Estimation of the Number of Relevant Images in Infinite Databases

Xiaoling Wang

Abstract

In image retrieval the most commonly used performance measures are precision and recall. However, to determine the number of relevant images in an infinite database presents a significant challenge as the relevant parameters are not directly observable. In our research, we use internet as a vehicle to investigate this problem, and evaluate search results from major image search engines. We also investigate whether the cumulative relevance of images in different results pages follows particular stochastic behaviors, such as the Independent Model or the Markov Chain Distribution. From such models, we shall estimate the total number of relevant images for major image search engines.

1. Introduction and Related Works

Due to the increased importance of the Internet, the use of image search engines such as Google, Yahoo, and Msn is becoming increasingly widespread. However, for many image search engines, it is difficult for users to make a decision as to which image search engine should be selected. It is obvious that the more effective the system is, the more it will offer satisfaction to the user. Therefore, retrieval effectiveness becomes one of the most important parameters to measure the performance of image retrieval systems [18, 19, 20]. As we know, the most commonly used performance measures are the precision P and recall R[1, 2, 3, 13, 14], but to compute recall R is rather difficult as the total number of relevant images is not directly observable in such a potentially infinite database.

Many researchers have conducted studies to evaluate the retrieval effectiveness of image search engines. Ece Çakır et al. [4] described the retrieval effectiveness of image search engines based on various query topics. Fuat Uluç et al. [10] described the impact of the number of query words on image search engines. However, none of these studies describe how to estimate the total number of relevant images for the image search engines. All of them only view the first two page results. In the study by Sprink and Jansen [20], data collected from Dogpile was analyzed and one of the findings was that the percentages of the users that viewed only the first page and those that viewed only the first two pages of document search results were about 71.9% and 15.8%, respectively. Although many works used recall as the measure to evaluate the image search engine, not many papers work on the estimation of the number of relevant images in infinite databases. An algorithm called sample-resample is presented in by Si and Callan [6]; in environments containing resource descriptions already created by query-based sampling, the sample-resample method uses several additional queries to provide an estimate of the database size. Therefore, if the database size has been known, then the distribution of relevant images can be estimated.

In our paper, we model the probabilistic behavior of the distribution of relevant images among the returned results by evaluating the performance of some widespread image search engines.

2. Basic Models and Empirical Evaluation

2.1 Independent Distribution

The cumulative relevance of images in different results pages may follow the independent distribution. We let P_k denote the probability that the cumulative relevance of all the images in Page *k*. In general, it is normally true that, for search engines, the first pages provide a larger probability, so that

$P_1 \geq P_2 \geq \ldots \geq P_k \geq P_{k+1} \geq \ldots$

And in practice, we may take that after a certain number of retrieved results, the inequalities in the above will become strict ones. Since the relevant outcomes of different ranked images are not mutually exclusive events and that the search results do not feasibly terminate, we have in general $\sum_{k=1}^{\infty} P_k >> 1$ and that $P_k \to 0$, as $k \to \infty$. We shall model such probability sequence by independent distribution laws which conform to the above characterization and validate these experimentally. We shall also investigate the usefulness of the quadratic formula:

$P_k = \beta_1 k^2 + \beta_2 k + \alpha$,

where $k=1, 2, 3\ldots$

in representing results relevance over the different pages. Therefore, after determining the parameters using the least square method, we can use the non-linear model to estimate the number of relevant images page by page.

Therefore, from this model we can estimate the total number of relevant images for the major image search engines.

2.2 Markov Chain Model

Since in internet image search, results are returned in units of pages, we shall focus on the integer-valued stochastic process $X_1, X_2, \ldots$, where X_J represents the aggregate relevance of all the images in page J, which may be estimated by $X_J = \sum_{i \in J} Y_{Ji}$, where $Y_{Ji}=1$ if the ith image on page J is relevant, and $Y_{Ji}=0$ if the ith image on page J is not relevant. For this stochastic modeling of cumulative page image relevance, we shall investigate in particular the Markov Chain Model.

Here, we model the sequence $X=\{X_1, X_2, \ldots\}$ by a Markov Chain [11, 16, 17]. That is, we assume that the number of relevant images X_J in a page J only depends on X_{J-1} and not on the cumulative number of relevant images returned in $X_1, X_2, \ldots X_{J-2}$. We take the conditional probability of the number of relevant images in X_J given the number of relevant images in X_{J-1} to be the transition probability $p_{(J-1),J}$. From this, we construct the transition probability matrix.

$$P = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0n} \\ p_{10} & p_{11} & \cdots & p_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n0} & p_{n1} & p_{n2} & p_{nn} \end{pmatrix},$$

Where n is the number of images contained in a page. We can effectively estimate the initial probabilities if the sample is reasonably large. The probabilities are placed in a vector of state probabilities:

$\pi(J)$ = vector of state probabilities for page J

$= (\pi_0, \pi_1, \pi_2, \pi_3, \cdots, \pi_n)$

Where π_k is the probability of having k relevant images

Therefore, from this model, we can estimate the number of relevant images by pages by using the formula in page J:

$\pi(J) = \pi(J-1)*P, J=1, 2, 3, \ldots, n$

2.3 Queries Selection

We shall evaluate the top image search engines, namely, Google [7], Yahoo [8], and Msn [9], whose market shares are 64%, 16.3%, and 9.9%, respectively [10].By using 72 example queries. The queries consist of one-word, two-word and more than three-word queries, which range from simple words like apple to more specific query like apple computers and finally progressing to rather specific search. In the table below I only list part of queries.

Categories	**Sample Queries**
One-word query	Apple
	Dolphin
	Octopus
	Facebook
	Roxy
	Wildlife
	Skiing
	Alleyway
	Maldives
	Puppy
Two-word query	Apple Computer
	Plane Crash
	Octopus Card
	Outer Space

	Night Scene
	Daisy Flower
	Street-Art
	Baby Cry
	Afghan Child
	Twin Towers
More than Three-word query	Man Wearing Hat
	Macro Fly Eyes
	Sunrise and Sunset
	Jordan Basketball Nike
	Black and White Portrait
	HongKong Night Scene
	Flowing in the Wind
	Michael Schumacher Ferrari
	Chinese Opera Mask
	Victoria Harbour Hong Kong

Table 1. Part of Sample Query list

Categories	Test Queries
One-word query	Volcano
Two-word query	Tibetan Girl
Three-word query	Desert Camel Shadow

Table 2. Test Query list

2.4 Test Image Search Engines

The Test Image Search Engines we have selected are listed as follows.

Google(www.google.com)

Yahoo(www.yahoo.com)

Msn(www.msn.com)

Each query is run on the selected image search engines separately. Based on Spink and Jansen's study [12], evaluating the images in first two pages is enough. Such a finding seems useful for the users who only want to find less than forty images, but not so useful for users who need a lot of images. Therefore, we will model the independent model or the Markov Chain Model to help us to estimate the number of relevant images per page while we have the initial probability and transition probability matrix in a specific ISE, and it will also guide us as to when we should stop viewing.

