Web Searching and SVD

Akshath Suresh R Rao, AM.EN.U4AIE20009
Department Of Computer Science (Artificial Intelligence)
School Of Engineering,Amrita Vishwa VidyaPeetham
Clappana PO ,Kollam,Kerala,India
amenu4aie20009@am.students.amrita.edu

Dhanush Krishna R, AM.EN.U4AIE20021
Department Of Computer Science (Artificial Intelligence)
School Of Engineering,Amrita Vishwa VidyaPeetham
Clappana PO ,Kollam,Kerala,India
amenu4aie20021@am.students.amrita.edu

Dhruv V S, AM.EN.U4AIE20022
Department Of Computer Science (Artificial Intelligence)
School Of Engineering,Amrita Vishwa VidyaPeetham)
Clappana PO ,Kollam,Kerala,India
amenu4aie20022@am.students.amrita.edu

Durgapu Sathvik, AM.EN.U4AIE20024
Department Of Computer Science (Artificial Intelligence)
School Of Engineering,Amrita Vishwa VidyaPeetham)
Clappana PO ,Kollam,Kerala,India
amenu4aie20024@am.students.amrita.edu

Pranav Jayasankar Nair, AM.EN.U4AIE20056
Department Of Computer Science (Artificial Intelligence)
School Of Engineering,Amrita Vishwa VidyaPeetham)
Clappana PO ,Kollam,Kerala,India
amenu4aie20056@am.students.amrita.edu

***Abstract*—Search engines present a variety of information, this information makes the user takes time to discover the appropriate data. The best web information not only dominated in a search engine, sometimes found on different machines with more information on specific topics. In this research, search engine optimization using Latent Semantic Indexing method of collaborative search engine. Information will be on the filter before it is sent to the user. The results of the study are expected to sort information on multiple search engines and provide a recommendation urls that correspond to the keywords to help users find the right information. There are many search engines in the web, but they return a long list of search results, ranked by their relevancies to the given query. Web users have to go through the list and examine the titles and (short) snippets sequentially to identify their required results. In this paper we present how usage of Singular Value Decomposition (SVD) as a very good solution for search results clustering.**

***Index Terms*—LSI, Big Data, LSI, Web Scraping, Search Engine, SVD**

I. Introduction

Web search engines are used to search for information on the Internet. Search results are sorted by relevance keywords and page content on a web server. Targeted search on the web server has various types of content such as text files, images, audio, and video. Search engines present a lot of text. It is a target that is often sought after by users in searching information on the internet.Search engines like Ask, Bing, Google, and Yahoo as an information center that has big data, they are needed by the user in search of information on the internet. If important information was not in the first page, web page change, and failure in the access server, the user takes a long time to select the information on the big data.For this reason we propose optimization techniques search engines, the search results from search engines will be crawling and re-indexed using Latent Semantic Indexing (LSI). In doing search engine optimization, LSI method concerned with the word contained in the document regardless of its linguistic characteristics.

The main part of the LSI is Singular Value Decomposition (SVD). SVD will compress the size of great information to form a smaller space, but represents the meaning of the value of the initial information. The major difficulties are the complicacy of the content and the classification of the huge information in the web, and identifying and naming topics and relationships between these topics. In this situation, clustering data gives us a good result for data analysis. We can use the search result clustering in width area from different fields. In this paper we present one of the methods for clustering data to be used in the search result clustering. We use the singular value decomposition as a mathematical method to reduce a big value of objects by combining the attributes of these objects.

A. *Introduction to SVD (Singular Value Decomposition)*

In linear algebra, the singular value decomposition (SVD) is a factorization of a real or complex matrix that generalizes the eigen decomposition, which only exists for square normal matrices.

Singular Value Decomposition (SVD) breaks a *m* x *n* matrix A into three matrices U, Σ and V such that A = UΣV^T. Where U is a (nk) orthogonal matrix whose column vectors are called left singular vectors of A, V is a (km) orthogonal matrix whose column vectors are termed the right singular vectors of A, and Σ is a (kk) diagonal matrix having the Singular values of A ordered decreasingly. Columns of U form an orthogonal basis for the column space of A.

