An empirical study on the effect demographic features and the choice of keywords have on searching success for Ananzi users

Supervisor: Weideman, M.
Student: Visser, E.B.

Abstract

Introduction: The main objective was to determine if demographic features as well as the choice of keywords used on a local search engine Ananzi, do affect searching success. In determining this, other factors of search engines need to be considered i.e. the use ofrobots and information retrieval methods.

Method: A list of keywords was obtained from Ananzi, based on searches of the previous month. The keywords were used to derive categories. These categories were integrated into the questionnaire and then hosted by Ananzi to elicit user responses. The first question asked of the user was whether or not

they used Ananzi during the month in question, which linked the categories derived from the previous month's list of keywords to the current user. Analysis: A quantitative method was used.

Results: The results of the questionnaire analysis indicate that age, race, spelling, use of Boolean operators and the number of keywords used, all affect searching success.

Conclusion: In conclusion it was determined that demographic features and the choice of keywords do affect searching success. Furthermore views on different perspectives were taken to determine what influential factors could cause demographic features to have an effect on searching success.

Introduction

After consulting with students and published journal articles, it was discovered that many people still struggle to find specific information on the internet. Although previous research has been done to ascertaining whether demographic features and the choice of keywords affect searching success, no research was found to have been conducted on a local search engine (Ananzi).

Before the Internet was born, information was generally obtained from published sources i.e. books, articles and newspapers. Considering libraries and the amount of books they stock, it seemed prudent to create a method in finding a particular book in the shortest amount of time. Indexing the title, author and even categorizing the book according to a particular set of standards or rules made the information in that particular book a lot easier to come by.

The Internet expanded exceptionally fast. There are thousands of webservers around, containing millions of websites. At the rate at which websites are created and/or destroyed, it is almost impossible to know exactly where to find specific information on the intemet. At first, websites were manually indexed creating small databases of web resources. This method was later no longer possible to maintain due to the amount of websites available on the internet. Search engines were designed with the intent of finding websites currently available on the internet, categorizing and indexing those websites automatically. Every time a user does a search on the search engine, the websites are listed according to what the search engine interprets the user is looking for. This would generally depend on factors such as keywords used, Boolean operators, searching methods and the way the searcher formulates his/her query using the above.

Literature Survey

Keywords

Many users approach Internet searching very casually. Few of them actually plan what keywords they might use, the possible query formulation options and whether or not the use of Boolean operators will increase searching success. A general approach would be to correctly nominate keywords that the user perceives to be the best term in describing what he/she is searching for (Debowski 2001). Depending on the searching success factor and based on the results found, the searcher might alter the query and retry the search that should hopefully increase the searching success factor. This becomes a loop whereby the searcher could have a few possible outcomes. Two of these possible outcomes could be; the searcher finding exactly what he/she was searching for, or time could take its toll whereby the searcher might give up entirely on the search.

According to Fidel (1991 a), a search query consists of two parts. The semantic part refers to the topic of the query, and the pragmatic part refers to how the information will be used. Once these two aspects are known and understood, the searcher can strategically start planning by selecting keywords to be used in the query formulation. The process of selecting keywords starts with the query being stated in a Natural Language. The request should then be broken up into individual components or concepts. A set of keywords could be created or derived from the concepts to represent these concepts. These keywords are then organised in a logical sequence to capture information that best represents the topic.

Keywords can be categorised into two types of search keys, namely text words and descriptors. Text words are keywords used in free text searching, whereas descriptors are keywords that form a controlled vocabulary. According to Fidel (1991 b), free text searching is easier to create due to automated generations of indexes as opposed to controlled vocabularies, which are more difficult and expensive to create - not to mention the labour intensive indexing process. Despite all these facts, controlled vocabularies are still created due to the belief that they can improve information retrieval. Increasing evidence shows that text word and descriptor searching complement one another in information retrieval. Fidel's work in the past was in the context of traditional information retrieval using electronic sources but not the internet i.e. online databases and the OPACS library systems that were used to store information.