3. Experimental Results

The search results for every query in Google, Yahoo and Msn are shown in Appendix 1, Appendix 2 and Appendix 3 respectively. In Appendix 1, the search results show that Google performs quite well except for the query "apple". The ISE Yahoo also shows us that it performs well except for the queries "apple", "roxy", "plane crash", "octopus card", and "outer space" in Appendix 2. Appendix 3 shows that the Msn Search Engine also performs well except for the queries "Apple", "octopus card", "outer space", "black and white portrait" and "HK night scene".

3.1 Independent Distribution

If the cumulative relevance of images in different results pages is independent, we try to create a formula to estimate the number relevant images in pages. In this experiment, we search 69 queries, and we totally review page by page to calculate the number of relevant images. The data are shown in Appendix 1. We can obtain a mean number of relevant images for each pages based on the data we collected. We plot the results for Google, Yahoo and Msn as follows.

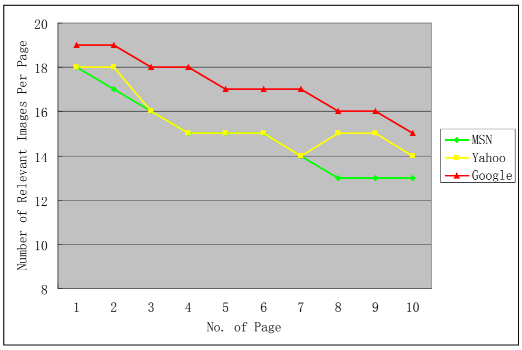

Figure 1. The Mean of the Number of Relevant Images for All the ISEs

The figure below shows the formula that we can use to estimate the number of relevant image page by page.

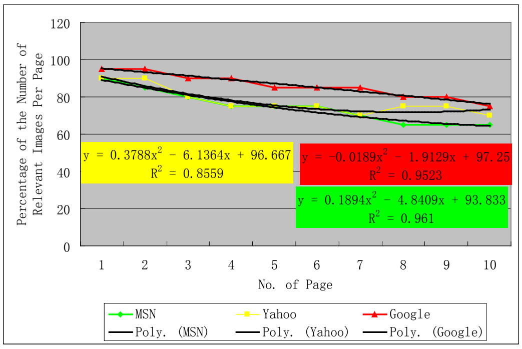

Figure 2. The Formula Created for All the ISEs

From the figure above, we can find that the formula for the image search engine Google is

$P_k = -0.0189 k^2 - 1.9129k + 97.25$ (1)

where k=1, 2, 3,…..

We can also get that the R-square is equal to *0.9523*, which represents the square of the sample correlation coefficient between the outcomes and their predicted values. And as we known, the larger the R^2 the better the model is.

Meanwhile, the formula for Yahoo and Msn are as follows respectively,

$P_k=0.3788 k^2- 6.1364k + 96.667$ (2)

$R^2 = 0.8559$

$P_k= 0.1894 k^2-4.8409k + 93.833$ (3)

$R^2 = 0.961$

Therefore, we can use these three formulas to estimate the number of relevant images page by page for the testing ISEs respectively. Meanwhile, according to R-square, the formula given by MSN is the most precise, following by Google and Yahoo provide.

Another fact that we can observe from the figure is the Google returned the best result, and it decline slowly. The second one is Yahoo, which decrease a little quicker than Google dose and the probability of the number of relevant images for each page are smaller than Google provides. Msn shows the worst result, it reduces more sharply and the probability of the number of relevant images per page is the smallest. Therefore, we can conclude that Google perform a best result if we think the cumulative relevance of images in different results pages is independent.

Now, we will testing these formulas by using the test query we have listed before. The charts are shown below and we will discuss the findings.

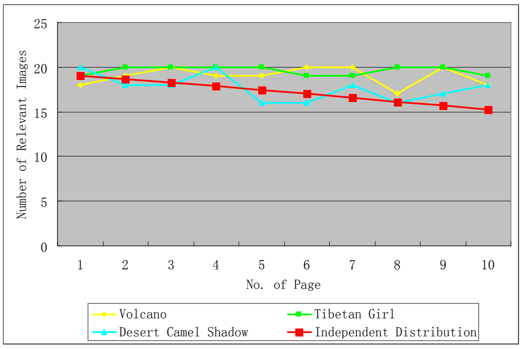

Figure 3. Search Result of Testing Queries and Independent Distribution Model for Google

Figure 3 shows us the searching result of testing queries and Independent Model we obtain from the example queries for Google. According to this figure, although the result we get from testing query is fluctuating while the independent model gives a gradual decline, we still can see that the Independent Model fits the testing queries quite well not only for the simple queries but especially for the rather specific searching term.

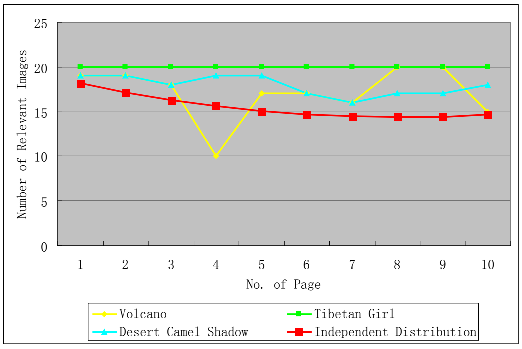

Figure 4. Search Result of Testing Queries and Independent Distribution Model for Yahoo

Figure 4 tells us that the searching result of testing queries and Independent Model we obtain from the example queries for Yahoo. From this figure, we can see that the independent model seems to provide a worse result than the actual data. However, in my opinion, we can still use this model to estimate the number of relevant images for engine, because apparently the distinction is not very large and we know the mean of the number is reasonable.

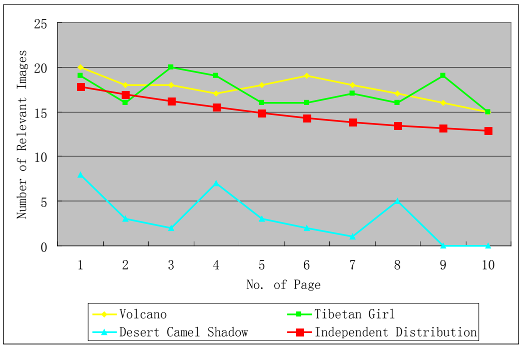

Figure 5. Search Result of Testing Queries and Independent Distribution Model for Msn

Figure 5 illustrates the searching result of testing queries and Independent Model we obtain from the example queries for Msn. The Figure shows us that the Independent Model seems fit the testing queries good too for one word query and two words query, and the trend seems quite the same although there are some fluctuation. But apparently, this model looks worse than we get from Google and Yahoo for rather specific searching term.

3.2 Markov Chain Model

According to these queries, we want to find whether the cumulative relevant images in different pages follows one-step Markov Chain Distribution. And we shall apply Markov Chain Model to estimate the number of relevant images in infinite databases.

