

**SEARCH ENGINE  
PERFORMANCE EVALUATION**

**A THESIS SUBMITTED TO  
THE GRADUATE SCHOOL OF APPLIED SCIENCES  
OF  
NEAR EAST UNIVERSITY**

**by**

**KEZBAN ALPAN**

**IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR  
THE DEGREE OF MASTER OF SCIENCE  
IN  
COMPUTER  
INFORMATION SYSTEMS**

**NICOSIA 2010**

**SEARCH ENGINE  
PERFORMANCE EVALUATION**

**A THESIS SUBMITTED TO  
THE GRADUATE SCHOOL OF APPLIED SCIENCES  
OF  
NEAR EAST UNIVERSITY**

**by**

**KEZBAN ALPAN**

**In Partial Fulfillment of the Requirements for  
the Degree of Master of Science  
in  
Computer Information Systems**

**NICOSIA 2010**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Kezban Alpan

Signature:

Date:

## **ABSTRACT**

The value, price and importance of information is increasing day by day in the information age that we are living in with internet becoming the major information resource for people. Because internet has become such a phenomenon, the number of documents has also increased rapidly. This situation makes finding information on the internet without the use of a web search engine impossible. Today, search engines are profitable sectors and engine owners are in great competition with each other to be the most popular. There are many web search engines which work with complex algorithms; however, the main topic of concern is which search engine can bring the greatest solution in the shortest time period. It is from this aspect that this study focused on the evaluation of 5 popular search engines. Additionally, the study includes research on search engine preferences of university students. Performance evaluation tests have been conducted in two different aspects. For the first aspects, 10 queries have been prepared related to favourite search queries of 2009, and another 10 questions have been prepared in relation to information technology and information systems for the second aspect. This query was searched through the use of selected search engines and the first 5 pages which cover 50 answers have been checked. Results have been categorized as relevant, irrelevant and broken link. At the end of the research, precision and currency values has been calculated according to sums and averages. In relation to search engine preferences of university students, a questionnaire was prepared and applied on 300 hundred students who were selected randomly from 15 different faculties. At the end of the research, results indicated that Google is the most widely used search engine by students, but results of performance evaluation test indicate Bing as the best search engine which gives the highest precision and lowest currency. The results of this study add empirical data to the relevant field and are expected to help computer science students, experts, instructors, and everyone else who wants to reach information via the internet.

*Keywords:* Search engine performance evaluation, Google, Yahoo, Bing, search engines, information retrieval systems

## ÖZ

Yaşamakta olduğumuz bilgi çağında bilginin değeri, fiyatı ve önemi gün geçtikçe artarken, internet de insanların başlıca bilgi kaynağı haline gelmiştir. İnternetin fenomen haline gelmesi barındırdığı döküman sayısının her geçen gün daha hızlı bir şekilde artmasına neden olmaktadır. Bu durum internet üzerinden bilgi arayışını arama motorları olmadan imkansız hale getirmektedir. İnternetin olmazsa olmazı arama motorları başlıca bir sektör ve ticari kazanç kaynağı halini alırken, arama motoru sahibi kuruluşlar da kendilerini geliştirmek ve internette popüler olmak adına amansız bir rekabet içine girmiş bulunmaktadırlar. Esas sorun hangi arama motorunun en doğru bilgi veya bilgileri en kısa süre içerisinde sağlayabileceğidir. Bu noktadan hareketle çalışma sektördeki en popüler beş arama motorunun performans değerlendirmesini içermektedir. Ayrıca üniversite öğrencilerinin hangi arama motorunu tercih ettiklerini ve tercih nedenlerini belirlemek amacıyla güden bir araştırma da bu çalışmada yer almıştır. Arama motorları performans değerlendirme testi iki farklı yönde yapılmıştır. İlk test için 2009'un en çok aranan terimleri yazar tarafından derlenerek on adet farklı sorgu cümlesi oluşturulmuş ve bu sorgu cümleleri tek tek arama motorlarında aranmıştır. İkinci test içinse Bilgi Teknolojisi ve Bilgi Sistemleri hakkında belirlenen on sorgu cümlesi ile arama motorları teste tabi tutulmuştur. Veri elde etmek için her bir arama sonucundan ilk beş sayfa yani ilk 50 link baz alınmış ve çağrılan sonuçlar ilgili, ilgisiz veya kırık link olarak sınıflandırılmıştır. Öğrenciler arasında yapılacak olan çalışma için veri toplamak amacı ile anket oluşturulmuş ve bu anket 15 farklı fakülteden rastgele seçilen toplam 300 öğrenci üzerinde uygulanmıştır. Performans değerlendirme testi verileri toplam ve avaraj olarak değerlendirilerek arama motorlarının duyarlılık ve yenilik oranları hesaplanarak en iyi performansı gösteren arama motoru belirlenmiştir. Öğrenciler üzerinde yapılan anket çalışması ise SPSS programı ile değerlendirilerek, öğrencilerin arama motoru kullanım frekansları ve nedenleri belirlenmiştir. Elde edilen bulgulara göre öğrencilerin büyük çoğunluğu arama motoru olarak Google kullanmaktadır fakat performans değerlendirme testi sonuçları Google'ı üçüncü sıraya iterken en iyi performans gösteren arama motoru olarak Bing'i belirlemiştir. Yapılan çalışmanın bilgisayar bilimleri öğrencilerine, eğitmenlerine, uzmanlarına ve internet üzerinden bilgi edinme ihtiyacı olan kişi veya kişilere yardımcı olması umulmaktadır.

