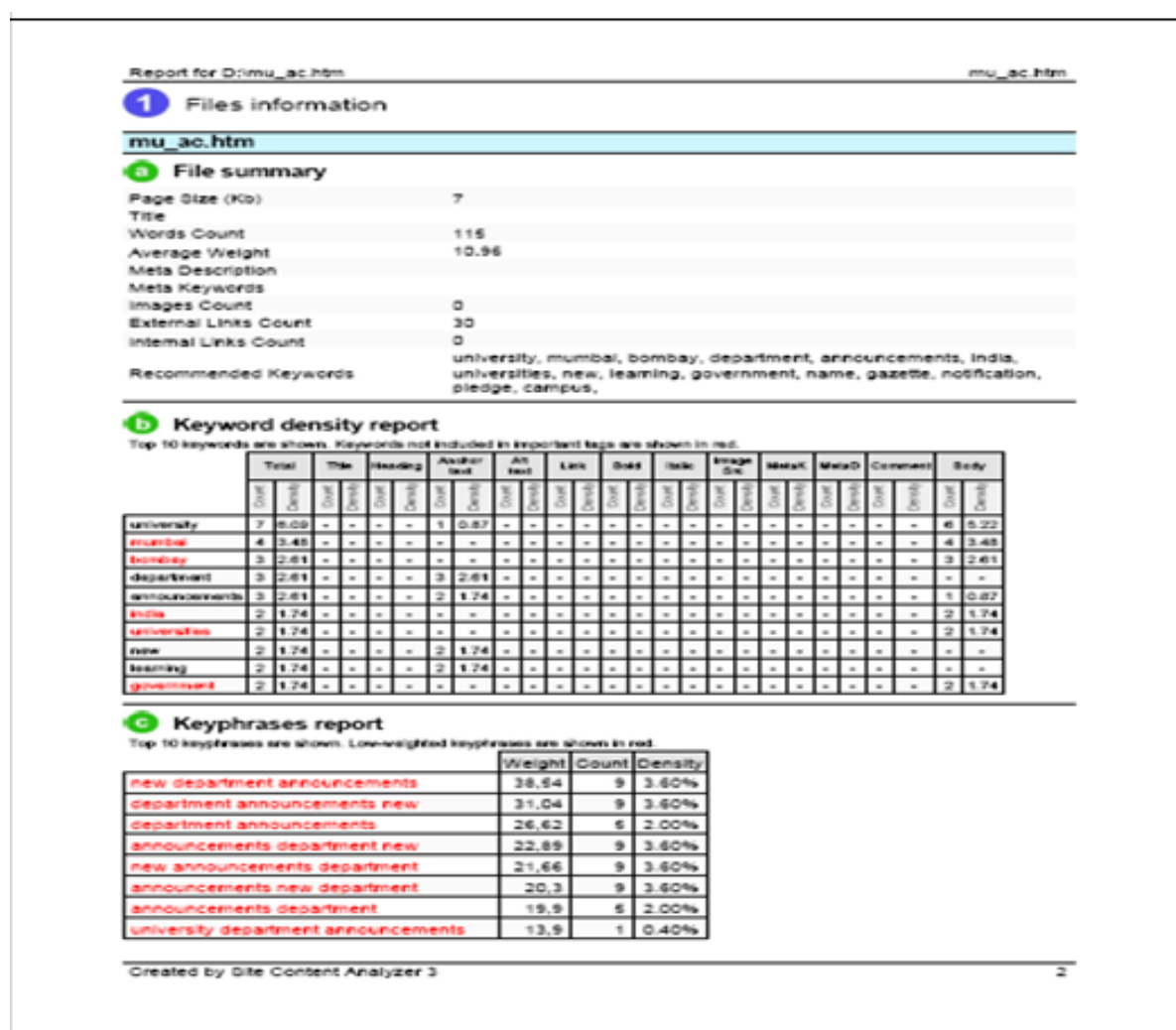


Fig 6.9 The output of File information.