The transition probability matrix and the vector of state probabilities for pages for all ISEs are shown in Appendix 1, Appendix 2 and Appendix 3 respectively.

In the following, the Markov Chain Model and test results will be shown. And we will discuss the findings.

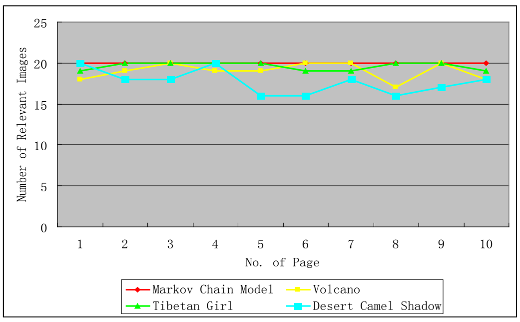

Figure 6. Search Result of Testing Queries and Markov Chain Model for Google

Figure 6 illustrates the searching result of testing queries and Markov Chain Model we obtain from the example queries for Google. According to this figure, we can see that the Markov Chain Model fits the testing queries rather well, no matter for the one-word query, two-word query or three-word query, although there are some distinctions. But apparently, we can't apply this model to query "apple" because this is a special case.

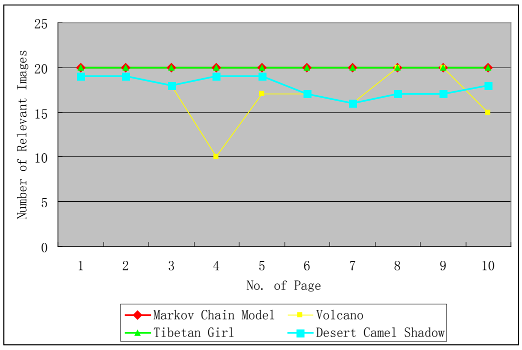

Figure 7. Search Result of Testing Queries and Markov Chain Model for Yahoo

Figure 7 illustrates the searching result of testing queries and Markov Chain Model we obtain from the example queries for Yahoo. The Figure shows us that the Markov Chain Model also fits the testing queries quite good. Especially for the two words query, the model fit the data perfectly, but for the single word query, some point is so bad apparently. But in whole, it is a good predicting model although it fits the data a little worse than Google does.

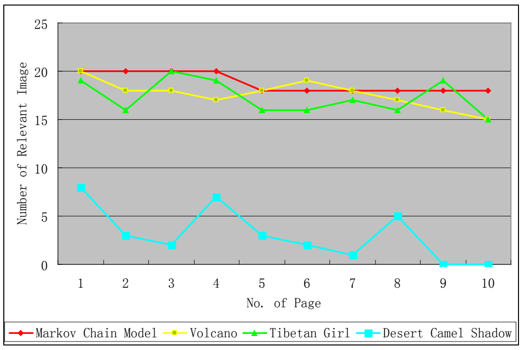

Figure 8. Search Result of Testing Queries and Markov Chain Model for Msn

Figure 8 illustrates the searching result of testing queries and Markov Chain Model we obtain from the example queries for Msn. The Figure tells us that the Markov Chain Model fits the testing queries well for the one-word query and two-word query but so bad for the three-word queries although the trend between them seems the same. Apparently, this model looks worse than we get from Google and Yahoo. And we can also get a conclusion that the Msn is not good at searching rather specific searching term.

Although all the equation can fit data quite well, but it can not fit all queries very well; the following is a special example. The Figure 9 shows us the searching result for query “apple” in all the ISEs we want to test.

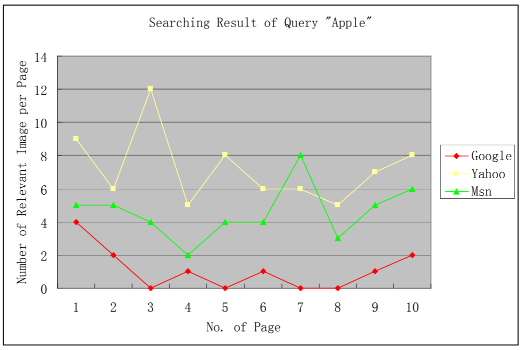

Figure 9. Search Result for Query “apple” in All the ISEs

We can see from the Fig 9 that the searching results are so bad in all ISE, although Yahoo perform better than Msn and Msn perform better than Google. For the query “apple”, it is equivocal for the ISE, because the ISE can’t differentiate what we want to search, the fruit or the brand of computer? Therefore, the searching results are so bad. But in all, for one-word query, we can see from the Appendix that the Google gives the best result and Msn displays the worst result.

For the two-word query, we choose query “apple computer” to compare the ISEs.

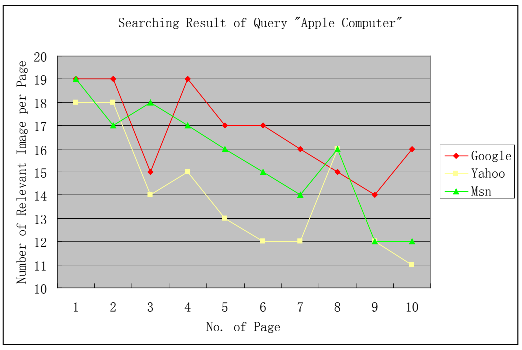

Figure 10. Search Result for Query “apple computer” in All the ISEs

From the Fig 10, we can see all the ISEs display good searching result when we give a more specific search term. And it shows us that Google give us the best result, Yahoo following it and Msn display the worst result.

For the three-word query, we pick query “black and white portrait” to compare.

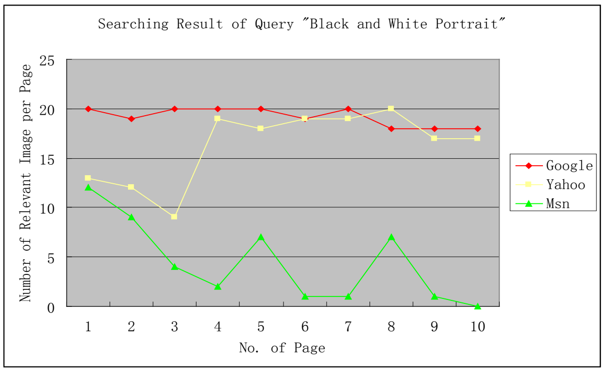

Figure 11. Search Result for Query “black and white portrait” in All the ISEs

When we give a rather specific query, Google perform rather well, the Yahoo displays well also, but Msn give us a quite bad result. Msn returned so many colorful images; it seems it can’t differentiate “black and white” and “colorful”. This illustrates Msn is not good at searching rather specific searching term but good at simple word.

4. Conclusion and Future Work

Currently, estimating the number of relevant images in the infinite image search engines is quite hard, but it is so important for us. Therefore, we develop a set of image queries to investigate models to estimate the number of relevant images in infinite ISEs. And using some queries to test the model we obtain.