*Anahtar Kelimeler:* Arama motoru performans değerlendirme, Google, Yahoo, Bing, arama motorları, Bilgi erişim sistemleri

To my parents

## **ACKNOWLEDGEMENTS**

I would like to thank to our Dean Prof.Dr. Aykut POLATOĞLU and Vice Dean and chairperson of Business Administration Department Assist.Prof.Dr. Şerife EYÜPOĞLU for their support and invaluable help through this thesis.

I would like to thank to chairperson of Computer Information Systems Department Assist.Prof.Dr. Yalçın AKÇALI for his substantial support and help.

I would like to sincerely thank to my advisor Assist.Prof.Dr.Nadire ÇAVUŞ for her great support and encouragement.

Also I would like to thank to Assist.Prof.Dr. Mustafa MENEKAY. More than being a teacher, he is the one who supported me during my master degree education and made me to walk through my dreams.

I would like to especially thank to my parents and my brother for their endless support during all my life. Also I would like to thank to someone special who was always beside me with his great support.

Finally I would like to thank my entire student friend who applied to research for this thesis.

## TABLE OF CONTENTS

PLAGIARISM.....	vi
ABSTRACT.....	vii
ÖZ.....	viii
DEDICATION.....	ix
ACKNOWLEDGMENTS.....	x
TABLE OF CONTENTS.....	xi
LIST OF TABLES.....	xv
LIST OF FIGURES.....	xvi
LIST OF ABBREVIATIONS.....	xvii

### CHAPTER I

<b>1. INTRODUCTION.....</b>	<b>1</b>
1.1 The problem.....	3
1.2 The purpose of the study.....	5
1.3 The significance of the study.....	6
1.4 Limitations of the study.....	6
1.5 Structure of the study.....	7
1.6 Summary.....	7

### CHAPTER II

<b>2. REVIEW OF LITERATURE.....</b>	<b>8</b>
2.1 Summary.....	12

### CHAPTER III

<b>3. CONCEPTUAL FRAMEWORK.....</b>	<b>13</b>
3.1 Information retrieval systems.....	13
3.1.1 Database.....	15
3.1.2 Terms.....	15



