

A Study on Performance Evaluation of International Web Search Engines Based on English, Arabic, and Turkish Languages in Various Categories

Mohamad Yahya J. Namk

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Master of Science
in
Computer Engineering

Eastern Mediterranean University
February, 2015
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Serhan Çiftçioğlu
Acting Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Master of Science in Computer Engineering.

Prof. Dr. Işık Aybay
Chair, Department of Computer Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Computer Engineering.

Asst. Prof. Dr. Yılmaz Bitirim
Supervisor

Examining Committee

1. Asst. Prof. Dr. Yılmaz Bitirim

2. Asst. Prof. Dr. Cem Ergün

3. Asst. Prof. Dr. Önsen Toygar

ABSTRACT

This study illustrates the performance evaluation of the three international Web search engines Google, Bing, and Yahoo based on the three languages English, Arabic, and Turkish in six various categories, i.e., “business”, “culture”, “health”, “sports”, “technology”, and “travel”. A total of 30 categorised queries (five per category) were determined and each of them was run in English, in Arabic, and in Turkish languages separately on each of the three Web search engines. At each run, every document retrieved in the first 10 was classified as “relevant” or “non-relevant” and, in each category, precision ratios and normalized recall ratios were calculated for each of the five queries at every Web search engine and language pair. Then, they were used in the evaluation. In all categories, generally, the Web search engines performed better performances in English document retrieval compared to Arabic document retrieval and Turkish document retrieval.

Keywords: Web Search Engine, English Language, Arabic Language, Turkish Language, Information Retrieval.

ÖZ

Bu çalışma üç uluslararası Web arama motoru Google, Bing, ve Yahoo'nun, altı farklı kategorideki, “ekonomi”, “kültür”, “sağlık”, “spor”, “teknoloji”, ve “seyahat”, olmak üzere üç dile dayalı, İngilizce, Arapça, ve Türkçe, performans değerlendirmesini göstermektedir. Toplamda kategorize edilmiş 30 sorgu (her kategori için 5 tane) belirlenmiştir ve her biri İngilizce, Arapça, ve Türkçe olarak üç Web arama motorunda ayrı ayrı çalıştırılmıştır. Her çalıştırmada, ilk 10' da alınan her belge “ilgili” yada “ilgisiz” olarak sınıflandırılmıştır ve her kategorideki her 5 sorgu için, her Web arama motoru ve dil çiftinin duyarlılık oranları ve normalize sıralama oranları hesaplanmıştır. Daha sonra, bunlar değerlendirmelerde kullanılmıştır. Tüm kategorilerde, Web arama motorları Arapça dokümana erişim ve Türkçe dokümana erişim performanslarına kıyasla genelde İngilizce dokümana erişimde daha iyi performans göstermiştir.

Anahtar Kelimeler: Web Arama Motoru, İngilizce Dili, Arapça Dili, Türkçe Dili, Bilgi Erişim.

To Türkmen of Iraq
To my precious family
To my beloved fiancée
To my late friend

ACKNOWLEDGMENT

In the name of Allah, the Beneficent, the Merciful. This study has been done under the supervision of Asst. Prof. Dr. Yıltan Bitirim. Therefore and foremost, I would like to express my heartfelt appreciation due to his continuous support, guidance, and endless encouragement that without his help I would not have reached the success I have so far.

I extend my sincere thanks to my family, especially my dear father "Jihad Ali", my dear mother "Fryal Ali", my dear sisters "Zainab, Maryam, and Rahmah", and my sweet nephews "Sarah, Ahmed, and Maria", I am beholden to all of you for everything of my life.

Finally, a special gratitude to my beloved fiancée Şeyda Türker, I would like to express my thanks by saying I love you forever.

TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZ	iv
DEDICATION	v
ACKNOWLEDGMENT	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xi
1 INTRODUCTION	1
2 A BRIEF OVERVIEW ON WEB SEARCH ENGINES.....	4
3 METHODOLOGY.....	6
3.1 Determination of Web Search Engines	6
3.2 Determination of Languages	6
3.3 Determination of Categories and Queries	7
3.4 Search Settings	7
3.5 Query Run and Document Evaluation	9
3.6 Precision Ratio	10
3.7 Normalized Recall Ratio	12
4 EXPERIMENTAL RESULTS	16
4.1 DDs, RDs, and SDs.....	16
4.2 Relevant Documents	19
4.3 Mean Precision Ratio Results	21
4.4 Mean Normalized Recall Ratio Results	28
4.5 Summary and Discussion.....	35

5 CONCLUSION	40
REFERENCES.....	41
APPENDICES	45
Appendix A: Mean Precision Ratios.....	46
Appendix B: Mean Normalized Recall Ratios.....	47

LIST OF TABLES

Table 1: The list of queries in English, Arabic, and Turkish languages for the six categories.....	8
Table 2: Number of DDs, RDs, and the SDs	18
Table 3: Number of “relevant” Documents Retrieved	20
Table 4: The highest performance displayed by the engines in English language	38
Table 5: The highest performance displayed by the engines in Arabic language.....	38
Table 6: The highest performance displayed by the engines in Turkish language	38

LIST OF FIGURES

Figure 2.1: List of the top most popular search engines	5
Figure 3.1: Evaluation of the first 10 documents retrieved for the query “brand licensing process”	11
Figure 3.2: Finding out R^+	13
Figure 3.3: Finding out R^-	14
Figure 3.4: The ideal retrieval output for “brand licensing process” query’s retrieval output	15
Figure 4.1: Mean precision ratios of Google Web search engine	21
Figure 4.2: Mean precision ratios of Bing Web search engine	22
Figure 4.3: Mean precision ratios of Yahoo Web search engine	23
Figure 4.4: Mean precision ratios for English language	24
Figure 4.5: Mean precision ratios for Arabic language.....	25
Figure 4.6: Mean precision ratios for Turkish language	26
Figure 4.7: Average mean precision ratios based on language for Web search engines	27
Figure 4.8: Mean normalized recall ratios of Google Web search engine.....	28
Figure 4.9: Mean normalized recall ratios of Bing Web search engine.....	29
Figure 4.10: Mean normalized recall ratios of Yahoo Web search engine	30
Figure 4.11: Mean normalized recall ratios for English language	31
Figure 4.12: Mean normalized recall ratios for Arabic language	32
Figure 4.13: Mean normalized recall ratios for Turkish language.....	33
Figure 4.14: Average mean normalized recall ratios based on language for Web search engines	34

LIST OF ABBREVIATIONS

URL	Uniform Resource Locator
-----	--------------------------

Chapter 1

INTRODUCTION

In recent years, Web search engines have become the most important tools for the Web users. The importance of the Web search engines may come as a consequence of the huge number of pages on the Web.

Gao et al. reported in 2013 that “search engines are the most important domain in Internet services in terms of the number of page views and daily visitors” [1]. Beside of these, Grosjean estimated in the study [2] that more than 50% of the world’s inhabitants use two or more than two languages (or dialects) in daily life; thus, it could be said that the Web users may use one or more than one languages in addition to their mother tongue. Furthermore, in the mid of 2014, there were almost 3 billion Internet users in the world which was around 40% of the total world’s inhabitants [3] and Internet users search for information in various categories according to their needs.

Bringing all these together let us to have a study on evaluation of the performance of the three international Web search engines Google [4], Bing [5], and Yahoo [6] based on the three languages English, Arabic, and Turkish in six various categories, i.e., “business”, “culture”, “health”, “sports”, “technology”, and “travel”.

The aims of this study are: to present information to the Web users who know one or more of these languages in order to have idea about the Web search engines and have advantage for better handling of the Web search engines; to motivate corresponding researchers in order to improve search engines; and to motivate search engine providers in order to improve their search engines.

Several related studies are as follows:

Efthimiadis et al. [7] evaluated the five international Web search engines A9, AltaVista, Google, MSN, and Yahoo and the five Greek Web search engines Anazitisi, Ano-Kato, Phantis, Trinity, and Visto by using the navigational queries. These queries were homepage finding queries for known Greek organizations from the ten categories “government departments”, “universities”, “colleges”, “travel agencies”, “museums”, “media (TV, radio, newspapers)”, “transportation”, and “banks”. Furthermore, they evaluated the freshness of the Web search engine indexes. One of their findings is that the international Web search engines performed better than the Greek Web search engines.