If the stochastic process follows independent distribution, we can get equation (1) (2) (3) respectively for Google, Yahoo and Msn. We can observe that the Google returned the best results, and it decline slowly. Meanwhile it fits the data very well. The second one is Yahoo, which decrease a little quicker than Google dose and the probability of the number of relevant images for each page are smaller than Google provides. Msn shows the worst results when three-word query is given, and it reduces more sharply and the probability of the number of relevant images per page is the smallest. Therefore, we can conclude that Google performs the best results if we think the cumulative relevance of images in different results pages is independent. And the most important thing is that we can use these formulas to estimate the number of relevant images for ISEs.

If the stochastic process follows Markov Chain Distribution, we can find that the Markov Chain Model we obtain for Google is the best, because it fits the data quite well, especially for the quite specific queries. Yahoo comes in 2nd. The model fits the specific query quite well, but for the special specific query, the result is a little worse than Google have done. Msn seems the worst among these three ISEs when a rather specific query is given. But it fits easy word and specific query quite good. In a word, according to our experiment, we can use Independent Model and Markov Chain Model to estimate the total number of relevant images in infinite image search engines. But apparently, the equivocal character of the query "apple" made all the Independent Model and Markov Chain Model incapable of estimating.

For infinite image repositories, it is generally impractical to step through the interminable set of search results presented. Therefore, our future work is to find an optimal stopping rule. According to what we have done above, it seems using Markov Chain could provide us a good stopping rule.

5. Reference

[1] K. Stevenson and C. Leung, "Comparative Evaluation of Web Image Search Engines for Multimedia Applications", *In Proceedings of IEEE International Conference on Multimedia and Expo*, July 2005

[2] Ishioka, T. 2003, "Evaluation of Criteria for Information Retrieval", *In Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence* (October 13 - 17, 2003). Web Intelligence IEEE Computer Society, Washington, DC, 425

[3]Raquel Kolitski Stasiu, Carlos A. Heuser, and Roberto da Silva, "Estimating Recall and Precision for Vague Queries in Databases", *In Proceedings of Springer-Verlag Berlin Heidelberg, CAiSE 2005, LNCS 3520, pp. 187-200, 2005*

[4] Ece Çakır, Hüseyin Bahçeci, Yıltan Bitirim, "An Evaluation of Major Image Search Engines on Various Query Topics," icimp, pp.161-165, 2008. *In Proceedings of the Third International Conference on Internet Monitoring and Protection*, 2008

[5] F. Uluç, E. Emirzade and Y. Bitirim, "The Impact of Number of Query Words on Image Search Engines", *In Proceedings of Second International Conference on Internet and Web Applications and Services* (ICIW), pp. 50, 2007

[6] Si, L. and Callan, J. 2003, "Relevant document distribution estimation method for resource selection", *In Proceedings of the 26th Annual international ACM SIGIR Conference on Research and Development in informaion Retrieval* Toronto, Canada, July 28 - August 01, 2003. SIGIR '03. ACM, New York, NY, 298-305

[7] Google Search Engine www.google.com

[8] Yahoo Search Engine www.yahoo.com

[9] Msn Search Engine www.msn.com

[10] Nielsen//NetRatings, "Goole accounts for more than half of all web searches, while Yahoo and Msn combined account more than one fourth, according to Nielson//Netrating", April 2009. http://www.polepositionmarketing.com/emp/april-2009-search-engine-2/

[11] Render, R., R. M. Stair Jr., and M. E. Hanna. "Quantitative Analysis for Management-Chapter 16", 10th Edition, Prentice Hall, 2009

[12] A.Sprink and B.J.Jansen, "Searching Multimedia Federated Content Web Collections", Online Information Review, Vol.30, No.5, pp. 485-495,2006

[13] Jin, X. and French, J. C. 2003. Improving ima ge retrieval effectiveness via multiple queries. *In Pr oceedings of the 1st ACM international Workshop o n Multimedia Databases* (New Orleans, LA, USA, November 07 - 07, 2003). MMDB '03. ACM, New York, NY, 86-93. DOI= http://doi.acm.org/10.1145/951676.951692

[14] Nottelmann, H. and Fuhr, N. 2003. Evaluating different methods of estimating retrieval quality for resource selection. In *Proceedings of the 26th Annual international ACM SIGIR Conference on Research and Development in informaion Retrieval* (Toronto, Canada, July 28 - August 01, 2003). SIGIR '03. ACM, New York, NY, 290-297. DOI=

http://doi.acm.org/10.1145/860435.860489

[15] Thomas W. Miller, *Data and Text Mining: A Business Applications Approach.* Upper Saddle River, N.J. : Pearson Prentice Hall, c2005

[16]W.S. Kendall, F. Liang, J.-S. Wang, *Markov chain Monte Carlo : innovations and applications,* Singapore ; Hackensack, NJ : World Scientific, c2005

[17]Dani Gamerman, *Markov chain Monte Carlo : stochastic simulation for Bayesian inference***,** London ; New York : Chapman & Hall, 1997

[18] Shirahatti, N. V. and Barnard, K. 2005. Evaluating Image Retrieval. *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cvpr'05) - Volume 1 - Volume 01* (June 20 - 26, 2005). CVPR. IEEE Computer Society, Washington, DC, 955-961. DOI= http://dx.doi.org/10.1109/CVPR.2005.147

[19] Black Jr, J. A., Fahmy, G., and Panchanathan, S. 2002. A Method for Evaluating the Performance of Content-Based Image Retrieval Systems. *In Proceedings of the Fifth IEEE Southwest Symposium on Image Analysis and interpretation* (April 07 - 09, 2002). SSIAI. IEEE Computer Society, Washington, DC, 96.

[20] Münevver Tugçe Elagöz , Mehtap Mendeli , Remziye Zeden Manioglulari , Yiltan Bitirim, "An Empirical Evaluation on Meta-Image Search Engines," ICDT, pp.135-139, 2008. *In Proceedings of the Third International Conference on Digital Telecommunications* (ICDT 2008), 2008

Appendix 1: The Searching Result for All Queries in Google

Page Number \ Query Number of Relevant Images	Apple	Dolphin	Octopus	Facebook	Roxy	Wildlife
1	4	20	18	20	20	18
2	2	20	20	20	19	17
3	0	20	19	20	18	18
4	1	20	17	20	18	16
5	0	19	17	20	17	13
6	1	19	18	20	17	15
7	0	19	15	20	16	10
8	0	19	14	19	15	5
9	1	18	14	19	12	11
10	2	18	13	19	12	11
	Apple Computer	Plane Crash	Octopus Card	Outer Space	Night Scene	Daisy Flower
1	19	19	20	20	20	20
2	19	18	18	17	20	18
3	15	18	16	16	20	18
4	19	20	17	18	20	18
5	17	19	15	14	18	16
6	17	18	13	13	17	17
7	16	18	15	15	18	17
8	15	17	10	10	15	17
9	14	18	7	9	14	15
10	16	17	4	3	15	12
	Man Wearing	Macro Fly Eyes	Sunrise and Sunset	Jordan Basketball	Black and White	HK Night Scene