Robots

When the internet was in the early stages of development, webpages were indexed manually and displayed on text-only browsers. Users could then browse through all the links available until they found what they were looking for. This concept eventually evolved into combining human and automatic indexing. The human aspect centred on the webmaster creating a file on the local server which he/she was able to maintain. The automatic aspects are whereby the robot software would pass through different webpages and index them, deposit the information in the file made available by the webmaster, with a description of the website. An example of such a combined indexing system is the ALIWEB (Archie-like Indexing of the Web) as cited by Chun (1999), and created by Koster. This combined indexing system paved the way to the next

generation of robots that emerged. These new robots were also known as WWWW, W4 or Web Worms, and were developed by McBryan at the University of Colorado.

A search robot typically has two main functions. The robot travels the web's Hypertext structure, retrieving and indexing documents, saving the information acquired into a database (Searching phase). Secondly the text in the document is broken down into keywords that are then incorporated into a controlled index document (Indexing phase). Furthermore robot software works constantly and will only stop indexing once the program has been terminated. Therefore when a user searches on a search engine, the search engine will access its database and display URLs of documents that have already been indexed (Chun 1999).

According to Koster (1995), these worms or spiders that are completely automated, would recursively retrieve documents and index their URLs (Uniform Resource Locators), HTML (Hypertext Markup Language) tags or even text within the web documents by making use of standard web protocols. This brings to light that there are many different types of robots that can perform different functions. The first robots were used to derive statistical analysis on the number of web servers available, average number of documents per server and average size of webpage on a particular server. Furthermore robots were also able to count the amount of a particular file type on a certain server. Other robots again could assist an author in maintaining his/her webpages, by identifying "dead links", which refer to a page that no longer exists or which has been removed. Dead links could cause user frustration or log failures. Some robots even act as mirroring tools whereby they retrieve webpages and then save them locally to allow the same page to be retrieved a lot quicker at the next access. It would seem that robots are very useful, but according to Koster (1995), not without a price. Firstly, the robots place a very high demand on the bandwidth, sometimes doubling the load on servers. Considering a robot collects documents and then decides to exclude a certain amount of documents, the price has already been paid in the usage of bandwidth. On the other hand, if a robot excludes a large number of documents to reduce the usage of bandwidth, the robots function becomes ineffective. However, robots still have some fundamental problems when indexing the web. Some advances have been made to make the robots more effective i.e. sharing results or have robots

mark webpages warning other robots thereby preventing redundancy.

According to Peacock (1998), webservers could implement the robot exclusion protocol (REP) that controls access to server resources by means of a site/robot.txt file. By implementing the REP server, resources could be managed more efficiently. Furthermore REP could aid in cutting down on spam being indexed and site administrators directing robots away from irrelevant information, stumbling blocks and black holes.

According to Green (2000), a technology had been developed by IBM, the Focused Crawler, which seeks out highly relevant information that is topic specific and ignores irrelevant sections.

Information Retrieval Methods

Information today largely exists in electronic form, stored in data repositories all over the world. Due to this electronic format, information retrieval can be accomplished by programming computers to collect webpage descriptions for indexing. Information retrieval occurs when a query made is associated with the indexed description. According to an earlier study by Blair (1990), there are 12 different information retrieval models tabulated in the Table 1.

Model	Method	Disadvantages
1	Information retrieval occurs by looking up the author, title or description in a catalogue. Only a single description query is permitted	Large quantities of information are retrieved. (information overload)
2	Multiple description queries are permitted	Cumbersome to install
3	Relaxes the retrieval constraint by allowing documents to be retrieved if the query matches only a subset of the description indexed	High possibility for redundancy - results will have equal amount of value
4	Ranking of retrieved documents depending on the query matching to the description	High possibility for redundancy - High overheads in processing
5	Permits the searcher to assign a weight to each query representing the relevance of the query	The user might find it confusing and difficult to assign these weights
6	Permits the indexer to assign a weight to each index representing the relevancy	The user may differ with regard to the weights assigned by the indexer
7	Combining model 5 and 6 permitting the indexer and the user to assign weights	High overheads in processing due to ranking
8	Creates a conceptual model to interpret the different information retrieval models	Could prevent aspects of information retrieval
9	Facilitated searchers to create complex arrangements of search terms as a query	Became too difficult for the average searcher
10	Queries passed through the entire stored text, to determine relevancy	Due to increase of database and storage size performance had to be improved
11	Provides the user with searching terms relating to the meaning of the user's initial searching terms	Terms provided may not accurately represent the searchers' perception
12	Allowed the user to see which of the related terms are more related to the initial searching terms	Processing time required to constantly update and maintain the term list