The SVD is not unique. It is always possible to choose the decomposition so that the singular values Σ_{ii} are in descending

order. In this case, Σ (but not always U and V) is uniquely determined by M.

Since U and V* are unitary, the columns of each of them form a set of orthonormal vectors, which can be regarded as basis vectors. The matrix M maps the basis vector V_i to the stretched unit vector $\sigma_i U_i$. By the definition of a unitary matrix, the same is true for their conjugate transposes U^* and V, except the geometric interpretation of the singular values as stretches is lost. In short, the columns of U, U^*, V, and V^* are orthonormal bases. When the M is a normal matrix, U and V are both equal to the unitary matrix used to diagonalize M. However, when M is not normal but still diagonalizable, its eigen decomposition and singular value decomposition are distinct.

Mathematical applications of the SVD include computing the pseudo inverse, matrix approximation, and determining the rank, range, and null space of a matrix. The SVD is also extremely useful in all areas of science, engineering, and statistics, such as signal processing, least squares fitting of data, and process control.

B. Introduction to LSI (Latent Semantic Indexing)

Latent Semantic Indexing – LSI is a popular term amongst SEO (Search Engine Optimization) experts and some reputed influencers. LSI is an excellent application for ensuring organic search achievement.

LSI is a process by which search engines (such as Google) discover ways to make content and terms work simultaneously. LSI helps in collating content that is beyond keywords, key phrases, or synonyms.Search engine optimization boosts website traffic through the exact keywords on your pages but, as time goes, keyword-stuffed content, paragraphs and meta tags are becoming outdated SEO techniques. And, that is where LSI comes to action- a necessary practice of using LSI keywords to keep sites relevant amidst the vast search engine world.

Points to remember while working with LSI

- Understanding LSI requires basic familiarity with internet searches, keywords, and SEO.
- Search engines are programmed in a way that it can recognize your content and identify related synonyms alongside your keywords.
- LSI are not just simple keywords synonyms, instead, it is a collection of words that are frequently searched relating to the same topic/content

Example 1

"Samsung" and "Android" are LSI keywords. Why? They are often found together; they share the same context but not synonyms.

Example 2

Suppose you are blogging about 'Apple' but how will the search engine differentiate if it is a 'fruit' or 'the company'. So what's the solution?

- If your page is about the fruit apple and encompasses keywords like 'red apple', 'green apple', 'nutrition', 'healthy fruit', 'eating an apple' then the search engine distinguishes that your page is about the 'apple' fruit.
- But, if your website consists of key phrases like 'I-pad', 'I-Tunes', 'Apple News' and 'Apple iPhone' and 'Apple Store' then the search engine understands will establish your page for the 'Apple' company.

Let the document collection be represented by a D x n matrix X =$[x1|...|xn]$, where the columns are document BOW vectors, D is the vocabulary size, and n is the number of documents.

LSI is SVD applied to X:

$X_{Dxn} = U_{Dxm} S_{mxm} V_{nxm}{}^T$

Furthermore, one only keeps the largest d ¡¡ m = min(D, n) singular values.That is, let U_{Dxd} be the first d columns of U, S_{dxd} be the first (d x d) submatrix of S, V_{nxd} be the first d columns of V . Then $U_{Dxd} S_{dxd} V_{nxd}{}^T$ is the best rank-d approximation to X in the least square sense.The d columns of U_{Dxd} defines the new, rotated, lower dimensional coordinate system. The n columns of $S_{dxd} V_{nxd}{}^T$ are the new coordinates of each document after dimensionality reduction.

The new coordinate system allows a natural way to perform dimensionality reduction for points not in the dataset. For a new test document x its new coordinate is $U^T x$. If xcoincide with an existing document x_i, it will have the same new coordinates $U^T x_i$ = S $V_i{}^T$ as above.