Çakır et al. [8] evaluated the international image search engines Google, Yahoo, Ask, and MSN, for retrieving images based on various topics, “IT & Internet”, “Food & Beverage Brands”, “Broadcast Media”, “Automotive Manufacturers”, “Movies”, “Pharmaceutical & Medical Products”, and “Travel Destination & Accommodations”, with five English queries per topic. Mean precision ratios at 10, 20, 30, and 40 cut-off point values for each engine and topic pair were found and used in the evaluations.

Goel and Yadav [9] presented an approach based on page level keywords for search engine evaluation. They used 40 educational queries and measured the performance of the international Web search engines Google, Yahoo, and Bing with their approach. Afterwards, they used the same queries and measured the performance of the same engines based on human ranking in terms of precision to verify their approach. Due to the similar results were obtained from the two measurements, they declared that page level keywords could be good criteria to evaluate search engines.

Moukdad [10] evaluated the performance of the international Web search engines AltaVista, AlltheWeb, and Google and the local Web search engines Al bahhar, Ayna, and Morfix (the arabic module) based on their ability to retrieve documents containing morphologically related Arabic terms. The international Web search engines were limited in retrieving Arabic documents. Using the exact forms of Arabic words as search terms caused the loss of high number of documents.

Note that this study is an extended version of our study [11] and additionally includes evaluations based on normalized recall ratios.

The organization of this study is as follows: The next chapter of this study contains general information about Web search engines; chapter 3 includes the methodology used in our study; chapter 4 demonstrates the experimental results; and the last chapter presents the conclusion of this study.

Chapter 2

A BRIEF OVERVIEW ON WEB SEARCH ENGINES

Web search engine is a Web-based information retrieval system which enables the Web users to search for information on the Web by writing one or more keywords in a query. Every Web user is well aware of the importance of Web search engines due to they shorten the time spend in searching for particular information. Beside their simple view, Web search engines have complex functions working at behind.

In the early 1990s, the world has witnessed the emergence of Web search engines for the first time. The first tool used for searching on the Internet was Archie search engine, which was created by Alan Emtage, Bill Heelan, and Peter J. Deutsch in 1990 [12]. In 1994, David Filo and Jerry Yang collected their favorite Websites and listed all the URL addresses of these Websites on a single guide page and they managed the page and grew it until they founded Yahoo [13]. In the early 1996, the students of Stanford University Larry Page and Sergey Brin worked together on a search engine that they named it “BackRub” at that time. In 1997, they registered “Google.com” as a domain for the search engine and in 1998 they established Google Company [14]. In the middle of 2009, Microsoft CEO Steve Ballmer unveiled Bing, which was previously known as "Live Search", "Windows Live Search", and "MSN Search" [15].

A list in Figure 2.1 which is obtained from the study of eBizMBA [16] shows the top five most popular Web search engines in nowadays.



Figure 2.1: List of the top most popular search engines

Chapter 3

METHODOLOGY

In this chapter, the methodology used to evaluate the Web search engines based on the three languages in the six categories is explained.

3.1 Determination of Web Search Engines

We initially selected the highest ranked first three international Web search engines Google, Bing, and Yahoo as derived from the study of eBizMBA [16].

3.2 Determination of Languages

Three languages English, Arabic, and Turkish were selected. The reasons behind choosing these three languages were as follows: One of the authors knows all these three languages; in the Q-Success' study [17], these three languages were recorded in the first 36 of "Usage-of-content-languages-for-websites" list which English language was ranked as the first (55.6%), Turkish language was ranked as eleventh (1.4%), and Arabic language was ranked as fourteenth (0.8%); and the numbers of English language speakers, Arabic language speakers, and Turkish language speakers were 335 million, 237 million, and 70.8 million, respectively [18].

3.3 Determination of Categories and Queries

Four popular international news providers CNN [19], euronews [20], Fox News [21], and The Huffington post [22] were selected. Then, six category topics, i.e., “business”, “culture”, “health”, “sports”, “technology”, and “travel”, were determined from these news providers. For each category, five specific queries were determined. Note that each query was especially considered to represent only its corresponding category that would be listed underneath. All queries (30 in total) in English, Arabic, and Turkish languages based on the six categories are given in Table 1.

3.4 Search Settings

Before run the queries, some preferences of the three Web search engines were changed:

- The language setting of the Web search engines was changed based on the query language to retrieve the search results in the same language of the query.
- SafeSearch feature were turned off to reach the search results without any filters.
- The location setting was specified to retrieve the search results from all over the world.

Table 1: The list of queries in English, Arabic, and Turkish languages for the six categories

Language	Query #	Business	Query #	Culture	Query #	Health	Query #	Sports	Query #	Technology	Query #	Travel
E.		brand licensing process		ancient greek gods		addiction treatment stages		billiards rules		cloud storage		first class flight features
A.	1	عملية ترخيص الماركة	1	الالهة اليونانية القديمة	1	مراحل علاج الادمان	1	قوانين البلياردو	1	تخزين السحابي	1	مزايا رحلات درجة الاولى
T.		marka lisanslama süreci		antik yunan tanrıları		bağımlılık tedavi aşamaları		bilardo kuralları		bulut depolama		birinci sınıf uçuş özellikleri
E.		free trade agreement		ethnic symbols		breast cancer		boxing techniques		smart watch features		safari tours
A.	2	اتفاقية تجارة الحرة	2	رموز الإثنية	2	سرطان الثدي	2	تقنيات الملاكمة	2	مميزات ساعات الذكية	2	رحلات السفاري
T.		serbest ticaret anlaşması		etnik semboller		meme kanseri		boks teknikleri		akıllı saat özellikleri		safari turları
E.		investment risks		famous temples		ebola symptoms		derby matches		social media		train trips
A.	3	مخاطر الاستثمار	3	معابد شهيرة	3	اعراض الايولا	3	مباريات الديربي	3	اعلام الاجتماعي	3	جولات القطار
T.		yatırım riskleri		ünlü tapınaklar		ebola belirtileri		derbi maçları		sosyal medya		tren gezileri
E.		job opportunities		old civilizations		heart angioplasty risks		international tennis tournament		three dimensional printers		travel tips
A.	4	فرص العمل	4	حضارات القديمة	4	مخاطر قسطرة القلب	4	بطولات تنس الدولية	4	طابعات ثلاثية الأبعاد	4	ارشادات السفر
T.		iş fırsatları		eski uygarlıklar		kalp anjiyoplasti riskleri		uluslararası tenis turnuvaları		üç boyutlu yazıcılar		seyahat ipuçları
E.		professional salesmanship		traditional clothes		vitamin c benefits		swimming styles		yosemite operating system		yacht voyages
A.	5	محترف فن البيع	5	ملابس التقليدية	5	فوائد فيتامين ج	5	انماط السباحة	5	نظام التشغيل يوسمايت	5	اسفار اليخت
T.		profesyonel pazarlamacılık		geleneksel giysiler		c vitamininin faydaları		yüzme stilleri		yosemite işletim sistemi		yat seferleri

Notes: “E.” stands for English; “A.” stands for Arabic; “T.” stands for Turkish

3.5 Query Run and Document Evaluation

The first query was run in English, in Arabic, and in Turkish languages one by one on each of the three Web search engines. At each run, the first 10 documents retrieved were evaluated. The same were done for the rest of the queries separately and, as a result, totally 270 (30 queries $\times$ 3 languages $\times$ 3 Web search engines) runs and retrieval output evaluations were done. During the evaluation of each retrieval output, every document retrieved in the first 10 was categorised as “relevant” or “non-relevant” based on the followings:

1. If the text content of a document was related to subject of the query, it was categorised as “relevant”.
2. If a document was repeated at least one time with different URL address –the URL addresses appeared in the address bars were considered after the documents were displayed–, all the documents (repeated-document and repeat-document(s)) were categorised as “relevant” or categorised as “non-relevant” based on the relation of the text content with the query subject (note that “SD” code will be used to represent a repeat-document in a such case).
3. If a document was repeated at least one time with the same URL address –the URL addresses appeared in the address bars were considered after the documents were displayed–, the repeated-document was categorised as “relevant” or “non-relevant” based on the relation of the text content with the query subject, while the repeat-document(s) was directly categorised as “non-relevant” (note that “RD” code will be used to represent a repeat-document in a such case).

4. The documents were categorised as “non-relevant” based on the followings:
- If a document’s text content was not related to subject of the query.
 - If a document’s text content was in different language other than the query language.
 - If a document included images or videos without any text description related to the query.
 - If a document required registration for accessing the content.
 - If a document was appeared with the message “Access Denied” or the message “509 Bandwidth limit exceeded, the server is temporarily unable to serve your request due to the site owner reaching his/her bandwidth limit”.
 - If a document was not accessible and one of the following messages was displayed: “This website is not available”; “No data received”; “URL unavailable, please contact the administration team if you think it’s an error”; “404 error”. Note that this situation was considered as a dead link situation and from now on, “DD” code will be used to represent such situation.