Table 1: 12 Different Models, Information Retrieval Methods and Disadvantages of Each

Today information retrieval methods have evolved to optimize information retrieval quality. One such way is for the developer to make use of meta-tags in websites. As a result, the developer exercises a degree of control over the indexing of his/her webpage. In turn this can improve the retrievability of information (Alimohammadi 2003). The concept of retrieving information depends not only on the method used by the search engine but shifts over to the user and the method he/she uses to retrieve information.

Searching Styles/Methods

The term information seeking refers to someone looking for relevant information in general but the term information searching refers to someone looking for relevant and useful online information. The way users search generally depends on the individual's perception on how to solve a problem, but according to Fidel (1984), there are two basic searching styles, the operationlist and conceptualist searching styles.

During a search, operationlists try to maximise system capabilities but refrain from altering the meaning of the search query. They try to conduct highprecision searches by optimizing the search strategy. On the other hand, conceptualist searchers might alter the meaning of the original search query to obtain the best results. They might use possible parts of the results found and combine them in different ways to gain a higher recall.

According to Fidel (1984), users often rely on a three-stage framework, dependent to each searching style, used as a search process. The preparation stage involves the user trying to understand what he/she is searching for. They attempt to create several different queries with the aim of finding the most accurate results. Furthermore, the user would decide as a search strategy which database he/she is planning to search. This stage usually occurs before the user accesses a terminal. Secondly, the search proper stage occurs as part of the initial searching in obtaining relevant results. Depending on the searching style, terms may be added or removed without altering the concept of the search in obtaining relevant information. By intentionally altering the meaning of the search, to not necessarily obtain relevant information but to view the information as a foundation to improve recall. This stage is

all about getting a feel for the response of the request in retrieving information and to narrow the playing field. The final stage is the quality control and assessment. It is in this stage that the way the user thinks could affect searching results.

During the search, users would generally create their own information retrieval methods that they might feel the need to follow, but if the methods are not working according to the user, the user might feel the need to adjust them to suit his/her needs. The operationlist assesses their results by ensuring the information is context related or the information is published in a well respected journal which might be related to the subject. Generally operationlists feel that they will never be able to obtain all the relevant information regarding the request (Fidel 1984). Conceptualists measure their accuracy of their results by ensuring the information obtained is subject related. They also feel they will be able to obtain all the relevant information regarding the request (Fidel 1984).

Fidel's research differs with other authors i.e. Marchionini (1995) which defines a seven step sequence in retrieving information. The sequence starts with a problem definition, then to source selection, query formulation, query execution, results examination, information extraction and lastly reflection.

Kuhlthau (1991) defines a six stage sequence a searcher passes through when doing information searching which includes the user's emotions and thoughts in each stage. Initiation is the first stage followed by selection, exploration, formulation, collection and presentation.

It would seem that the way a searcher thinks, their current emotions and his/her thoughts all contribute to the searching process when searching for information on the Internet (Weideman & Strümpfer 2004).

Fidel and Kuhlthau's work is based on the traditional information retrieval methods. Even through the methods might be old they might still be relevant in today's searching environment. In the past, users needed an intermediator to do searching for them on the older electronic sources, as it was relatively specialised. Today, with the Internet, that intermediator is removed leaving only the search engine database and the user. Therefore almost any method of

searching could give the searcher acceptable results including the older methods.