	Hat			Nike	Portrait	
1	20	18	19	20	20	19
2	19	18	20	19	19	19
3	18	16	20	19	20	20
4	18	16	20	20	20	19
5	19	16	19	18	20	20
6	16	16	20	17	19	19
7	17	15	16	19	20	20
8	18	15	17	20	18	18
9	17	10	15	20	18	19
10	20	11	15	19	18	18

Query / Number of Relevant Images / Page Number	Starbucks	Skiing	Alleyway	Maldives	Puppy	Twilight
1	20	19	20	19	20	20
2	20	19	20	19	20	20
3	20	20	20	18	20	20
4	20	19	18	18	20	20
5	19	17	20	18	19	20
6	20	16	19	20	19	20
7	18	17	20	20	19	20
8	18	18	19	17	20	18
9	16	14	19	18	19	17
10	18	13	17	13	18	19
	Macro Abstract	Street-art	Baby Cry	Afghan Child	Twin Towers	Toilet Icon
1	19	19	17	20	19	14
2	18	20	19	20	19	8
3	18	20	16	19	18	14
4	17	20	13	20	19	9
5	16	19	10	20	17	13
6	18	20	10	20	18	5
7	17	19	4	17	19	9
8	18	17	8	19	18	3
9	14	16	4	17	19	4
10	18	17	6	19	18	2
	Eagle Catching Fish	Blowing in the Wind	Michael Schumacher Ferrari	Chinese Opera Mask	Victoria Harbour Hong	Star Wars and Pepsi

					Kong	
1	19	20	19	19	20	19
2	18	20	19	20	19	19
3	17	18	19	19	18	18
4	19	18	18	18	19	18
5	18	18	20	19	18	17
6	20	19	19	20	18	17
7	17	19	19	19	20	17
8	18	20	18	18	19	19
9	18	18	18	17	18	15
10	18	19	18	16	19	17

Query / Number of Relevant Images / Page Number	Headphone	Gundam	Yelling	Transformers	Chair	Scrat
1	19	20	20	20	19	20
2	20	20	20	20	20	19
3	19	20	19	20	20	20
4	18	20	19	20	20	20
5	19	20	18	20	19	20
6	20	20	18	20	20	20
7	20	20	16	20	18	19
8	18	19	15	19	20	20
9	17	19	14	16	20	19
10	20	19	14	18	19	20
	Kiehl's Lotion	Colorful Candy	Jennifer Aniston	Doraemon Figure	Yoga Poses	Amazon Rainforest
1	19	19	20	19	20	20
2	19	18	20	20	20	20
3	19	16	20	19	20	20
4	18	16	20	19	19	19
5	19	15	19	19	17	19
6	19	17	20	18	16	20
7	18	16	20	19	18	19
8	19	14	20	14	16	19
9	19	12	20	18	16	18
10	15	13	20	15	14	18
	Great White Shark	Rock n Roll	Heart Shaped Cookies	Drink Vending Machine	Butterfly on Yellow Flower	Lily of the Valley

1	20	20	19	20	20	16
2	20	20	18	20	19	18
3	20	20	16	19	19	17
4	20	20	16	19	15	18
5	20	20	16	16	14	16
6	19	20	18	16	13	16
7	20	20	16	15	13	15
8	18	20	17	18	11	16
9	20	20	16	16	9	15
10	20	18	16	15	16	15

Query / Number of Relevant Images / Page Number	Tornado	Tired	Fisheye	Windmill	Steak	Piano
1	19	18	20	20	18	14
2	18	18	20	20	20	12
3	20	14	20	18	18	9
4	17	16	20	18	18	11
5	17	15	20	18	18	3
6	19	15	20	17	19	4
7	11	12	20	17	18	6
8	15	14	20	17	20	3
9	15	11	20	15	18	4
10	8	9	20	15	18	2
	Outdoor Wedding	Dior Catwalk	BMW Z4	African Art	Solar Eclipse	Icy Tree
1	20	20	20	20	18	20
2	20	20	20	20	17	19
3	19	20	20	20	19	20
4	20	20	20	20	19	20
5	20	20	20	19	20	19
6	20	19	20	19	18	17
7	20	19	20	19	20	20
8	19	17	20	20	20	20
9	19	16	20	19	20	20
10	20	16	20	20	20	20
	Bee on Sunflower	Sailing on San Francisco	Front View of Garden Cottage	Couple in Beach Chairs	Pink Room Design	Messy Working Desk

		Bay				
1	20	19	17	16	20	16
2	20	19	16	15	19	13
3	20	17	14	15	19	14
4	17	19	15	17	14	12
5	16	17	14	17	15	15
6	18	16	17	15	15	15
7	16	15	13	12	15	14
8	16	17	12	10	14	13
9	18	19	14	11	11	13
10	18	20	15	7	10	14

Appendix 1: Google Transition Probability Matrix

Number of Relevant Images	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.5	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0.6667	0	0.3333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0.5	0	0	0	0.3333	0	0.1667	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
6	0	0	0	0.5	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0.3333	0	0	0	0	0.3333	0	0	0	0	0.3333	0	0	0	0	0	0
9	0	0	0	0.4	0	0	0	0	0	0	0	0.2	0	0.2	0	0	0.2	0	0	0	0
10	0	0	0	0	0.1429	0.1429	0	0.1429	0	0.1429	0.1429	0.2857	0	0	0	0	0	0	0	0	0
11	0	0	0	0.1429	0	0	0	0.1429	0	0.1429	0.1429	0.2857	0	0	0	0.1429	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0.1429	0.1429	0	0.1429	0.1429	0.2857	0.1429	0	0	0	0	0
13	0	0	0	0	0	0	0.0909	0	0	0	0.0909	0.0909	0.0909	0.1818	0.1818	0.2727	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0.04	0.04	0	0.04	0.12	0.2	0.08	0.24	0.08	0.08	0.08	0	0
15	0	0	0	0	0	0	0	0	0.0294	0	0.1176	0	0.1176	0.0294	0.2353	0.2941	0.0294	0.0882	0.0294	0.0294	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0682	0.0682	0.2273	0.2727	0.1591	0.2045	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0175	0	0.0877	0.2456	0.193	0.2105	0.193	0.0526
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0105	0.0421	0.0316	0.1368	0.1789	0.3053	0.1579	0.1368
19	0	0	0	0	0	0	0	0	0	0	0	0.0074	0	0	0.0148	0.0296	0.0296	0.0889	0.2593	0.2963	0.2741
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0051	0.0303	0.0909	0.2626	0.6111

Appendix 1: The table below shows us that the number of relevant images for Sixteen Pages. The result shows us the probability and index, which get from the program we have done with Matlab.