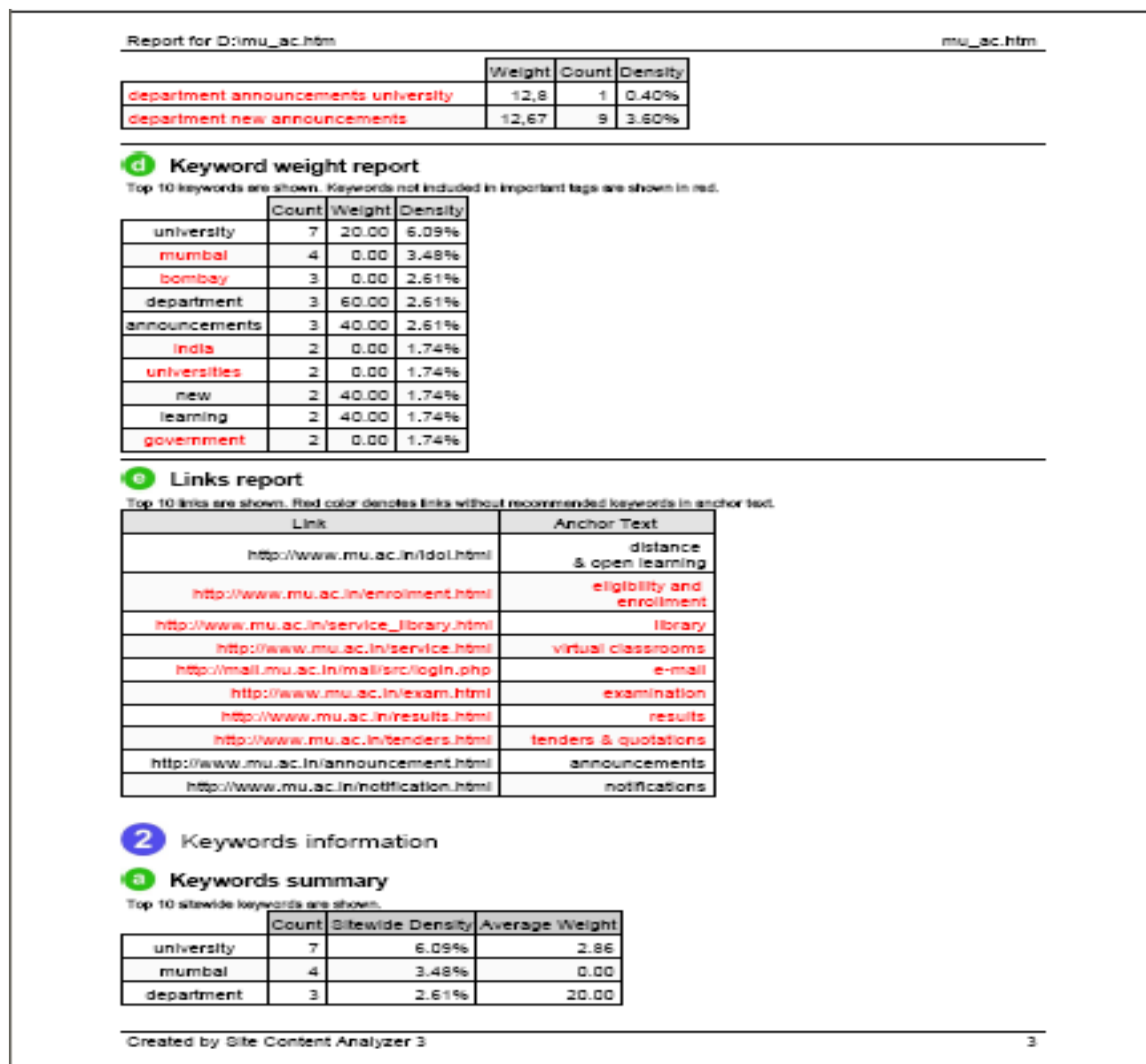


Fig 6.10 The output of File Information.