Boolean Operators

A large number of search engines exist, mostly based on different information retrieval systems. Operators have been made available to apply focus on the search query to increase searching success by way of the information retrieval systems. According to Weideman & Strümpfer (2004), the most commonly used operators included quotes, the inclusive operator (+) and the exclusive operator (-). The less often used Boolean operators included the 'AND', 'OR', 'NOT and 'NEAR' operators.

Frants, et al. (1999) claims that many search engines support the use of Boolean operators. Surprisingly, not too many searchers use these operators to increase searching success. For example: according to Hoelscher, in 1998 as cited by Ford, et al. (2002) some 16 million queries from the German search engine Fireball was studied to find that only 2.6% of those queries used Boolean operators. One year later, Silverstein as cited by Ford, et al. (2002), studied 993,208, 159 queries used on the AltaVista search engine and found that 20.4% of searchers use modifiers. It was concluded that the use of Boolean operators ranged from 2 to 8%.

According to Hirsh (1999), the younger generation generally does not use Boolean operators for information retrieval and when they do they tend to experience problems. According to Kreymer (2002), all searchers should be programmers because it is difficult to explain to an information retrieval system exactly what the searcher is looking for, by making use of keywords and Boolean operators. Recently it was found that information retrieval systems are moving towards the viewpoint of the searcher. This is also referred to as Natural Language Processing (NLP) whereby the user is able to communicate to the system in the form of natural language and thus exclude the complex query syntax environment.

Jansen and Pooch as cited by Ford, et al. (2002), concluded that the average searcher generally uses two terms in a query, does not use complex query syntax, does not view more than 10 results per result list and rarely uses the advanced searching features available.

Methodology

A list of more than 10 000 search queries as used during September 2004 was acquired from the Ananzi search engine. From this list, 25 categories were derived. The 25 categories were included in an initial pilot questionnaire that was created to test the initial structure. Four post-graduate students commented on the presentation, use of language and structure of the questionnaire. The questionnaire was refined by addressing the comments made to create a short, simple and easy to answer questionnaire. Before the official release the second pilot questionnaire was answered by 23 post-graduate students at the Cape Peninsula University of Technology, after the latest modifications have been applied. The questionnaire was then refined one last time, again addressing presentation, language and structure of the questionnaire. At this time a second list of search queries of June 2005 was requested from the Ananzi search engine to update the categories. The list contained more than 13 000 search queries from June 2005 that were categorised into 25 categories. The questionnaire was then placed on the Ananzi search engine from July 4, 2005 to July 31 , 2005. The questionnaire extracted keywords, success and demographic features of searchers into a data repository. See Appendix A for details on questionnaire structure and questions.

Results and Analysis

The questionnaire was placed on the Ananzi result page. This ensured that any user doing a search on Ananzi in July 2005 would see the questionnaire link. According to Ananzi, 445 883 unique users were identified using the Ananzi search engine during the month of July 2005 which meant that just as many users were exposed to the questionnaire. Of this figure, 510 questionnaires were submitted of which 482 were usable. 28 Questionnaire respondents were unusable as most of these questionnaires were submitted containing no information at all. The 28 respondents that contained information were still deemed unusable as the answers submitted were unrelated to the questions. Of the remaining usable 482 respondents, one filter needed to be applied to ensure that all information gathered from the respondents were within the scope of the methodology. This meant dividing the results into two groups: the results that were obtained from users that had used the Ananzi search engine during the month of June 2005, and those users that did not. The questionnaire results of the latter group were omitted as there was no connection to the categories supplied in the questionnaire, which was derived from the previous month's (June

2005) keywords. This left a grand total of 452 questionnaire results to work with. A second filter was applied to the actual results found by the searchers, which was divided into two categories: one of which was No Result and the other, Suitable Results.

Age versus Searching Success

The Age Group categories of successful and unsuccessful searching results are tabulated in Table 2.