3.1.3 Documents.....	16
3.1.4 Queries.....	18
3.1.5 User interface.....	19
3.1.6 Retrieval rules.....	19
3.2 Search engines technology.....	23
3.2.1 Architecture.....	24
3.2.1 Indexing.....	25
3.2.3 Presentation of documents.....	26
3.2.4 Efficiency.....	27
3.2.5 Ranking and retrieval function.....	28
3.3 Evaluated search engines.....	30
3.3.1 Google.....	30
3.3.2 Yahoo!.....	30
3.3.3 Bing.....	31
3.3.4 Ask.....	31
3.3.5 AOLSearch.....	32
3.3 Summary.....	32
<b>CHAPTER IV</b>	
<b>4. METHODOLOGY.....</b>	<b>33</b>
4.1 Research model.....	33
4.2 Students.....	33
4.3 Data collection.....	35
4.4 Data analysis.....	36
4.5 Duration and resources.....	37
4.6 Application.....	37
4.7 Summary.....	39
<b>CHAPTER V</b>	
<b>5. RESULTS.....</b>	<b>40</b>
5.1 SE usage of students.....	40
5.2 Differences between SEs.....	42
5.3 Students criteria for SE prefers.....	43
5.4 Students' opinions about SEs.....	44

5.5 Other prefers of students' request for SEs.....	45
5.6 Findings and interpretations of search engine performance evaluation test.....	47
5.6.1 Precision and currency ratios for top search queries of 2009 .....	48
5.6.2 Precision and currency ratios for IT and IS queries.....	50
5.6.3 Precision and currency ratios for general test results.....	51
5.8 Summary.....	54
 <b>CHAPTER VI</b>	
<b>6. CONCLUSION, DISCUSSIONS AND RECOMMENDATIONS.....</b>	<b>55</b>
6.1 Conclusion and discussions.....	55
6.2 Recommendations.....	57
 <b>REFERENCES.....</b>	<b>58</b>
 <b>APPENDICES</b>	
A. Performance test results for favourite queries of 2009.....	68
B. Performance test results for IT and IS queries.....	69

## LIST OF TABLES

<b>Table 3.1:</b> Summary of retrieval rules.....	20
<b>Table 3.2:</b> Presentation of search results.....	27
<b>Table 4.1:</b> Faculty and department details of students.....	34
<b>Table 4.2:</b> Time schedule.....	37
<b>Table 5.1:</b> SE usage frequencies of students.....	40
<b>Table 5.2:</b> One-sample <i>t</i> -test for SE usage frequencies among students.....	42
<b>Table 5.3:</b> Precision ratios for top search queries of 2009.....	48
<b>Table 5.4:</b> Dead link ratios for top search queries of 2009.....	49
<b>Table 5.5:</b> Precision ratios for IT and IS queries.....	50
<b>Table 5.6:</b> Dead link ratios for IT and IS queries.....	51
<b>Table 5.7:</b> General results for precision ratios.....	52
<b>Table 5.8:</b> General results for currency ratios.....	53

## LIST OF FIGURES

<b>Figure 3.1:</b> Traditional IR system.....	14
<b>Figure 3.2:</b> Vector space IR system mode.....	21
<b>Figure 3.3:</b> High level architecture of a standard web crawler.....	25
<b>Figure 5.1:</b> Favorite SEs of students in numbers.....	41
<b>Figure 5.2:</b> Important criterions for SE preferring.....	43
<b>Figure 5.3:</b> General opinions for SEs.....	44
<b>Figure 5.4:</b> User's request for homepage.....	45
<b>Figure 5.5:</b> General information or quotation prefers of users.....	46
<b>Figure 5.6:</b> Percentage for users when leave search.....	47

## LIST OF ABBREVIATIONS

<b>AOL</b>	American Online
<b>ARPANET</b>	Advanced Research Project Agency Network
<b>DARPA</b>	Defense Advanced Research Project Agency
<b>e-commerce</b>	Electronic commerce
<b>e-mail</b>	Electronic mail
<b>FTP</b>	File transfer protocol
<b>IR</b>	Information retrieval
<b>IS</b>	Information systems
<b>IT</b>	Information technology
<b>KWIC</b>	Key-Word-In-Context
<b>NEU</b>	Near East University
<b>SD</b>	Standard deviation
<b>SE</b>	Search engine
<b>SEO</b>	Search engine optimization
<b>TRNC</b>	Turkish Republic of Northern Cyprus
<b>URL</b>	Uniform resource locator
<b>Veronica</b>	Very Easy Oriented Net-Wide Index to Computerized Archives
<b>web</b>	World Wide Web
<b>WWW</b>	World Wide Web