An alternative would be to add x to the dataset, and compute SVD on the n+ 1 documents. In general, this will produce a slightly different dimension reduction solution. It is much more computational intensive.

II. Literature Review

Over the last 20 years the number of internet users has grown exponentially with time. Trying to extract information from this exponentially growing resource of material can be a daunting task. Information retrieval is an interdisciplinary science, which is concerned with automated storage and retrieval of documents.

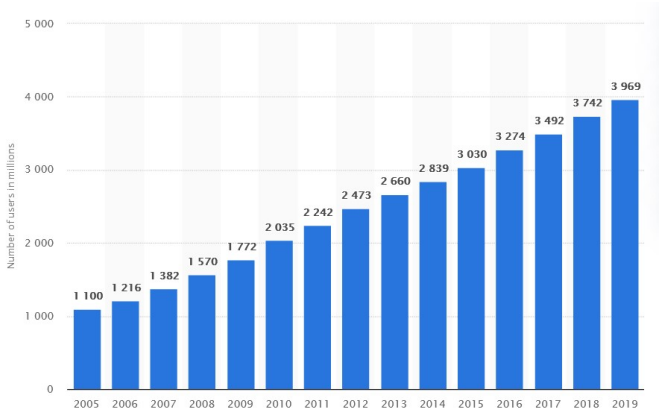


Fig. 1. Data obtained from the website https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/

One of the largest digital data sets is the Internet with more than 20 billion web pages.Extraction of a small subset

of useful information from this huge digital data set requires a web search engine. The grandfather of search engines, which was really not a search engine, was called Archie and was created in 1990 by Peter Deutsch, Alan Emtage, and Bill Wheelan. After Archie,followed a whole array of search engines; World Wide Web Wanderer (1993), Excite (1993), Yahoo(1994), Infoseek (1994), Lycos (1994), AltaVista (1995), and Google (1997) to name a few.

A search engine is a software system that is designed to carry out web searches. They search the World Wide Web in a systematic way for particular information specified in a textual web search query. The search results are generally presented in a line of results, often referred to as search engine results pages (SERPs) The information may be a mix of links to web pages, images, videos, infographics, articles, research papers, and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler.

To handle this, we used the LSI Method to get a good results of the input which we are searching from the data we already stored.

III. DATASETS

A data set is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. Data sets can also consist of a collection of documents or files.

The data set which we are using in this article to do the web searching using SVD is ”cars for sale”. This data set contains a title, link and snippet as columns and each row contains the records of the cars for sale. The data set contains about 2000 rows and 3 columns.

The data set is as follows.