3.6 Precision Ratio

Precision ratio is one of the most widely used metric in order to evaluate the retrieval quality of information retrieval system [23], it used in this study and its formula that obtained from [24] displayed below.

$$\text{Precision Ratio} = \frac{\text{Number of Relevant Documents Retrieved}}{\text{Total Number of Documents Retrieved}} \times 100 \quad (1)$$

In each category, precision ratios were calculated for each of five queries at every Web search engine and language pair. For each Web search engine, mean precision ratios were calculated for each category and language pair by using the precision ratios of the five queries in the corresponding category. Additionally, average mean

precision ratios of the six categories for each Web search engine and language pair were calculated.

Note that in our study, for each query run, more than ten documents were retrieved. Since, the first ten documents retrieved were evaluated; therefore, the total number of documents retrieved was considered as 10 in every precision ratio calculations.

For example, in “business” category, the first 10 documents retrieved by Google for the first query “brand licensing process” in English language were evaluated as shown in Figure 3.1.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
<i>document</i>	<i>document</i>	<i>document</i>	<i>document</i>	<i>document</i>	<i>document</i>	<i>document</i>	<i>document</i>	<i>document</i>	<i>document</i>
+	+	+	-	+	-	+	-	+	+

Figure 3.1: Evaluation of the first 10 documents retrieved for the query “brand licensing process”

The first cell from the left side represents the first document retrieved, the second cell represents the second document retrieved and so on. The “relevant” documents are represented by “+” and the “non-relevant” documents are represented by “-”. After run the query “brand licensing process”, seven documents retrieved were “relevant” (“+”) and the rest were “non-relevant” (“-”). Since the number of “+” which are 7 divided by total number of documents retrieved which are 10, and all multiplied by 100, so precision ratio equals to 70%.

$$Precision\ Ratio = \frac{7}{10} \times 100 = 70\%$$

3.7 Normalized Recall Ratio

Beside precision ratio, normalized recall ratio is also used in this study. The normalized recall ratio is a metric which evaluates the information retrieval system based on performance of showing relevant documents retrieved in higher ranks. The formula of normalized recall which obtained from [25] is presented below.

$$R_{norm}(\Delta) = \frac{1}{2} \left(1 + \frac{R^+ - R^-}{R_{max}^+} \right) \quad (2)$$

- R^+ is the number of document pairs that “relevant” documents come before "non-relevant" documents.
- R^- is the number of document pairs that “non-relevant” documents come before "relevant" documents.
- R_{max}^+ is the number of maximal R^+ .

Normalized recall ratio is the value of normalized recall that multiplied by 100. Like precision ratio, in each category, normalized recall ratios were calculated for each of five queries at every Web search engine and language pair. For each Web search engine, mean normalized recall ratios were calculated for each category and language pair by using the normalized recall ratios of the five queries in the corresponding category. In addition, average mean normalized recall ratios of the six categories for each Web search engine and language pair were calculated.

For instance, in “business” category, the first 10 documents retrieved by Google for the first query “brand licensing process” in English language were evaluated as shown in Figure 3.1. Unlike precision ratio, the rankings of documents retrieved are important in normalized recall ratio. First of all, the number of document pairs that

Then, the number of document pairs that “non-relevant” documents come before “relevant” documents were found out as shown in Figure 3.3. It can be seen that R^- is 9.

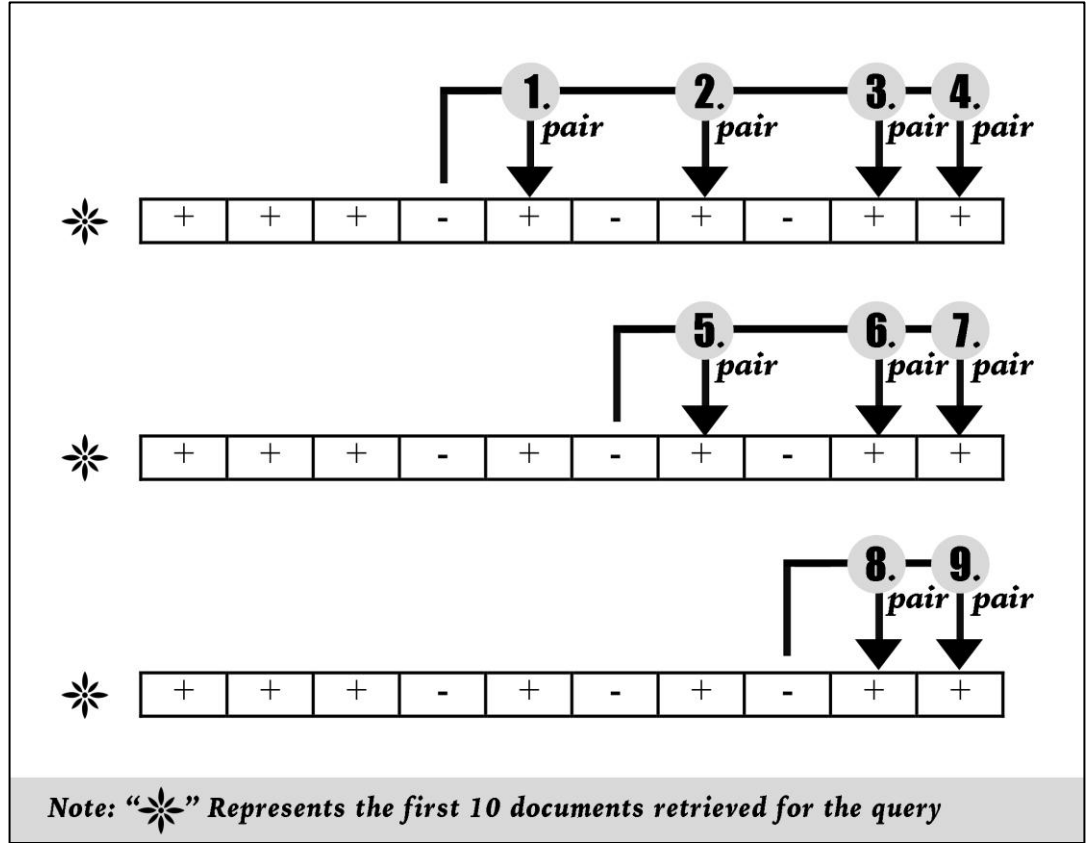


Figure 3.3: Finding out R^-

Afterwards, to find out $R_{\max}^+$: Firstly, documents’ positions were rearranging in order to have the ideal retrieval output (all “relevant” documents (“+”) come before “non-relevant” documents (“-”)) for the existing retrieval output of the query “brand licensing process” as shown in Figure 3.4; Secondly, R^+ was found out for the ideal retrieval output for $R_{\max}^+$. $R_{\max}^+$ is 21. While $R^+ = 12$, $R^- = 9$, and $R_{\max}^+ = 21$,

$$\text{Normalized recall ratio} = \frac{1}{2} \left(1 + \frac{12 - 9}{21} \right) \times 100 = 57.14\%$$

The Retrieval Output

+	+	+	-	+	-	+	-	+	+
---	---	---	---	---	---	---	---	---	---



The ideal Retrieval Output

+	+	+	+	+	+	+	-	-	-
---	---	---	---	---	---	---	---	---	---

Figure 3.4: The ideal retrieval output for “brand licensing process” query’s retrieval output

Chapter 4

EXPERIMENTAL RESULTS

When considering all query runs, 2700 documents were retrieved in total. Twenty two of them were DDs, 20 of them were RDs, and 33 of them were SDs. Excluding DDs, 2678 documents were examined and 1840 of them were categorised as “relevant”.

4.1 DDs, RDs, and SDs

For each Web search engine, Table 2 shows the total number of DDs, RDs, and SDs retrieved from 5 queries per a category for each language.

For all three languages, Google retrieved the lowest total number of DDs (5), RDs (2), and the SDs (5) among the engines. Contrary to Google, Yahoo retrieved the highest total number of DDs (9) and RDs (11) for all three languages, while both Yahoo and Bing retrieved the same total number of SDs (14) for all three languages.

Every Web search engine retrieved at least total one DD for all queries per language. Google retrieved total zero number of RDs and SDs for all queries in English and Turkish languages, respectively. On the other hand, Google retrieved total one RD for all queries in both Arabic and Turkish languages and at least total two SDs for all queries in both English and Arabic languages. Both Bing and Yahoo retrieved at least total one RD and at least total one SD for all queries in each language.