Page Number	Probability	Index	Number of Relevant Images
1	0.5139	21.0000	20
2	0.4068	21.0000	20
3	0.3406	21.0000	20
4	0.2965	21.0000	20
5	0.2649	21.0000	20
6	0.2410	21.0000	20
7	0.2222	21.0000	20
8	0.2067	21.0000	20
9	0.1938	21.0000	20
10	0.1828	21.0000	20
11	0.1732	21.0000	20
12	0.1648	21.0000	20
13	0.1573	21.0000	20
14	0.1506	21.0000	20
15	0.1445	21.0000	20
16	0.1389	21.0000	20

The original vector is as follows:

π (0)=[0, 0, 0, 0, 1/72, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2/72, 0, 3/72, 2/72, 6/72, 21/72, 37/72]

The testing result for Google:

Model and Testing Query / Page Number	Markov Chain Model	Volcano	Tibetan Girl	Desert Camel Shadow
1	20	18	19	20
2	20	19	20	18
3	20	20	20	18
4	20	19	20	20
5	20	19	20	16
6	20	20	19	16
7	20	20	19	18
8	20	17	20	16
9	20	20	20	17
10	20	18	19	18

Appendix 2: The Searching Result for All Queries in Yahoo

Query / Number of Relevant Images / Page Number	Apple	Dolphin	Octopus	Facebook	Roxy	Wildlife
1	9	19	19	19	20	15
2	6	19	19	17	20	16
3	12	20	18	9	15	17
4	5	15	13	5	5	13
5	8	15	16	6	10	13
6	6	13	11	7	12	14
7	6	16	14	5	14	15
8	5	16	9	5	6	12
9	7	11	8	6	9	13
10	8	16	14	4	4	10
	Apple Computer	Plane Crash	Octopus Card	Outer Space	Night Scene	Daisy Flower
1	18	18	13	19	20	19
2	18	18	3	17	19	20
3	14	16	8	4	19	18
4	15	4	2	1	20	19
5	13	11	5	1	18	19
6	12	9	8	3	19	16
7	12	8	7	0	20	18
8	16	7	6	1	19	18
9	12	7	4	0	19	17
10	11	7	8	0	18	15
	Man Wearing Hat	Macro Fly Eyes	Sunrise and Sunset	Jordan Basketball Nike	Black and White Portrait	HK Night Scene
1	19	20	18	19	13	20
2	17	19	16	20	12	20
3	15	19	17	20	9	20
4	17	19	15	20	19	20
5	18	20	19	20	18	19
6	18	20	19	20	19	18
7	19	18	18	18	19	20
8	17	19	19	20	20	20
9	18	19	19	19	17	19
10	18	19	19	19	17	19

Query / Number of Relevant Images / Page Number	Starbucks	Skiing	Alleyway	Maldives	Puppy	Twilight
1	19	19	19	19	20	20
2	17	12	18	19	20	20
3	19	11	17	19	20	19
4	13	9	17	19	19	6
5	17	5	19	19	18	3
6	18	10	17	20	17	5
7	16	9	20	20	18	5
8	18	9	18	19	15	2
9	19	12	17	17	20	5
10	16	10	17	19	19	5
	Macro Abstract	Street-art	Baby Cry	Afghan Child	Twin Towers	Toilet Icon
1	20	19	5	17	19	20
2	20	20	5	11	20	20
3	20	19	8	17	19	20
4	18	20	8	19	19	20
5	19	19	8	12	14	20
6	19	20	10	18	17	20
7	18	19	8	17	13	20
8	19	20	11	16	12	20
9	19	20	6	16	12	20
10	17	20	10	17	13	20
	Eagle Catching Fish	Blowing in the Wind	Michael Schumacher Ferrari	Chinese Opera Mask	Victoria Harbour Hong Kong	Star Wars and Pepsi
1	19	19	20	20	20	20
2	18	19	19	19	20	17
3	19	20	20	20	20	17
4	18	20	20	20	20	16
5	16	20	20	20	20	16
6	15	19	20	19	20	15
7	15	20	20	20	20	9
8	14	20	20	18	20	0
9	13	20	20	20	20	0
10	14	19	19	19	19	0

Query / Number of Relevant Images / Page Number	Headphone	Gundam	Yelling	Transformers	Chair	Scrat
1	19	9	18	20	20	20
2	12	17	18	20	20	20
3	7	8	17	18	20	14
4	6	0	17	18	20	12
5	7	1	16	18	19	7
6	7	4	14	18	18	6
7	9	8	13	17	19	8
8	9	15	13	18	19	8
9	12	20	6	20	19	10
10	11	17	7	18	20	5
	Kiehl's Lotion	Colorful Candy	Jennifer Aniston	Doraemon Figure	Yoga Poses	Amazon Rainforest
1	6	17	20	19	20	20
2	19	20	20	20	16	17
3	19	19	20	18	14	20
4	18	16	20	13	20	20
5	19	19	18	11	19	17
6	12	19	20	8	17	15
7	14	20	20	7	18	15
8	13	18	20	0	17	17
9	13	17	20	0	18	17
10	14	17	20	0	17	20
	Great White Shark	Rock n Roll	Heart Shaped Cookies	Drink Vending Machine	Butterfly on Yellow Flower	Lily of the Valley
1	20	20	20	19	20	16
2	20	19	20	20	17	16
3	20	14	20	19	15	17
4	19	18	20	19	11	20
5	19	20	19	19	13	17
6	20	20	19	18	14	15
7	18	7	18	19	12	11
8	18	10	16	18	15	16
9	19	18	16	20	13	20
10	19	20	17	19	11	18

Query / Number of Relevant Images / Page Number	Tornado	Tired	Fisheye	Windmill	Steak	Piano
1	18	18	20	20	20	17
2	19	17	20	18	16	18
3	11	17	20	18	17	16
4	8	17	20	16	18	13
5	8	16	20	19	18	15
6	11	17	20	18	17	11
7	6	16	20	16	17	9
8	5	18	20	16	17	13
9	7	16	20	18	17	11
10	6	18	20	16	16	12
	Outdoor Wedding	Dior Catwalk	BMW Z4	African Art	Solar Eclipse	Icy Tree
1	19	20	20	20	17	20
2	19	19	20	20	20	20
3	19	20	20	18	17	20
4	20	20	20	18	11	20
5	19	19	20	20	17	19
6	18	12	20	20	20	20
7	18	17	20	19	16	20
8	17	19	20	20	17	20
9	19	19	19	20	17	20
10	19	19	19	20	14	20
	Bee on Sunflower	Sailing on San Francisco Bay	Front View of Garden Cottage	Couple in Beach Chairs	Pink Room Design	Messy Working Desk
1	20	19	11	11	16	18
2	18	19	15	9	17	16
3	18	19	12	3	12	16
4	18	18	12	5	11	15
5	20	19	11	4	10	16
6	20	20	12	5	4	16
7	20	19	13	0	6	15
8	18	19	12	2	14	15
9	19	18	8	2	6	14
10	20	17	9	0	11	13