Age Group	Successful Results		Unsuccessful Results		Total Results	
	No	%	No	%	No	%
19 and below	3	1.19%	8	3.17%	11	4.37%
20-25	15	5.95%	38	15.08%	53	21.03%
26 and above	48	19.05%	140	55.56%	188	74.60%
Total	66	26.19%	186	73.81%	252	100.00%

Table 2: Distribution of Results Obtained by Respondents per Age Group

As Table 2 indicates, there were more unsuccessful than successful results. Furthermore, searchers aged 26 and above seem to be the most successful, followed by searchers aged 20 to 25. Searchers aged 19 and below seem to be the least successful. These results slightly differs from Weideman & Strümpfer (2004), that found searchers aged 20 to 25 were the most successful followed by searchers aged 19 and below. Their least successful searchers were aged 26 and above.

Gender versus Searching Success

The relationship between gender and searching success was extracted from the results and is displayed in Table 3.

Gender	Successful Results		Unsuccessful Results		Total Results	
	No	%	No	%	No	%
Female	27	10.71%	82	32.54%	109	43.25%
Male	39	15.48%	104	41.27%	143	56.75%
Total	66	26.19%	186	73.81%	252	100.00%

Table 3: Distribution of Results Obtained by Respondents per Gender

Once again there were more unsuccessful than successful results. Of the successful results it was found that male searchers are more successful than female searchers. Weideman & Strümpfer (2004) results were unsurprisingly similar.

Ethnic Group versus Searching Success

The ethnic group was categorised into four general races and organised into alphabetical order. Both result types are tabulated in Table 4.

Ethnic Group	Successful Results		Unsuccessful Results		Total Results	
	No	%	No	%	No	%
Asian	3	1.21%	14	5.67%	17	6.88%
Black	6	2.43%	15	6.07%	21	8.50%
Coloured	2	0.81%	10	4.05%	12	4.86%
White	54	21.86%	143	57.89%	197	79.76%
Total	65	26.32%	186	73.68%	247	100.00%

Table 4: Distribution of Results Obtained by Respondents per Ethnic Group

Again unsuccessful results were higher than those of successful results. The White ethnic group had the majority of successful results followed by the Black ethnic group, thereafter the Asian and Coloured ethnic group. In this instance, Weideman & Strümpfer (2004) results were in complete contrast of these results with the exception of one. Their White ethnic group portrayed to be the most successful searchers, followed by the Asian and Coloured ethnic

groups that were equal to each other. The Black ethnic group followed as the least successful in searching.

Spelling of Keywords versus Searching Success

The Ananzi search engine did not, at the time of the results obtained, have a spell check program applied to the search queries. Therefore it was deemed necessary to determine whether the spelling of the keywords in the query affected searching results. The spelling of keywords and its relationship with regards to searching success is indicated in Table 5.

Spelling	Successful Results		Unsuccessful Results		Total Results	
	No	%	No	%	No	%
Correct	52	22.81%	129	56.58%	181	79.39%
Incorrect	7	3.07%	40	17.54%	47	20.61%
Total	59	25.88%	169	74.12%	228	100.00%

Table 5: Distribution of Results Obtained by Respondents per Spelling

Overall the number of unsuccessful results was more than double on that of the successful results. Furthermore with regard to the spelling of keywords, the searchers using correct spelling had a significantly higher success rate than those using incorrect spelling.

Use of Boolean Operators versus Searching Success

Boolean operators are made available by the search engine for searchers to apply focus on search queries to increase the possibility of searching success (Weideman & Strümpfer 2004). The relationship between both the use and non use of Boolean operators with regards to successful and unsuccessful searching results is portrayed in Table 6.

Use of Boolean Operators	Successful Results		Unsuccessful Results		Total Results	
	No	%	No	%	No	%
Yes	7	2.83%	38	15.38%	45	18.22%
No	58	23.48%	144	58.30%	202	81.78%
Total	65	26.32%	182	73.68%	247	100.00%

Table 6: Distribution of Results Obtained by Respondents per use of Boolean Operators

Very few searchers made use of Boolean operators. When they did, it would seem to be problematic as indicated by a higher number in unsuccessful searching results. Searchers not making use of Boolean operators also seem to have a high unsuccessful result rate.