When all Web search engines are considered, the lowest category-based total numbers of DDs, RDs, and SDs for all queries in all three languages were 1 in “culture”, “health”, and “technology” categories, 1 in “business”, and “travel” categories, and 2 in “technology” category, respectively, while the highest category-based total numbers of DDs, RDs, and SDs for all queries in all three languages were 9 in “business”, 7 in “technology”, and 11 in “travel”, respectively.

Table 2: Number of DDs, RDs, and the SDs

		# of DDs								# of RDs								# of SDs							
W.S.E.	Language	Business	Culture	Health	Sports	Technology	Travel	Total	W.S.E. Total	Business	Culture	Health	Sports	Technology	Travel	Total	W.S.E. Total	Business	Culture	Health	Sports	Technology	Travel	Total	W.S.E. Total
Google	E.	0	0	0	1	0	0	1	5	0	0	0	0	0	0	0	2	0	0	0	0	0	2	2	5
	A.	1	0	0	1	0	1	3		0	1	0	0	0	0	1		1	0	0	0	0	2	3	
	T.	0	0	0	0	0	1	1		0	0	0	1	0	0	1		0	0	0	0	0	0	0	
Bing	E.	1	0	0	0	0	0	1	8	0	0	2	0	1	0	3	7	0	1	1	0	1	1	4	14
	A.	1	0	1	0	0	2	4		0	1	1	0	1	0	3		0	0	0	1	0	1	2	
	T.	2	0	0	0	0	1	3		0	0	0	1	0	0	1		0	3	1	0	1	3	8	
Yahoo	E.	1	0	0	1	0	0	2	9	0	0	1	0	4	1	6	11	1	1	1	0	0	1	4	14
	A.	1	1	0	0	1	2	5		0	0	0	0	1	0	1		0	0	0	2	0	0	2	
	T.	2	0	0	0	0	0	2		1	0	0	3	0	0	4		1	5	1	0	0	1	8	
C. Total		9	1	1	3	1	7			1	2	4	5	7	1			3	10	4	3	2	11		

Notes: “W. S. E.” stands for Web Search Engine; “C.” stands for Category; “E.” stands for English; “A.” stands for Arabic; “T.” stands for Turkish.

4.2 Relevant Documents

As Table 3 shows, in all categories, Google retrieved at least one “relevant” document for every query and language pair. Bing retrieved zero “relevant” documents only for query 5 of the “travel” category in Arabic language. Yahoo retrieved zero “relevant” documents for queries 5, 2, and 4 of “business”, “culture”, and “travel” categories, respectively in Turkish language, and for queries 2 and 5 of “sports” and “travel” categories, respectively in Arabic language.

Google retrieved 10 “relevant” documents for one or more queries in every category, whereas Bing and Yahoo have the same except in “culture” and “technology” categories, respectively.

Only Yahoo retrieved the same number of “relevant” documents in all languages for the fourth query of “sports” category and fifth query of “technology” category.

When each language-based-totals are considered, it is seen that Bing has the highest total number of “relevant” documents (45) in “health” category for English language. In contrast, Yahoo has the lowest total number of “relevant” documents (16) in “business” category for Turkish language.

Table 3: Number of “relevant” Documents Retrieved

Search Engine	Language	Business						Culture						Health						Sports						Technology						Travel					
		Query 1	Query 2	Query 3	Query 4	Query 5	Total	Query 1	Query 2	Query 3	Query 4	Query 5	Total	Query 1	Query 2	Query 3	Query 4	Query 5	Total	Query 1	Query 2	Query 3	Query 4	Query 5	Total	Query 1	Query 2	Query 3	Query 4	Query 5	Total	Query 1	Query 2	Query 3	Query 4	Query 5	Total
Google	E.	7	10	10	5	5	37	10	6	10	10	8	44	10	7	9	8	8	42	10	7	4	8	7	36	10	9	5	8	10	42	10	10	10	7	3	40
	A.	2	9	8	4	3	26	9	6	2	10	5	32	9	9	10	5	9	42	8	1	8	6	3	26	9	7	4	8	9	37	5	8	5	6	2	26
	T.	6	9	7	5	5	32	6	1	10	7	9	33	8	9	10	6	8	41	9	4	7	7	6	33	9	6	1	9	10	35	6	10	8	10	9	43
Bing	E.	7	10	9	5	8	39	9	3	8	8	9	37	10	8	9	9	9	45	10	10	7	8	9	44	10	9	4	10	10	43	10	8	10	4	4	36
	A.	1	6	8	8	4	27	4	1	4	9	3	21	10	8	10	8	8	44	8	2	10	9	3	32	8	8	5	10	8	39	8	10	7	8	0	33
	T.	2	9	6	3	2	22	8	1	8	7	8	32	4	7	9	6	9	35	7	8	8	8	8	39	9	8	3	8	9	37	2	9	7	3	10	31
Yahoo	E.	8	10	10	6	4	38	10	4	10	9	9	42	8	8	6	9	10	41	10	10	6	8	9	43	7	9	6	8	9	39	7	6	10	5	4	32
	A.	1	7	7	7	5	27	5	1	3	9	4	22	9	7	10	6	9	41	6	0	4	8	3	21	9	7	3	9	9	37	7	9	4	6	0	26
	T.	4	6	4	2	0	16	3	0	4	5	7	19	4	5	7	4	8	28	5	3	4	8	9	29	9	8	2	8	9	36	1	9	8	0	7	25

Notes: “E.” stands for English; “A.” stands for Arabic; “T.” stands for Turkish

4.3 Mean Precision Ratio Results

Figure 4.1 demonstrates mean precision ratios of Google for every category and language pair. In “travel” category, Google presented the best performance in Turkish language, in “health” category, it displayed the best performance in both English and Arabic languages, while it presented the best performance in English language for the rest of the categories. Besides, in “health” and “technology” categories, Google presented the lowest performance in Turkish language, while for the rest of the categories, it presented the lowest performance in Arabic language. Google’s mean precision ratios range for all languages was between 52.00% and 88.00%. The best mean precision ratio (88.00%) was shown in “culture” category in English language and the lowest mean precision ratio (52.00%) was shown in both “business” and “travel” categories in Arabic language.

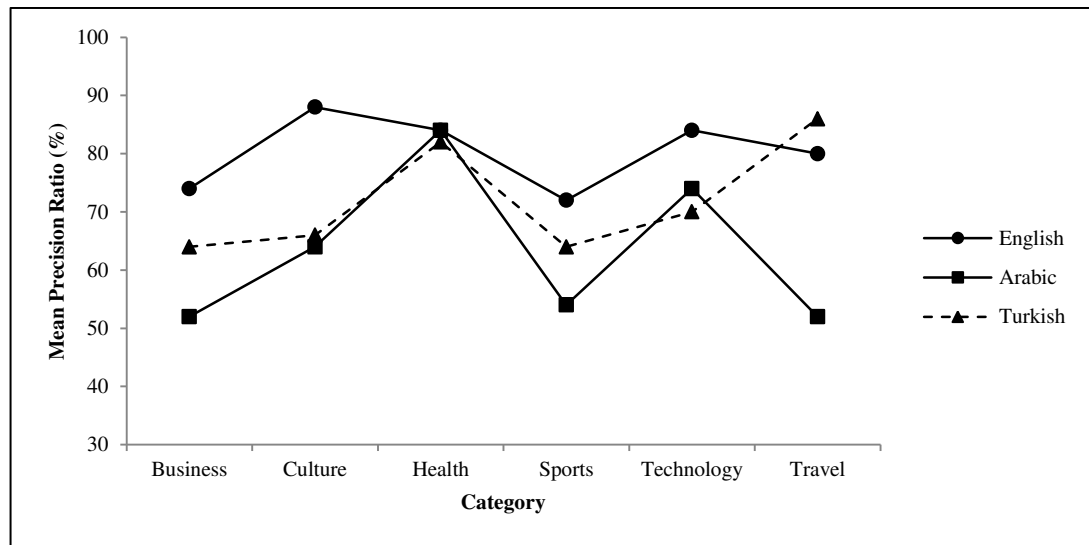


Figure 4.1: Mean precision ratios of Google Web search engine

Figure 4.2 clarifies mean precision ratios of Bing for every category and language pair. In all categories, Bing showed the best performance in English language among all languages. In “culture” and “sports” categories, Bing showed the lowest performance in Arabic language, while for the rest of the categories, it showed the lowest performance in Turkish language. Bing’s mean precision ratios range for all languages was between 42.00% and 90.00%. The best mean precision ratio (90.00%) was shown in “health” category in English language and the lowest mean precision ratio (42.00%) was shown in “culture” category in Arabic language.