	A	B	C
1	Used Chevrolet Malibu for Sale (with Photos) - CARFAX	https://www.carfax.com/Used-Chevrolet-Malibu_w117	Every used car for sale comes with a free CARFAX Report. We have 9344 Chevrolet Malibu vehicles for sale that are reported accident free, 7718 1-Owner cars,Â ...
2	Used Chevrolet Malibu for Sale (with Photos) - CarGurus	https://www.cargurus.com/Cars/l-Used-Chevrolet-Malibu-d622	Listings 1 - 15 of 24608 ... Save $4486 on a used Chevrolet Malibu near you. Search over 36800 listings to find the best local deals. We analyze millions of usedÂ ...
3	Used Chevrolet Malibu for Sale Near Me | Cars.com	https://www.cars.com/shopping/chevrolet-malibu/	Shop Chevrolet Malibu vehicles for sale at Cars.com. Research, compare and save listings, or contact sellers directly from 16947 Malibu models nationwide.
4	Used Chevrolet Malibu for Sale Near You | Edmunds	https://www.edmunds.com/used-chevrolet-malibu/	Save up to $9463 on one of 11598 used Chevrolet Malibus near you. Find your perfect car with Edmunds expert reviews, car comparisons, and pricing tools.
5	Used Chevrolet Malibu for Sale	https://www.carmax.com/cars/chevrolet/malibu	Used Chevrolet Malibu for Sale on carmax.com. Search new and used cars, research vehicle models, and compare cars, all online at carmax.com.
6	Used Chevrolet Malibus for Sale | TrueCar	https://www.truecar.com/used-cars-for-sale/listings/chevrolet/malibu/	No accidents, 1 Owner, Personal use Â· car image. 2018 Chevrolet Malibu. Premier with 2LZ. Great Price. $689 off avg. list price Â ...
7	Used Chevrolet Malibu For Sale - Carsforsale.comÂ®	https://www.carsforsale.com/chevrolet-malibu-for-sale-C999008	Results 1 - 15 of 13441 ... Find 13441 used Chevrolet Malibu as low as $1700 on Carsforsale.comÂ®. Shop millions of cars from over 21000 dealers and findÂ ...
8	50 Best Used Chevrolet Malibu for Sale, Savings from $2,839	https://www.autolist.com/chevrolet-malibu	Aug 27, 2019 ... Here are the top Chevrolet Malibu listings for sale ASAP. Check the carfax, find a low miles Malibu, view Malibu photos and interior/exteriorÂ ...
9	2019 Chevrolet Malibu for Sale - Autotrader	https://www.autotrader.com/cars-for-sale/2019/Chevrolet/Malibu	Find 2019 Chevrolet Malibu for Sale . Find car prices, photos, and more. Locate car dealers and find your car at Autotrader!
10	Used Chevrolet Malibu for Sale - Hertz Certified | Hertz Car Sales	https://www.hertzcarsales.com/used-chevrolet/malibu.htm	Shop a Hertz Certified Chevrolet Malibu online. We offer a quality selection, incredible prices, a warranty and a buy back guarantee. Schedule a test drive!
11	Used Chevrolet Malibu for Sale (with Photos) - CarGurus	https://www.cargurus.com/Cars/l-Used-Chevrolet-Malibu-d622	Listings 1 - 15 of 24608 ... Save $4486 on a used Chevrolet Malibu near you. Search over 36800 listings to find the best local deals. We analyze millions of usedÂ ...
12	Used Chevrolet Malibu for Sale (with Photos) - CARFAX	https://www.carfax.com/Used-Chevrolet-Malibu_w117	Every used car for sale comes with a free CARFAX Report. We have 9344 Chevrolet Malibu vehicles for sale that are reported accident free, 7718 1-Owner cars,Â ...
13	Used Chevrolet Malibu for Sale	https://www.carmax.com/cars/chevrolet/malibu	Used Chevrolet Malibu for Sale on carmax.com. Search new and used cars, research vehicle models, and compare cars, all online at carmax.com.
14	Used Chevrolet Malibu for Sale Near You | Edmunds	https://www.edmunds.com/used-chevrolet-malibu/	Save up to $9463 on one of 11598 used Chevrolet Malibus near you. Find your perfect car with Edmunds expert reviews, car comparisons, and pricing tools.
15	Used Chevrolet Malibu for Sale Near Me | Cars.com	https://www.cars.com/shopping/chevrolet-malibu/	Shop Chevrolet Malibu vehicles for sale at Cars.com. Research, compare and save listings, or contact sellers directly from 16947 Malibu models nationwide.

IV. METHODS

This Study uses several steps in processing the information from search engines, stages are as follows:

1) Web Scraping
2) Tokenization
3) Stop word
4) Stemming
5) LSI

These components will be built on Matlab.

The system will send a series of keywords to search engines. Then the search engine will reply with a list of websites. The web page will be stored temporarily by changing into a text file. This include the process of changing the words to lowercase, tokenization, stop word removal and stemming.

A. *Web Scraping*

Web scraping (or data scraping) is a technique used to collect content and data from the internet. This data is usually saved in a local file so that it can be manipulated and analyzed as needed. Common data types organizations collect include images, videos, text, product information, customer sentiments and reviews (on sites like Twitter, Yell, or Tripadvisor), and pricing from comparison websites.