Number of Keywords used in the Search Query versus Searching Success

The number of keywords are categorised from one keyword up to six keywords used per query as indicated in Table 7. Although the use of more than six keywords were detected in the results, they were not listed due to the 0% searching success rate.

Word Count	Successful Results		Unsuccessful Results		Total Results	
Number of Words	No	%	No	%	No	%
1 Word	15	6.88%	35	16.06%	50	22.94%
2 Words	32	14.68%	56	25.69%	88	40.37%
3 Words	5	2.29%	33	15.14%	38	17.43%
4 Words	2	0.92%	19	8.72%	21	9.63%
5 Words	3	1.38%	12	5.50%	15	6.88%
6 Words	2	0.92%	4	1.83%	6	2.75%
Total	59	27.06%	159	72.94%	218	100.00%

Table 7: Distribution of Results Obtained by Respondents per Word Count

As displayed in Table 7, the use of two keywords during a search seems to be the most successful, followed by the use of one keyword. Thereafter searching success decreases as the use of keywords increase with the exception of the use of four and five keywords respectively. Weideman & Strümpfer (2004) found a similar result whereby the use of two keywords during a search significantly increased the searching success factor.

Statistical Results of Internet Searching

The final comparison indicates the different categories of search topics with regards to being successful and unsuccessful.
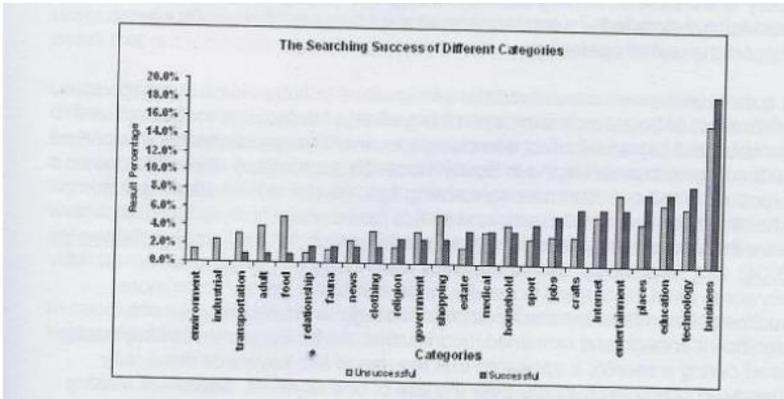
Graph 1 :Distribution of Results Obtained by Respondents per Different
  Category

Two distinct colours were used in Graph 1. The dark red colour represents the success factor per category. Adding all the percentages of the dark red colour will give you 100% and therefore represents all the useable questionnaire respondents that were successful during their search. The light blue colour represents the unsuccessful factor per category. Adding all the percentages of the light blue colour will give you 100% and therefore represents all the useable questionnaire respondents that were unsuccessful during

their search. Due to this every participant features somewhere on the graph, either in the successful or unsuccessful categories.

 Conclusion

Previous research has revealed that searchers in general are largely unsuccessful in obtaining accurate results. Many things could contribute to this fact, including search engines syntax. There are a few major factors that could influence searching success, one which is the choice of keywords used during a search. The use of keywords is crucial to searching success as they are the only guide to the search engine to retrieve relevant information. Users should be reminded that the order in which keywords are placed during a search could be interpreted differently on different search engines. Other factors are: user searching methods and search engine information retrieval methods. Searching methods used are dependent on searcher preference and could differ from user to user. Information retrieval methods are used by the search engine and could differ from search engine to search engine. Another factor is robots software used by search engines. Searchers experience much frustration due to irrelevant information being retrieved by search engines. This is sometimes a result of robot software incorrectly interpreting a webpage as relevant information and then indexing that webpage. Another factor is users making use of Boolean operators during a search. Different search engines make different operators available to the user. These operators apply focus to the search query to increase searching success. Some search engines automatically include operators in the search query, where other search engines do not support the use of operators.