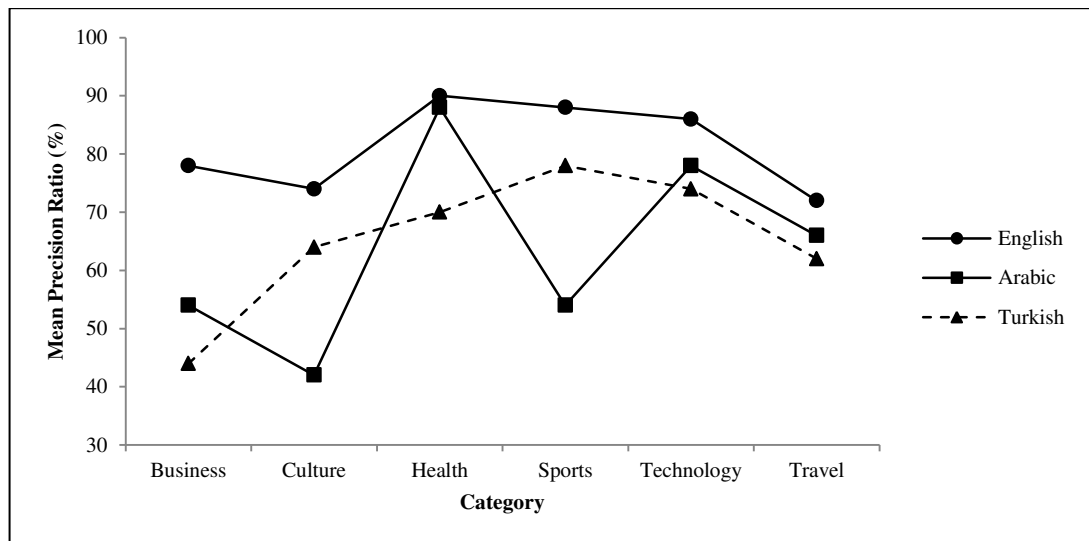


Figure 4.2: Mean precision ratios of Bing Web search engine

Figure 4.3 illustrates mean precision ratios of Yahoo for every category and language pair. In “health” category, Yahoo displayed the best performance in both English and Arabic languages, while it displayed the best performance in English language for the rest of the categories. Excluding “sports” category, in all categories, Yahoo has the lowest performance in Turkish language. In “sports” category, the lowest performance was displayed in Arabic language. Yahoo’s mean precision ratios range for all languages was between 32.00% and 86.00%. The best mean precision ratio (86.00%) was shown in “sports” category in English language and the lowest mean precision ratio (32.00%) was shown in “business” category in Turkish language.

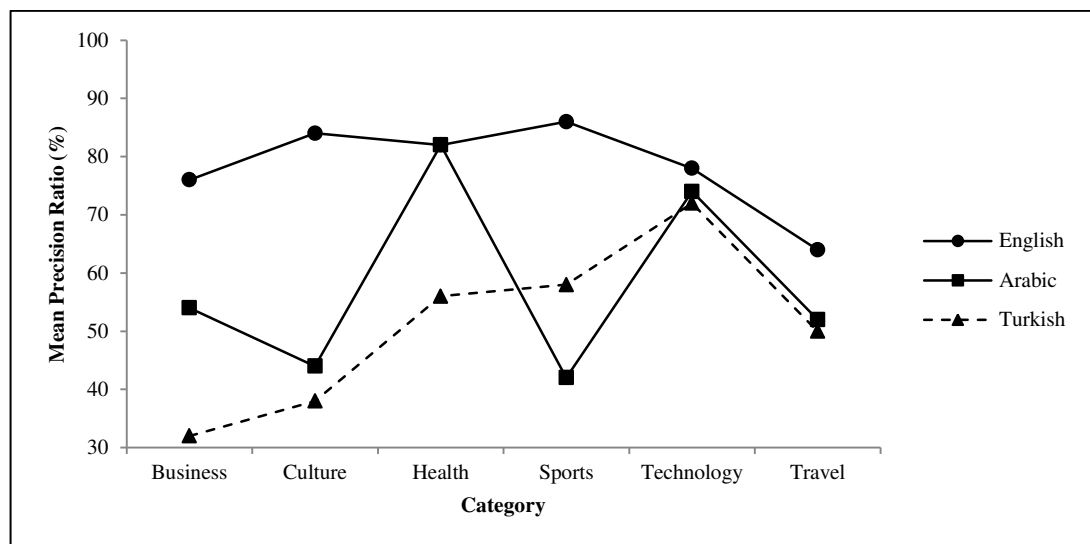


Figure 4.3: Mean precision ratios of Yahoo Web search engine

Figure 4.4 demonstrates mean precision ratios of Google, Bing, and Yahoo in English language for all categories. In all categories except “culture” and “travel”, Bing presented the best performance. In both “culture” and “travel” categories, Google presented the best performance. In both “business” and “sports” categories, Google presented the worst performance. In “health”, “technology”, and “travel” categories, Yahoo showed the worst performance. In “culture” category, Bing has the worst performance. The mean precision ratios range for English language was between 64.00% (displayed by Yahoo in “travel” category) and 90.00% (displayed by Bing in “health” category).

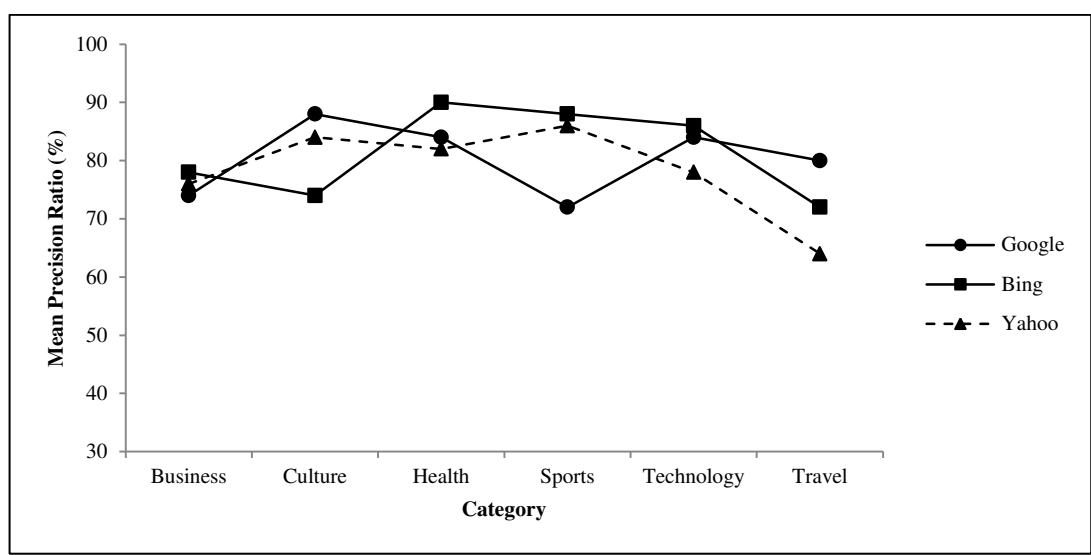


Figure 4.4: Mean precision ratios for English language

Figure 4.5 clarifies mean precision ratios of Google, Bing, and Yahoo in Arabic language for all categories. In “business” category, the best performance was shown by both Bing and Yahoo, in “sports” category, by both Google and Bing, and in “culture” category, by Google, while in the rest of the categories, Bing showed the best performance. In “culture” and “health” categories, Bing and Yahoo displayed the worst performance, respectively. In both “technology” and “travel” categories, Google and Yahoo have shared the worst performance. The mean precision ratios range for Arabic language was between 42.00% (displayed by Bing in “culture” category and displayed by Yahoo in “sports” category) and 88.00% (displayed by Bing in “health” category).