Appendix 2: Yahoo Transition Probability Matrix

Number of Relevant Images	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.625	0.25	0.125	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0.25	0.25	0	0.25	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0.25	0	0.25	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0.25	0	0	0	0	0.5	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0.1667	0	0	0	0.1667	0.1667	0	0.3333	0	0	0.1667	0	0	0	0	0	0	0	0	0
5	0.0625	0	0.0625	0	0.0625	0.25	0.125	0.125	0.1875	0	0.125	0	0	0	0	0	0	0	0	0	0
6	0	0	0.0625	0.0625	0.125	0.125	0.0625	0.1875	0.0625	0.0625	0.0625	0.0625	0	0	0.0625	0	0	0	0	0.0625	0
7	0.0833	0	0	0	0	0.0833	0.3333	0.25	0.0833	0.0833	0.0833	0	0	0	0	0	0	0	0	0	0
8	0.0625	0	0.0625	0	0	0	0.0625	0.1875	0.25	0	0.125	0.125	0	0	0.0625	0.0625	0	0	0	0	0
9	0.0667	0	0	0.0667	0.0667	0.1333	0.0667	0	0.1333	0.1333	0	0	0.1333	0.0667	0	0	0	0.0667	0	0.0667	0
10	0	0	0	0	0.1667	0.1667	0	0	0.1667	0.1667	0	0	0.1667	0	0	0	0	0	0.1667	0	0
11	0	0	0	0	0	0	0.1111	0	0.1111	0.2222	0.0556	0	0.1111	0.0556	0.0556	0.0556	0.1111	0.1111	0	0	0
12	0	0	0	0	0	0.04	0.04	0.08	0.04	0.08	0.04	0.2	0.12	0.12	0.08	0.04	0.04	0.04	0.04	0	0
13	0	0	0	0.05	0	0	0	0	0	0	0.05	0.15	0.2	0.15	0.2	0.05	0.1	0.05	0	0	0
14	0	0	0	0	0	0	0.1538	0	0	0.0769	0	0	0.1538	0.2308	0	0.1538	0	0.0769	0.0769	0	0.0769
15	0	0	0	0	0	0.0435	0	0	0	0.0435	0	0.1304	0.087	0.1304	0.087	0.1739	0.087	0.087	0	0.0435	0.087
16	0	0	0	0	0.0294	0	0	0	0	0	0	0.0588	0.0294	0.0294	0.0588	0.1176	0.2647	0.2059	0.1176	0.0588	0.0294
17	0	0	0	0	0.0175	0	0	0	0.0175	0.0175	0	0.0351	0.0175	0.0351	0.0175	0.1053	0.1053	0.2456	0.1404	0.1053	0.1404
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0274	0.0137	0.0137	0.1507	0.1918	0.2466	0.2329	0.1233
19	0	0	0	0	0	0	0.0088	0	0	0	0	0.0088	0.0439	0.0088	0.0175	0	0.0263	0.0877	0.1754	0.3596	0.2632
20	0	0	0	0	0	0	0	0.006	0	0	0	0	0	0	0.006	0.0119	0.0179	0.0536	0.1131	0.25	0.5417

Appendix 2: The table below shows us that the number of relevant images for Sixteen Pages. The result shows us the probability and index, which get from the program we have done with Matlab.

Page Number	Probability	Index	Number of Relevant Images
1	0.4028	21.0000	20
2	0.3282	21.0000	20
3	0.2809	21.0000	20
4	0.2498	21.0000	20
5	0.2278	21.0000	20
6	0.2112	21.0000	20
7	0.1980	21.0000	20
8	0.1872	21.0000	20
9	0.1781	21.0000	20
10	0.1703	21.0000	20
11	0.1636	21.0000	20
12	0.1577	21.0000	20
13	0.1525	21.0000	20
14	0.1479	21.0000	20
15	0.1438	21.0000	20
16	0.1402	21.0000	20

The original vector:

π (0)= [0,0,0,0,0,1/72,1/72,0,0,2/72,0,2/72,0,2/72,0,1/72,2/72,4/72,7/72,21/72,29/72]

The Testing Result for Yahoo:

Model and Testing Query / Page Number	Markov Chain Model	Volcano	Tibetan Girl	Desert Camel Shadow
1	20	19	20	19
2	20	19	20	19
3	20	18	20	18
4	20	10	20	19
5	20	17	20	19
6	20	17	20	17
7	20	16	20	16
8	20	20	20	17
9	20	20	20	17
10	20	15	20	18

Appendix 3: The Searching Result for All Queries in Msn

Query / Number of Relevant Images / Page Number	Apple	Dolphin	Octopus	Facebook	Roxy	Wildlife
1	5	20	19	18	20	20
2	5	17	19	17	18	19
3	4	18	18	18	18	19
4	2	16	17	17	17	18
5	4	14	18	17	18	16
6	4	11	16	15	17	15
7	8	10	16	18	16	13
8	3	8	15	18	14	15
9	5	7	15	18	14	15
10	6	11	13	17	14	14
	Apple Computer	Plane Crash	Octopus Card	Outer Space	Night Scene	Daisy Flower
1	19	17	18	16	19	17
2	17	17	14	13	16	16
3	18	17	6	11	17	15
4	17	16	11	8	16	16
5	16	16	9	11	15	13
6	15	15	6	5	16	13
7	14	17	8	4	16	12
8	16	12	8	1	12	10
9	12	18	6	7	14	9
10	12	16	5	7	15	9
	Man Wearing Hat	Macro Fly Eyes	Sunrise and Sunset	Jordan Basketball Nike	Black and White Portrait	HK Night Scene
1	18	15	19	12	18	18
2	17	12	19	9	14	17
3	16	14	16	4	13	16
4	14	13	19	2	10	14
5	15	16	16	7	4	15
6	15	15	16	1	7	15
7	13	11	17	1	7	13
8	11	14	17	7	0	11
9	6	11	19	1	0	6
10	10	12	20	0	0	10

Query / Number of Relevant Images / Page Number	Starbucks	Skiing	Alleyway	Maldives	Puppy	Twilight
1	20	19	19	20	18	19
2	17	20	17	20	16	18
3	17	20	16	19	14	18
4	16	19	16	20	15	15
5	17	19	13	15	13	17
6	16	19	14	17	17	19
7	14	19	14	18	13	14
8	17	15	12	17	14	13
9	16	18	14	17	12	12
10	16	19	11	16	15	14
	Macro Abstract	Street-art	Baby Cry	Afghan Child	Twin Towers	Toilet Icon
1	18	20	7	19	20	0
2	18	20	5	17	17	1
3	16	19	3	15	18	1
4	16	17	3	15	17	0
5	15	16	3	11	15	2
6	14	16	2	10	14	1
7	16	15	1	10	16	0
8	18	15	2	10	14	3
9	15	12	1	10	10	1
10	16	16	2	6	16	1
	Eagle Catching Fish	Blowing in the Wind	Michael Schumacher Ferrari	Chinese Opera Mask	Victoria Harbour Hong Kong	Star Wars and Pepsi
1	20	18	20	20	13	20
2	18	17	20	17	12	18
3	16	16	18	18	11	18
4	15	12	20	14	9	19
5	12	8	20	13	12	18
6	11	13	20	7	17	18
7	10	10	20	10	15	16
8	8	9	20	11	13	14
9	7	8	18	7	9	12
10	9	7	20	10	17	12