A questionnaire was considered the best method to accumulate anonymous information of the searcher in determining whether demographic features and the choice of keywords affect searching success. The questionnaire was placed on the largest search engine in South Africa. No parameters were enforced on exposure of the questionnaire concerning age, gender and/or ethnic group. After the cleansing of information, statistics was derived from it. The statistics were then compared to a previous but similar research (Weideman & Strümpfer

2004). Not surprisingly, similar statistics were found on gender and number of keywords used. According to results, male searchers appear to be more successful than female searchers. Unfortunately

439

the numbers were not significant enough and remained inconclusive. As for the number of keywords used during a search, it appeared that the use of two keywords drastically increased searching success over the use of one keyword. Searchers making use of more than two keywords during a search could still be successful, even though the results revealed that these searchers generally tend to be less successful. This could be due to search engines not accurately interpreting Natural Language searching. This could also mean that users do not possess the knowledge and/or skills required to be efficient searchers.

Statistics on age revealed that searchers aged 26 and above seem to be the most successful, followed by searchers aged 20 to 25. Searchers aged 19 and below seem to be the least successful. This could be due to the older generation being around during the development of the Internet. A few years ago, technology was not as user friendly as it is today. This meant users really had to understand technology details to find something on the Intemet. The younger generation, even though they are educated on the use of the Internet, could be more interested in online communication, games and entertainment. As a result, it might distort their holistic view of information retrieval.

Some surprising results were derived from searching success concerning the ethnic groups. The White ethnic group was the most successful in searching success, followed by the black ethnic group. Weideman & Strümpfers (2004) results were slightly different concerning the black ethnic group as they were the least successful in Internet searching. Even through the Black ethnic group is the largest ethnic group in South Africa, it could be assumed that resources and Internet availability is limited due to a lack of funds. On the other hand it could also be assumed that the development and education of the Black ethnic group is well on its way when bearing in mind the sample size.

Statistics portray that searchers are more successful when spelling keywords correctly. As mentioned before, keywords are crucial to searching success for the one reason that search engines need to interpret keywords correctly to retrieve accurate results. To ensure that search queries are spelt correctly it could be suggested to Ananzi to incorporate a dictionary spell check when requesting results from the search engine.

Results revealed that searchers using Boolean operators appear to more unsuccessful. One could assume that users are aware of operators but it would seem that not too many users know how to use operators correctly.

Lastly, Graph 1 represents the successful and unsuccessful searching in different categories. It is clearly portrayed that entertainment, places, education, technology and business are the five most popular searched categories. Taking a closer look at the business category one could see that this category has the highest success rate. This could be due to businesses ensuring that their websites are created correctly which in turn could assist their website being indexed correctly. This is an essential aspect for a business to increase revenue via Internet shopping.

In conclusion it was determined that demographic features and the choice of keywords do affect searching success.

Further Research

- Ascertaining whether or not Fidel's (1984, 1991a, 1991b) research in the past, regarding selection of keywords and searching methods can still be applied today, to retrieve accurate results.
- Determining which categories have the highest searching success rate and why there success rates are so high.

Acknowledgements

References

- Alimohammadi, D. (2003), Meta-tag: a means to control the process of Web indexing. Online Information Review 27(4), 238-242.

- Blair, D.C. (1990), Language and representation in information retrieval.
  Amsterdam, Netherlands: Elsevier Science Publishers B.V.
- Chun, T. Y. (1999), World Wide Web Robots: An Overview. Online & CDROM Review 23(3), 135-142.
- Debowski, S. (2001 Wrong way: go back! An exploration of novice search behaviour while conducting as information search. The Electronic Library 19(6), 371-382.
- Fidel, R. (1984), Online Searching Styles: A Case-Study Model of Searching Behaviour. Joumal of the American Society for Information Science 35(4), 211-221.
- Fidel, R. (1991 a), SearcherS Selection of Search Keys: I. The Selection.
  Journal of the American Society for Information Science 42(7), 490-500.
- Fidel, R. (1991 b), Searchers' Selection of Search Keys: II. Controlled Vocabulary of Free-Text Searching. Journal of the American Society for Information Science 42(7), 501-514.
- Ford, N., Miller, D. & Moss, N. (2002), Web search strategies and retrieval effectiveness: an empirical study. Journal of the American Society for Information Science 58(1 30-48.
- Frants, V.I., et al. (1999), Boolean Search: Current State and Perspectives. Journal of the American Society for Information Science 50(1), 86-95.
- Green, D. (2000), The evolution of searching. Online Information Review 24(2), 124-137.
- Hirsh, S.G. (1999), Children's Relevance Criteria and Information Seeking on Electronic Resources. Journal of the American Society for Information Science 50(14), 1265-1283.
- Koster, M. (1997), Robots in the Web: threat or treat? Retrieved 26