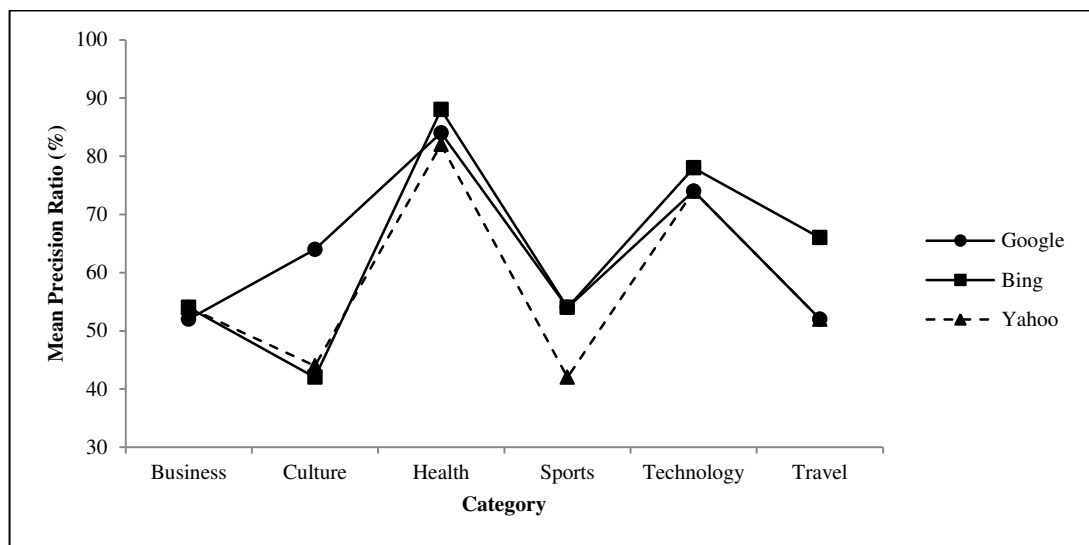


Figure 4.5: Mean precision ratios for Arabic language

Figure 4.6 illustrates mean precision ratios of Google, Bing, and Yahoo in Turkish language for all categories. In all categories except “sports” and “technology”, Google displayed the best performance. In both “sports” and “technology” categories, Bing displayed the best performance. In all categories except “technology” which Google has the worst at it, Yahoo has the worst performance. The mean precision ratios range for Turkish language was between 32.00% (displayed by Yahoo in “business” category) and 86.00% (displayed by Google in “travel” category).

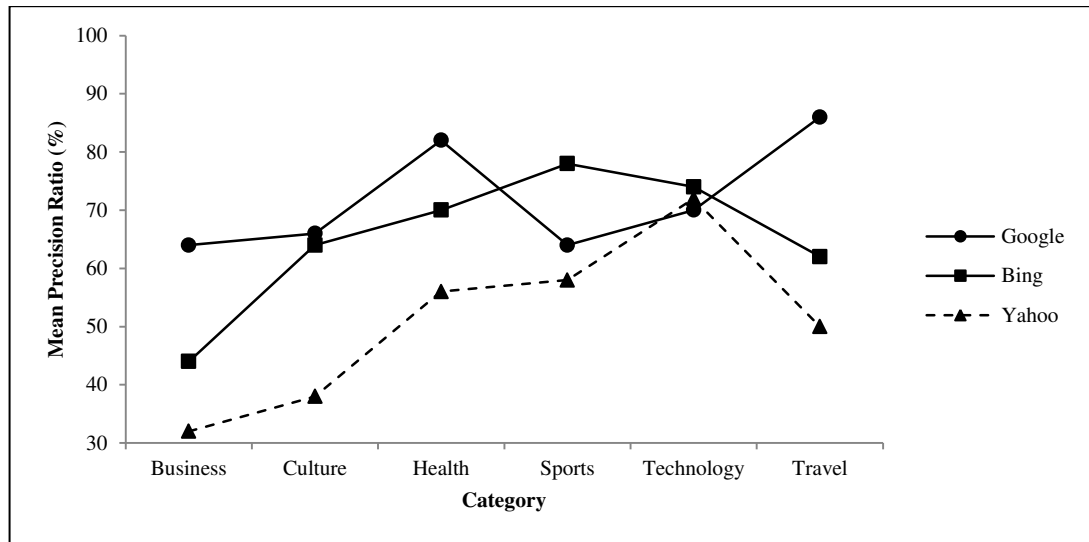


Figure 4.6: Mean precision ratios for Turkish language

Figure 4.7 demonstrates average mean precision ratios based on English, Arabic, and Turkish languages for Google, Bing, and Yahoo. All three Web search engines presented their highest performances in English language. Google's performance in Turkish language was better than its performance in Arabic language. Like Google, Bing acted in the same manner. In opposite of these, the performance of Yahoo in Arabic language was better than its performance in Turkish language. Bing's performance was better than Google's and Yahoo's performances in both English and Arabic languages, while Google's performance was better than the others' performances in Turkish language.

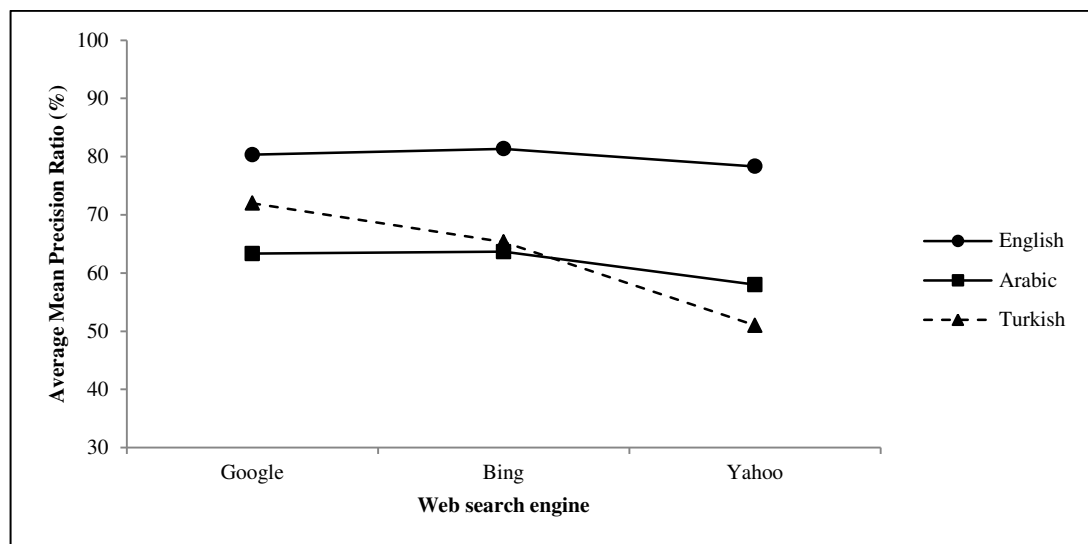


Figure 4.7: Average mean precision ratios based on language for Web search engines

4.4 Mean Normalized Recall Ratio Results

Figure 4.8 demonstrates mean normalized recall ratios of Google for every category and language pair. In “health” category, Google presented the best performance in Turkish language, while it presented the best performance in English language for the rest of the categories. Besides, in “sports” category, Google presented the lowest performance in Turkish language, while for the rest of the categories, it presented the lowest performance in Arabic language. Google’s mean normalized recall ratios range for all languages was between 41.85% and 87.62%. The best mean normalized recall ratio (87.62%) was shown in “travel” category in English language and the lowest mean normalized recall ratio (41.85%) was shown in “business” category in Arabic language.

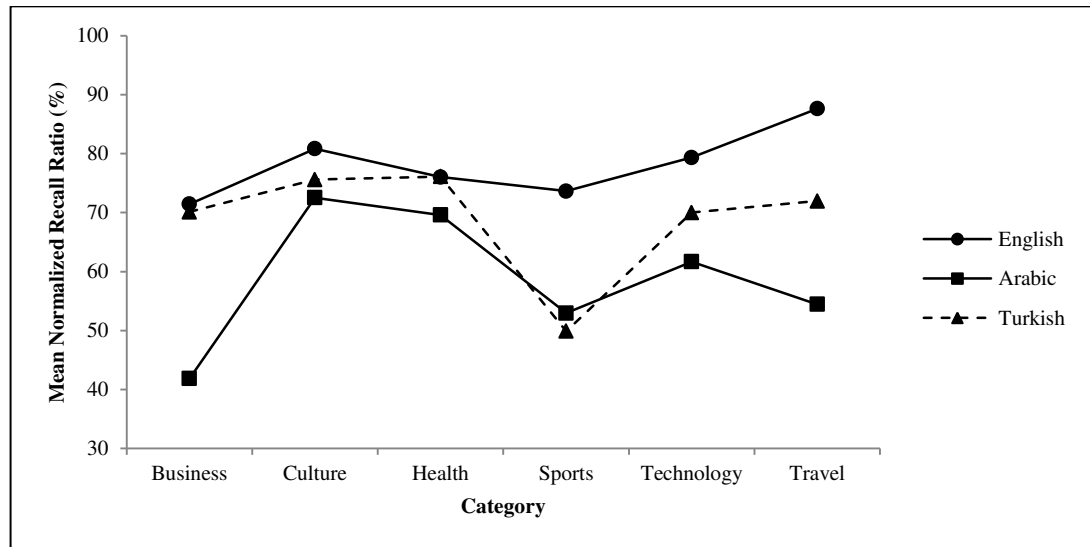


Figure 4.8: Mean normalized recall ratios of Google Web search engine

Figure 4.9 clarifies mean normalized recall ratios of Bing for every category and language pair. In “health” category, Bing presented the best performance in Arabic language, in “travel” category, it displayed the best performance in Turkish languages, while it showed the best performance in English language for the rest of the categories. Further, in both “sports” and “travel” categories, Bing showed the lowest performance in Arabic language, while for the rest of the categories, it showed the lowest performance in Turkish language. Bing’s mean normalized recall ratios range for all languages was between 45.34% and 83.25%. The best mean normalized recall ratio (83.25%) was shown in “sports” category in English language and the lowest mean normalized recall ratio (45.34%) was shown in “business” category in Turkish language.