Steps of Web Scraping

- Step 1 : Making an HTTP request to a server
- Step 2 : Extracting and parsing (or breaking down) the website’s code
- Step 3 : Saving the relevant data locally.

A brief explanation on the steps are given below:

*1) **Making an HTTP request to a Server**:* As an individual, when you visit a website via your browser, you send what’s called an HTTP request. This is basically the digital equivalent of knocking on the door, asking to come in. Once your request is approved, you can then access that site and all the information on it. Just like a person, a web scraper needs permission to access a site. Therefore, the first thing a web scraper does is send an HTTP request to the site they’re targeting.

*2) **Extracting and Parsing the website’s code**:* Once a website gives a scraper access, the application can read and extract the site’s HTML or XML code. This code determines the website’s content structure. The scraper will then parse the code (which basically means breaking it down into its constituent parts) so that it can identify and extract elements or objects that have been predefined by the owner of the application. These might include specific text, ratings, classes, tags, IDs, or other information.

*3) **Saving the data Locally**:* Once the HTML or XML has been accessed, scraped, and parsed, the web scraper will then store the relevant data locally. As mentioned, the data extracted is predefined by the owner (having told the application what they want it to collect). Data is usually stored as structured data, often in an Excel file, such as a .csv or .xls format.

B. *Tokenization*

Tokenization is a way of separating a piece of text into smaller units called tokens. Here, tokens can be either words, characters, or sub words. Hence, tokenization can be broadly classified into 3-types- word, character and sub word tokenization.Similarly, tokens can be either characters or sub words.

Example:

Consider the sentence ”Never give up”.

The most common way of forming tokens is based on space. Assuming space as a delimiter, the tokenization of the sentence results in 3 tokens- Never-give-up. As each token is a word it becomes an example of Word tokenization.

Similarly, tokens can be either characters or sub words. For example, consider ”smarter”:

1) Character tokens: s-m-a-r-t-e-r
2) Sub word tokens: smart-er

C. *Stopword*

Stop words are commonly used words that are excluded from searches to help index and parse web pages faster. While most internet search engines and Natural language processing(NLP) utilize stop words, they do not prevent a user from using them, but they are ignored.

Generally speaking, most stop words are function words, which are words with little or no meaning that help form a sentence. Content words like adjectives, nouns and verbs are often not considered as stop words. However a programmer may choose to add very common words.

Some common stop words include the following:

- a
- and
- but
- how
- or
- what
- will

D. *Stemming*

Stemming is the process of producing morphological variants of a root/base word. Stemming programs are commonly referred to as stemming algorithms or stemmers. A stemming algorithm reduces the words “chocolates”, “chocolatey”, “choco” to the root word, “chocolate” and “retrieval”, “retrieved”, “retrieves” reduce to the stem “retrieve”. Stemming is an important part of the pipelining process in Natural language processing. The input to the stemmer is tokenized words. How do we get these tokenized words? Tokenization involves breaking down the document into different words.

Errors in Stemming

There are mainly two errors in stemming –

- over-stemming
- under-stemming

Over-stemming occurs when two words are stemmed from the same root that are of different stems. Over-stemming can also be regarded as false-positives.

Under-stemming occurs when two words are stemmed from the same root that are not of different stems. Under-stemming can be interpreted as false-negatives.