Query / Number of Relevant Image / Page Number	Headphone	Gundam	Yelling	Transformers	Chair	Scrat
1	20	20	19	17	20	20
2	18	20	19	19	20	19
3	20	20	18	20	18	19
4	20	19	15	19	20	19
5	18	20	17	19	20	18
6	20	20	17	20	20	19
7	20	20	16	20	20	19
8	17	20	16	20	19	17
9	16	20	17	20	20	17
10	17	20	17	20	19	15
	Kiehl's Lotion	Colorful Candy	Jennifer Aniston	Doraemon Figure	Yoga Poses	Amazon Rainforest
1	19	15	19	19	20	20
2	19	15	20	17	20	20
3	18	12	19	16	19	20
4	17	8	20	14	17	19
5	18	8	20	11	18	18
6	16	9	20	13	19	20
7	15	6	20	5	15	20
8	13	6	20	5	15	18
9	13	7	20	2	18	18
10	12	8	20	2	15	19
	Great White Shark	Rock n Roll	Heart Shaped Cookies	Drink Vending Machine	Butterfly on Yellow Flower	Lily of the Valley
1	20	20	17	19	15	13
2	19	20	17	19	8	10
3	19	20	17	15	6	13
4	18	20	15	20	6	14
5	16	20	14	19	6	9
6	18	20	16	16	6	11
7	17	20	15	17	6	8
8	18	18	16	18	4	12
9	17	20	15	17	2	10
10	18	19	9	16	4	6

Page Number \ Number of Relevant Images \ Query	Tornado	Tired	Fisheye	Windmill	Steak	Piano
1	18	19	20	18	19	17
2	18	17	20	18	20	17
3	14	17	19	18	20	17
4	12	16	19	18	16	17
5	13	17	19	18	19	14
6	13	17	20	18	18	17
7	10	17	19	16	18	17
8	13	15	17	16	17	15
9	13	16	18	16	18	17
10	10	15	18	20	16	15
	Outdoor Wedding	Dior Catwalk	BMW Z4	African Art	Solar Eclipse	Icy Tree
1	19	17	20	20	16	19
2	18	19	20	20	12	18
3	19	16	20	19	15	16
4	20	18	20	19	14	14
5	18	19	20	20	13	17
6	18	19	20	19	10	16
7	18	18	20	20	11	18
8	15	15	19	17	8	17
9	17	17	20	19	12	17
10	13	14	20	20	11	12
	Bee on Sunflower	Sailing on San Francisco Bay	Front View of Garden Cottage	Couple in Beach Chairs	Pink Room Design	Messy Working Desk
1	20	17	16	9	11	20
2	19	15	16	6	10	18
3	20	16	15	5	8	17
4	17	17	14	3	9	10
5	16	19	16	4	3	13
6	17	17	15	4	4	12
7	16	17	12	2	2	9
8	18	18	14	1	4	8
9	16	15	14	1	6	8
10	13	17	12	2	3	5

Appendix 3: Msn Transition Probability Matrix

Number of Relevant Images	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.4	0.2	0.2	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0.25	0.3333	0.25	0	0	0	0	0.1667	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0.4444	0.1111	0	0.3333	0	0	0.1111	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0.1429	0.1429	0.2857	0.2857	0.1429	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0.0833	0.4167	0	0.1667	0	0.1667	0.0833	0.0833	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0.1429	0.2857	0.2857	0.2857	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
7	0.0909	0.1818	0	0	0	0.0909	0	0.1818	0.0909	0.0909	0.1818	0.0909	0	0	0	0	0	0	0	0	0
8	0	0	0	0.0625	0	0.0625	0.125	0.1875	0.1875	0.125	0	0.0625	0.125	0.0625	0	0	0	0	0	0	0
9	0	0	0	0.0909	0.0909	0	0.2727	0	0.1818	0.0909	0	0.0909	0.0909	0	0	0	0	0.0909	0	0	0
10	0	0	0	0	0.0588	0	0.1176	0	0.1765	0.1176	0.1765	0.1176	0	0.1765	0	0	0.0588	0	0	0	0
11	0	0	0	0	0	0.0667	0.0667	0.0667	0.2	0.1333	0.2667	0	0.0667	0.0667	0.0667	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0.0952	0.0952	0.0952	0.1429	0.0476	0.0476	0.2381	0.0952	0.0476	0.0476	0.0476	0	0
13	0	0	0	0	0	0.037	0	0.037	0	0.037	0.2222	0.0741	0.2222	0.1481	0.1111	0.037	0.037	0.037	0	0	0
14	0	0	0	0	0	0	0.0313	0	0	0.0313	0.0313	0.125	0.1563	0.1563	0.125	0.0938	0.1563	0.0938	0	0	0
15	0	0	0	0	0	0	0	0	0.0208	0.0208	0	0.0417	0.1042	0.125	0.1458	0.1458	0.1458	0.1667	0.0625	0	0.0208
16	0	0	0	0	0	0	0	0	0	0	0	0	0.0625	0.0625	0.1406	0.2656	0.1875	0.1563	0.0781	0.0313	0.0156
17	0	0	0	0	0	0	0	0	0	0	0.0123	0	0.0247	0.0247	0.0247	0.1235	0.2593	0.2469	0.2099	0.0741	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0533	0.0933	0.2133	0.24	0.2	0.1067	0.0933
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0145	0.0435	0.087	0.1304	0.2174	0.2609	0.2464
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0108	0.0108	0.0753	0.129	0.2258	0.5484

Appendix 3: The table below shows us that the number of relevant images for Sixteen Pages. The result shows us the probability and index, which get from the program we have done with Matlab.

Page Number	Probability	Index	Number of Relevant Images
1	0.3194	21.0000	20
2	0.2500	21.0000	20
3	0.1927	21.0000	20
4	0.1527	21.0000	20
5	0.1304	18.0000	18
6	0.1245	18.0000	18
7	0.1179	18.0000	18
8	0.1114	18.0000	18
9	0.1051	18.0000	18
10	0.0992	18.0000	18
11	0.0937	18.0000	18
12	0.0886	18.0000	18
13	0.0839	18.0000	18
14	0.0796	18.0000	18
15	0.0757	18.0000	18
16	0.0720	18.0000	18

The Original Vector:

π (0)= [1/72,0,0,0,0,1/72,0,1/72,0,1/72,0,1/72,1/72,2/72,0,3/72,3/72,7/72,9/72,18/72,23/72]

The Testing Result for Msn:

Page Number \ Model and Testing Query	Markov Chain Model	Volcano	Tibetan Girl	Desert Camel Shadow
1	20	20	19	8
2	20	18	16	3
3	20	18	20	2
4	20	17	19	7
5	18	18	16	3
6	18	19	16	2
7	18	18	17	1
8	18	17	16	5
9	18	16	19	0
10	18	15	15	0