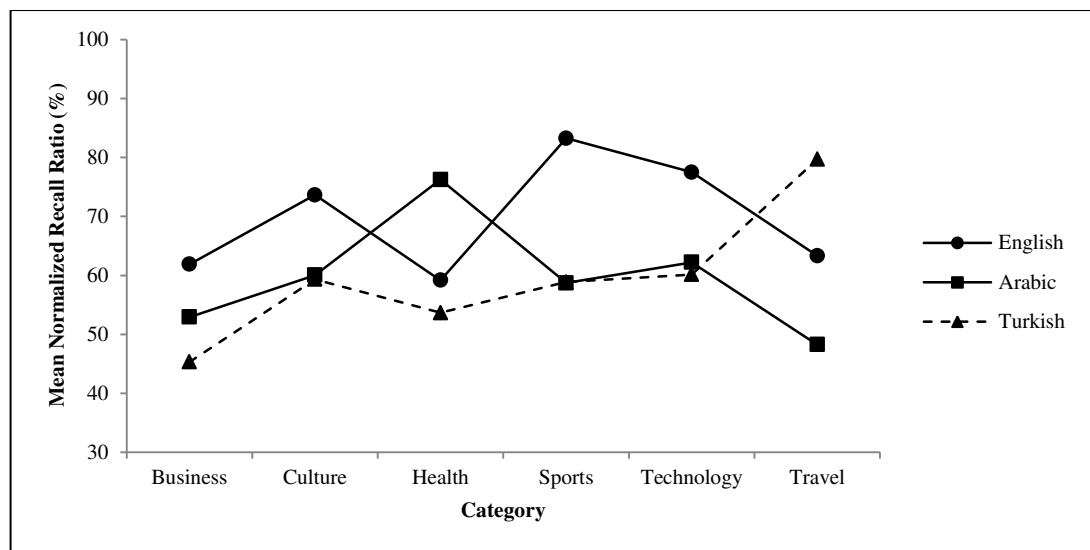


Figure 4.9: Mean normalized recall ratios of Bing Web search engine

Figure 4.10 illustrates mean normalized recall ratios of Yahoo for every category and language pair. In all categories, Yahoo displayed the best performance in English language. Excluding “sports” category, in all categories, Yahoo has the lowest performance in Turkish language. In “sports” category, the lowest performance was displayed in Arabic language. Yahoo’s mean normalized recall ratios range for all languages was between 37.78% and 90.83%. The best mean normalized recall ratio (90.83%) was shown in “health” category in English language and the lowest mean normalized recall ratio (37.78%) was shown in “travel” category in Turkish language.

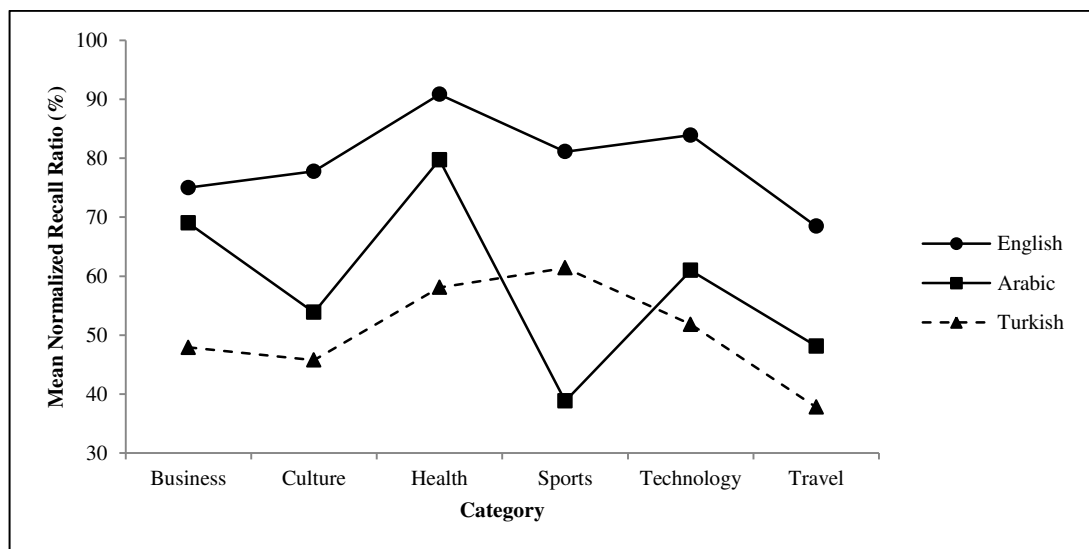


Figure 4.10: Mean normalized recall ratios of Yahoo Web search engine

Figure 4.11 demonstrates mean normalized recall ratios of Google, Bing, and Yahoo in English language for all categories. In “business”, “health”, and “technology” categories, Yahoo presented the best performance. In both “culture” and “travel” categories, Google presented the best performance. In “sports” category, Bing presented the best performance. While Google showed the worst performance in “sports” category, in the rest of the categories, Bing showed the worst performance. The mean normalized recall ratios range for English language was between 59.17% (displayed by Bing in “health” category) and 90.83% (displayed by Yahoo in “health” category).

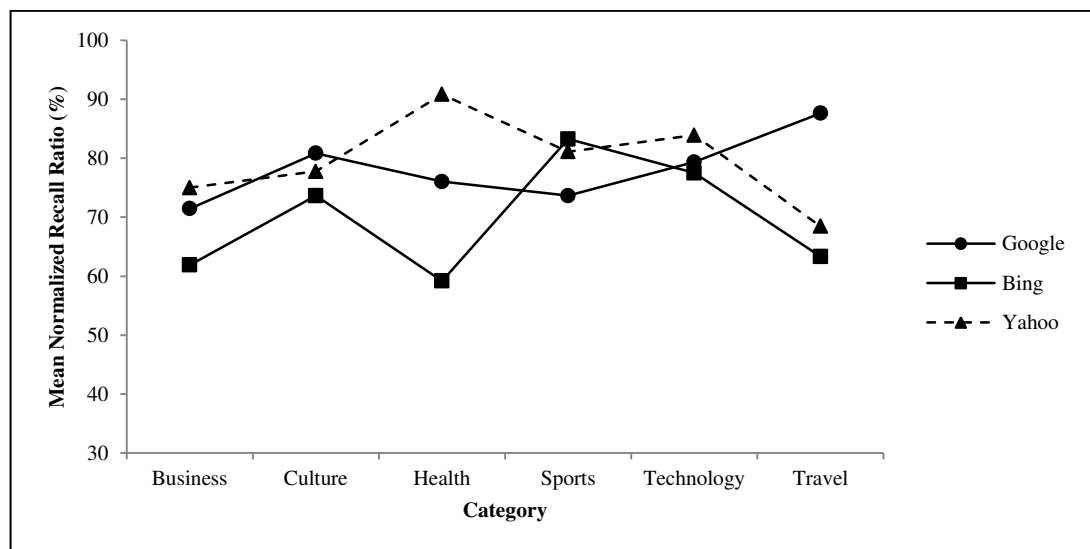


Figure 4.11: Mean normalized recall ratios for English language

Figure 4.12 clarifies mean normalized recall ratios of Google, Bing, and Yahoo in Arabic language for all categories. In both “business” and “health” categories, Yahoo showed the best performance. In both “culture” and “travel” categories, Google showed the best performance. In both “sports” and “technology” categories, Bing showed the best performance. In the rest of the categories, Bing showed the best performance. While Google displayed the worst performance in both “business” and “health” categories, Yahoo displayed the worst performance in the rest of the categories. The mean normalized recall ratios range for Arabic language was between 38.81% (displayed by Yahoo in “sports” category) and 79.72% (displayed by Yahoo in “health” category).