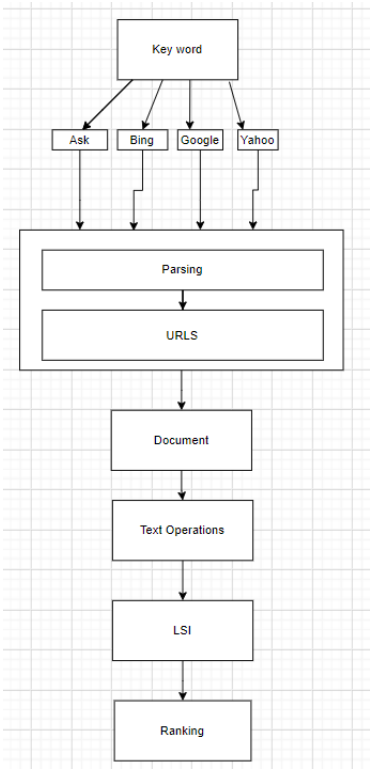
Stemming Algorithm Used

Porter Stemmer It is one of the most popular stemming methods proposed in 1980. It is based on the idea that the suffixes in the English language are made up of a combination of smaller and simpler suffixes. This stemmer is known for its speed and simplicity. The main applications of Porter Stemmer include data mining and Information retrieval. However, its applications are only limited to English words. Also, the group of stems is mapped on to the same stem and the output stem is not necessarily a meaningful word. The algorithms are fairly lengthy in nature and are known to be the oldest stemmer.

Example:

EED $\rightarrow$ EE : *means “if the word has at least one vowel and consonant plus EED ending, change the ending to EE” as ‘agreed’ becomes ‘agree’.*

The block diagram of search engine optimisation is:



V. Results and Discussion

The results that are shown below are, when we given in search bar to search for ”Chevrolet for sale” the below

images are the resulting matrices.

```
search1 =

    "Chevrolet for sale"


matrix =

    75     0
     1     1
    83    14
     0     0
    22     0
     4     1
    17    32
     1     0
     4     1
     3     5
```

Now, By performing the SVD of the count matrix.

```
U =

   -0.6375    0.3288   -0.6479         0   -0.2070   -0.0281    0.1443   -0.0094   -0.0281    0.0193
   -0.0097   -0.0269   -0.2369         0    0.0530   -0.0216   -0.9577    0.0024   -0.0216   -0.1488
   -0.7222   -0.0734    0.6647         0   -0.0779   -0.0171   -0.1547   -0.0035   -0.0171   -0.0254
         0         0         0    1.0000         0         0         0         0         0         0
   -0.1870    0.0965   -0.0626         0    0.9732   -0.0021    0.0671   -0.0012   -0.0021    0.0101
   -0.0352   -0.0137   -0.0185         0   -0.0032    0.9990   -0.0169   -0.0001   -0.0010   -0.0027
   -0.1827   -0.9250   -0.2754         0    0.0319   -0.0225    0.1200    0.0015   -0.0225   -0.1370
   -0.0085    0.0044   -0.0028         0   -0.0012   -0.0001    0.0030    0.9999   -0.0001    0.0005
   -0.0352   -0.0137   -0.0185         0   -0.0032   -0.0010   -0.0169   -0.0001    0.9990   -0.0027
   -0.0315   -0.1430   -0.0440         0    0.0046   -0.0035   -0.1365    0.0002   -0.0035    0.9788
```

```
S =

  116.5025         0
         0   31.7043
         0         0
         0         0
         0         0
         0         0
         0         0
         0         0
         0         0
         0         0
```

```
V =

   -0.9903    0.1390
   -0.1390   -0.9903
```

By using the reduction of dimensions of the matrix we will get the following matrices.

```
Uk =

   -0.6375    0.3288
   -0.0097   -0.0269
   -0.7222   -0.0734
         0         0
   -0.1870    0.0965
   -0.0352   -0.0137
   -0.1827   -0.9250
   -0.0085    0.0044
   -0.0352   -0.0137
   -0.0315   -0.1430

Sk =

  116.5025         0
         0   31.7043

Vk =

   -0.9903    0.1390
   -0.1390   -0.9903
```

The new matrix is obtained from the multiplication of the reduced matrices we get the following new matrix.

```
new =

   72.1017  -20.6482
    1.2367    0.6860
   83.6469   -9.3917
         0         0
   21.1498   -6.0568
    4.1207   -0.1399
   25.1530   26.0831
    0.9614   -0.2753
    4.1207   -0.1399
    4.2606    3.9808
```