## 2.2 Keywords information

- Keywords summary
- Keywords details

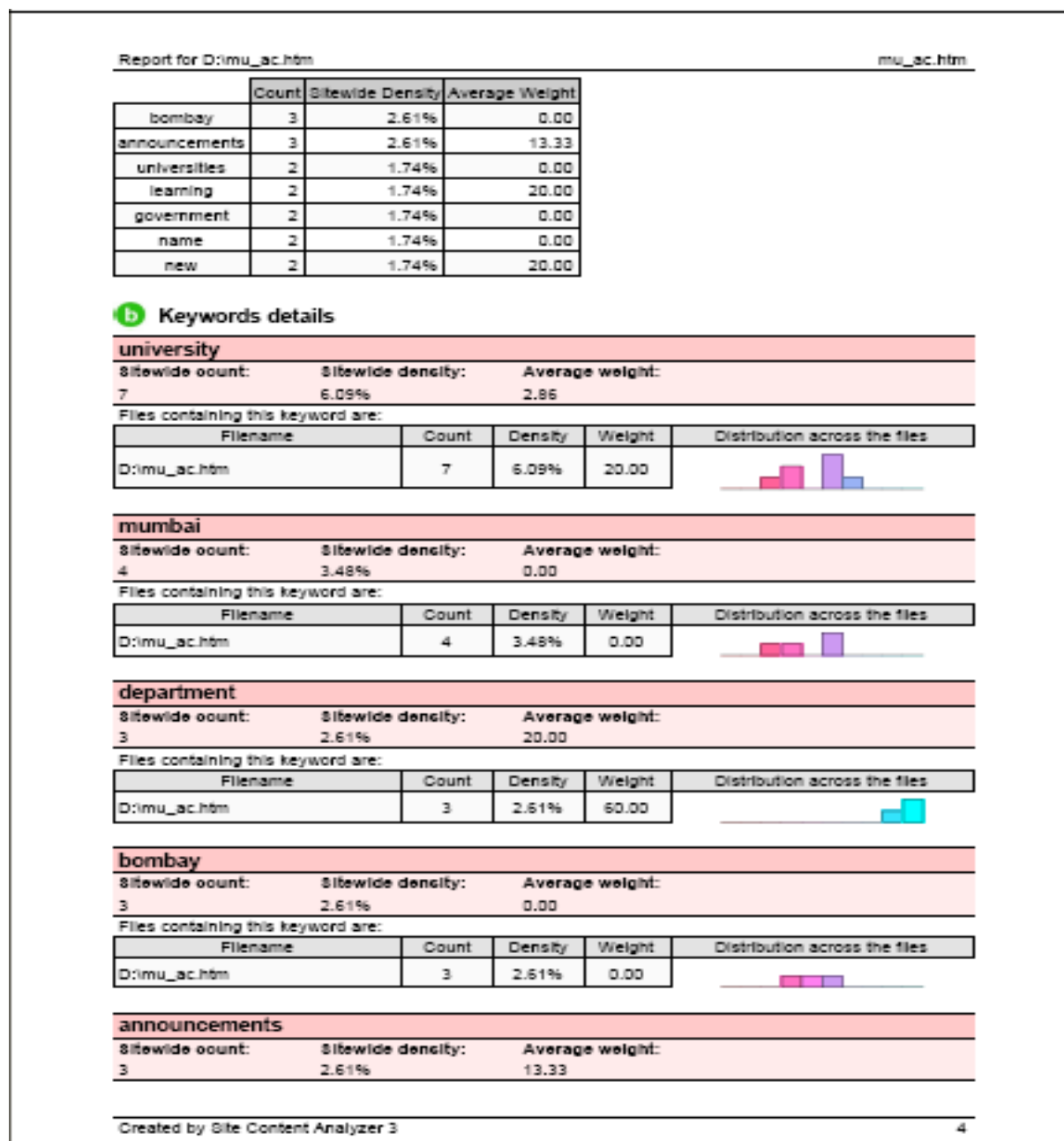


Fig 6.11 The output of keywords Information

## 2.3 Key phrases information

- Key phrases summary
- Key phrases details

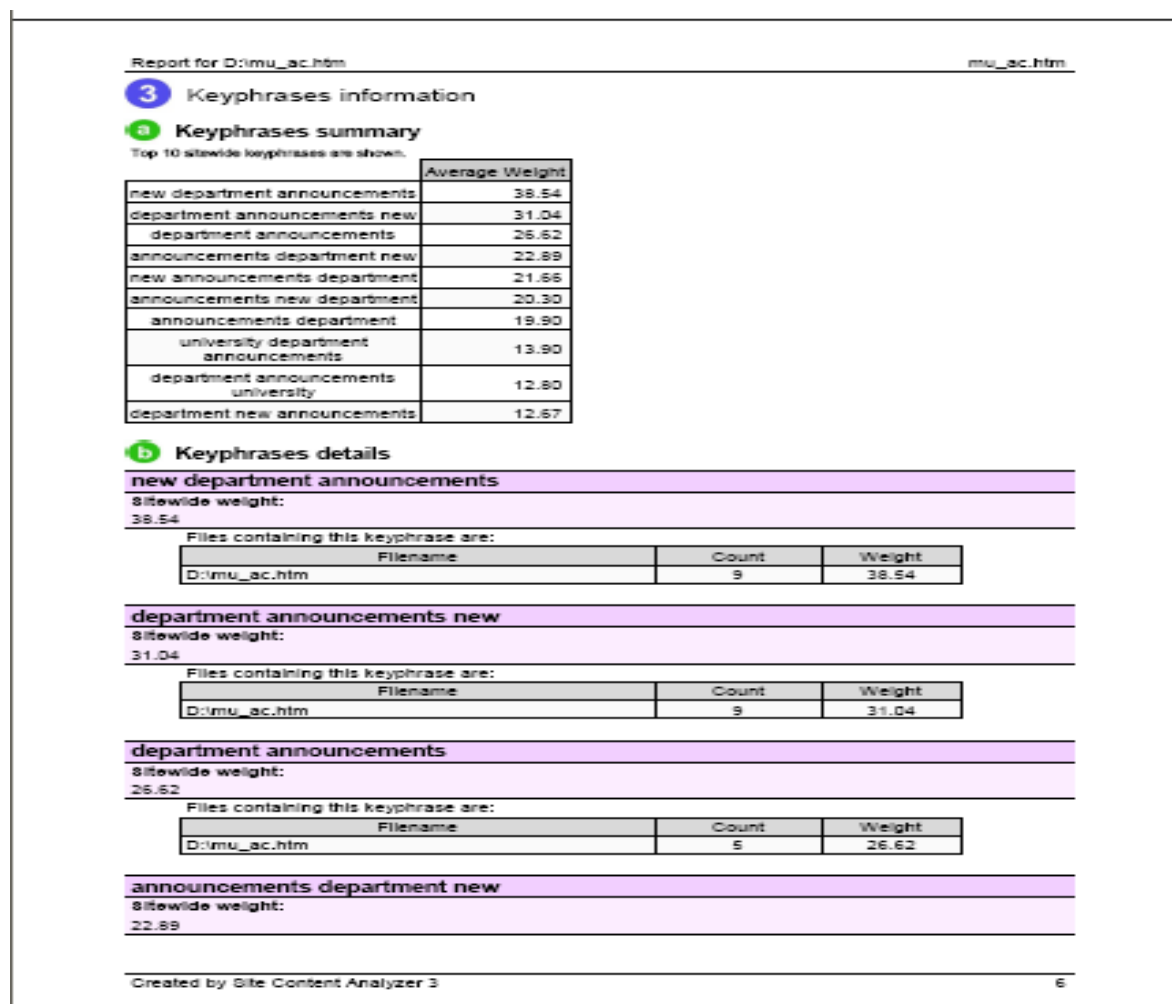


Fig 6.12 The output of key phrases information.

## CONCLUSION

At last the summary of this paper has defined the two tool which has major importance in our work. The Site Content Analyser tool is used to define the keyword density and weight of keyphrases. The Backlink analyser tool used for analysis of internal links and quality of internal links linked to the website.

## REFERENCES

- [1] Wayne Hulbert, “Keyword Density: SEO Considerations”, May.2005  
[www.webpronews.com/news/ebusinessnews/wpn-4520050501KeywordDensitySEOconsiderations.html](http://www.webpronews.com/news/ebusinessnews/wpn-4520050501KeywordDensitySEOconsiderations.html)
- [2] Kevin Curran, “Tips for achieving high positioning in the results pages of the major search engines”, 2004
- [3] iProspect, “iProspect Search Engine User Attitudes”, May.2004;  
[www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf](http://www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf)
- [4] Lee Underwood, “A Brief History of Search Engines”;  
[www.webreference.com/authoring/search\\_history/](http://www.webreference.com/authoring/search_history/)