September, 2005 from http://www.robotstxt.org/wc/threat-or-treat.html • Kreymer, O. (2002), An evaluation of help mechanisms in natural language information retrieval systems. Online Information Review 26(1)30-39.

- Kuhlthau, C.C. (1991), Inside the Search Process: Information Seeking from the User's Perspective. Journal of the American Society for Information Science 42(5), 361-371.
- Marchionini, G. (1995), Information Seeking in Electronic Environments. New York: Cambridge University Pr.

- Peacock, I. (1998), <u>Showing Robots the Door.</u> Retrieved 26 September,
  2005 from http://www.ariadne.ac.uk/issuel 5/robots/
- Weideman, M., Strümpfer, C. (2004), The Effect of Search Engine Keyword Choice and Demographic Features on Internet Searching Success. Information Technology and Libraries 23(2), 58-65.

How to cite this paper:

Eugene B. Visser, Melius Weideman, (2005) "An empirical study on the effect demographic features and the choice of keywords have on searching success for Ananzi users." Information Research, x(x) paper xxx [Available at http://InformationR.net/ir/xx/paperxxx.htmt

ananzi

Ananzi Search Engine Questionnaire

This is a survey questionnaire used for research done at the Cape Peninsula University of Technology on search engines. The results of the questionnaires will ascertain whether or not demographic features affect searching results. The information provided on the questionnaire will be used for statistical analysis, and treated as confidential as the questionnaire is anonymous. Your involvement in this study is voluntary. Please answer all the questions as accurate as possible.

1. Have you used the Ananzi search engine for keyword searching during the month of June 2005?
   - ○ Yes
   - ○ No

Please answer the following questions about yourself.

2. Age range:
   - ○ 15 and below
   - ○ 16-25
   - ○ 26 and above

3. Gender:
   - ○ Male
   - ○ Female

4. Ethnic group:
   - ○ Asian
   - ○ Black
   - ○ Coloured
   - ○ White
   - ○ Other

The following questions relate to the last search you have done during June 2005 on the Ananzi search engine:

5. Please type your entire search query of this search in, in the box provided below:

6. How did you feel about the result of this search?
   - ○ Did not find what I was looking for
   - ○ Found results close but not exactly what I was looking for
   - ○ Found what I was looking for after more than one search
   - ○ Found what I was looking for on one search after some time
   - ○ Immediately found what I was looking for

7. Did you make use of Boolean operators during this search? (AND, OR, +, -, ")
   - ○ Yes
   - ○ No

8. Kindly tick (one or more) boxes below to indicate the category of this search.
   - ☐ Adult
   - ☐ Business
   - ☐ Clothing and accessories
   - ☐ Crafts and Hobbies
   - ☐ Education
   - ☐ Entertainment
   - ☐ Environment
   - ☐ Fauna and Flora
   - ☐ Food and Drink
   - ☐ Government and History
   - ☐ Household
   - ☐ Industrial equipment
   - ☐ Internet
   - ☐ Jobs
   - ☐ Medical
   - ☐ News
   - ☐ Places and Traveling
   - ☐ Real Estate
   - ☐ Relationship
   - ☐ Religion
   - ☐ Shopping
   - ☐ Sport
   - ☐ Technology
   - ☐ Transportation
   - ☐ Other

9. Would you recommend the Ananzi search engine to other searchers?
   - ○ Yes
   - ○ No

10. Please motivate your answer of question nine.

Thank you for your time!

[Submit] [Reset]

The Questionnaire as it was on the Ananzi result page (2005).