The best 10 results obtained from the above operations for the given search results are

```
search1 =

    "Chevrolet for sale"

'Used Chevrolet Malibu for Sale Near Me | Cars.com'
'Used Chevrolet Malibu for Sale (with Photos) - CARFAX'
'Used Chevrolet Malibu For Sale - Carsforsale.comÂ®'
'Used Chevrolet Malibu For Sale - Carsforsale.comÂ®'
'Used Chevrolet Malibu for Sale'
'Used Chevrolet Malibu for Sale - Hertz Certified | Hertz Car Sales'
'Used Chevrolet Malibus for Sale | TrueCar'
'2019 Chevrolet Malibu for Sale - Autotrader'
'Used Chevrolet Malibu for Sale - Hertz Certified | Hertz Car Sales'
'Used Chevrolet Malibu for Sale (with Photos) - CarGurus'
```

The other search examples which will give the best 10 results are

```
search1 =

    "hyundai for sale"

'50 Best Used Hyundai Sonata for Sale, Savings from $2,725'
'Used Hyundai Santa Fe for Sale Near Me | Cars.com'
'Used Hyundai Elantra For Sale Online | Carvana'
'Used Hyundai Sonata For Sale Online | Carvana'
'Used Hyundai Elantra for Sale Near Me | Cars.com'
'Used Hyundai Elantra for Sale Near Me | Cars.com'
'Used Hyundai Santa Fe Sport For Sale Online | Carvana'
'Used Hyundai Santa Fe for Sale (with Photos) - CARFAX'
'Used Hyundai Santa Fe Sport for Sale (with Photos) - CarGurus'
'Used Hyundai Santa Fe Sport for Sale (with Photos) - CarGurus'
```

```
search1 =

    "nissan for sale"

'Used Nissan Sentra for Sale Near Me | Cars.com'
'Used Nissan Altima For Sale Online | Carvana'
'Used Nissan Altima For Sale Online | Carvana'
'Used Nissan Rogue For Sale Online | Carvana'
'Used Nissan Altima for Sale (with Photos) - CARFAX'
'Used Nissan Altima for Sale (with Photos) - CARFAX'
'Used Nissan Rogue for Sale (with Photos) - CARFAX'
'Used Nissan Rogue for Sale (with Photos) - CARFAX'
'Used 2009 Nissan Sentra for Sale & Salvage Auction Online - IAA'
'Used Hyundai Sonata For Sale - Carsforsale.comÂ®'
```

VI. Conclusions:

The results that can be drawn from this study are:

1) The results of search engines may change from server, but the search engines don't have the capability to renewing their index.
2) Changes in the content of the website can reduce the accuracy of the search engines.
3) Optimizing search engines give good results.
4) SVD is able to show the semantic relationships between keywords and documents.

References

[1] Nanda Zannibua Harisma and Anak Agung Putri Ratna, Implementation of automated essay grading system using LSA method with three keywords. Universitas Indonesia. Fakultas Teknik, 2008.

[2] Heryanto, Ahmad and Christyowidiasmoro, Christyowidiasmoro and Hariadi, Mochamad. (2014). Search Engine Optimization Based On Latent Semantic Indexing Using Web Scraping.

[3] Snasel, Vaclav and Abdulla, Hussam. (2009). Solution for Search Result Clustering by Using Singular Value Decomposition. 647 - 651. 10.1109/ICACTE.2008.143.

[4] S. K. Malik and S. A. M. Rizvi, "Information Extraction Using Web Usage Mining, Web Scrapping and Semantic Annotation," in 2011 International Conference on Computational Intelligence and Communication Networks (CICN), 2011, pp. 465–469.

[5] Scott Deerwester, Richard Harshman, Susan T, George W, and Thomas K, "Indexing by Latent Semantic Analysis," J. Am. Soc. Inf. Sci., vol. 41, no. 6, pp. 391–407, 1990.

[6] http://pages.cs.wisc.edu/ jerryzhu/cs769/latent.pdf