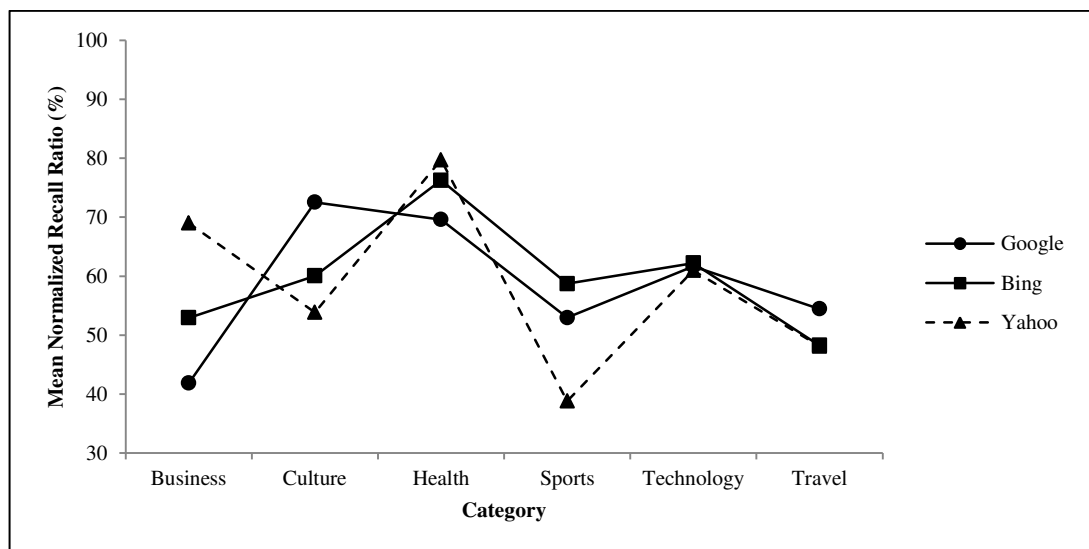


Figure 4.12: Mean normalized recall ratios for Arabic language

Figure 4.13 illustrates mean normalized recall ratios of Google, Bing, and Yahoo in Turkish language for all categories. In all categories except “sports” and “travel”, Google displayed the best performance. In “sports” and “travel” categories, the best performance displayed by Yahoo and Bing, respectively. In “business” and “health” categories, Bing has the worst performance. In “sports” category, Google has the worst performance. In the rest of the categories, Yahoo has the worst performance. The mean normalized recall ratios range for Turkish language was between 37.78% (displayed by Yahoo in “travel” category) and 79.68% (displayed by Bing in “travel” category).

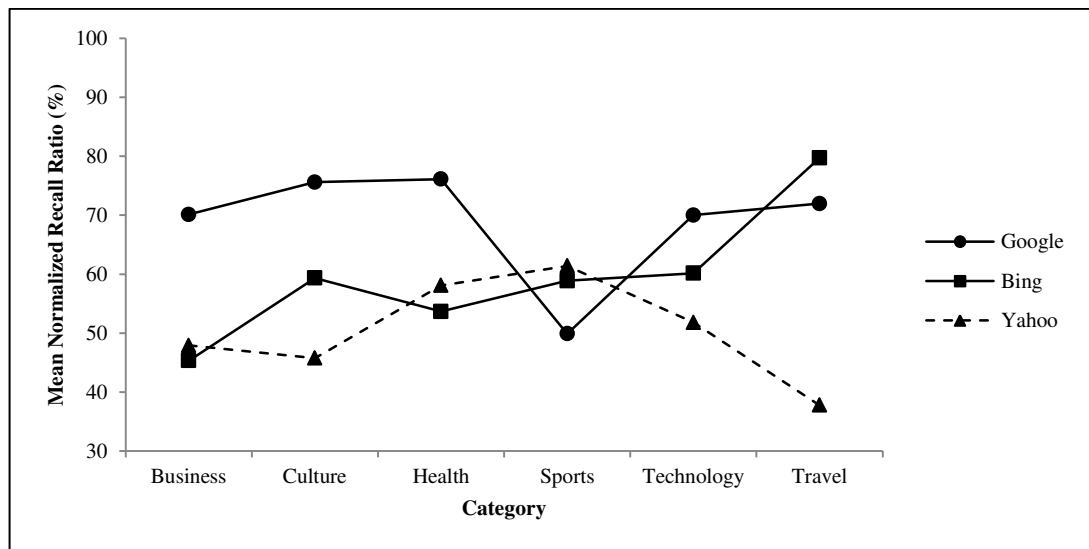


Figure 4.13: Mean normalized recall ratios for Turkish language

Figure 4.14 demonstrates average mean normalized recall ratios based on English, Arabic, and Turkish languages for Google, Bing, and Yahoo. All three Web search engines presented their highest performances in English language. Bing's performance in Arabic language was better than its performance in Turkish language. Like Bing, Yahoo acted in the same manner. In opposite of these, the performance of Google in Turkish language was better than its performance in Arabic language. Each Web search engine displayed better performance than the others' performance in certain language. For instance, Google's performance was better than the others' performances in Turkish language. Bing's performance was better than the others' performances in Arabic language. Yahoo's performance was better than the others' performances in English language.

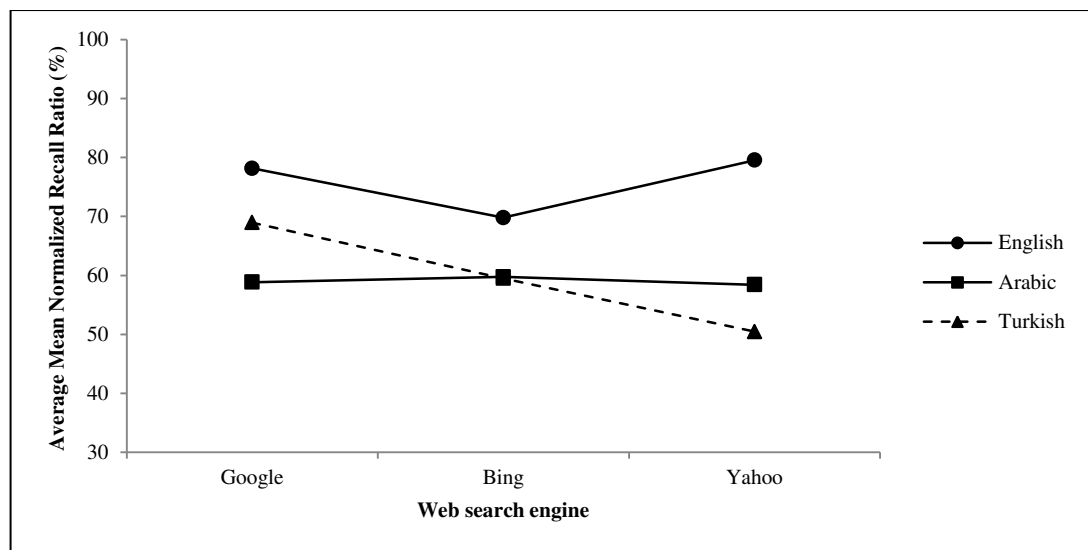


Figure 4.14: Average mean normalized recall ratios based on language for Web search engines

4.5 Summary and Discussion

The highest category-based total number of DDs was in “business” category, RDs in “technology” category, and SDs in “travel” category, while the lowest category-based total number of DDs was in “culture”, “health”, and “technology” categories, RDs was in “business” and “travel” categories, and SDs in “technology” category.

The highest total number of “relevant” documents was retrieved by Bing in “health” category for English language. In contrast, the lowest total number of “relevant” documents was retrieved by Yahoo in “business” category for Turkish language.

When number of relevant documents retrieved (mean precision ratio) is considered:

1. In “health” category, both Google and Yahoo displayed their best performances in both English and Arabic languages, while Bing showed its best performance only in English language. In “travel” category, Google showed its best performance in Turkish language, while both Bing and Yahoo showed their best performances in English language. In the rest of the categories, all three Web search engines showed their best performances in English language.
2. When the languages are considered:
 - For English language:
 - Google presented the highest performance in both “culture” and “travel” categories,
 - Bing presented the highest performance in the rest of the categories.
 - For Arabic language:
 - Both Bing and Yahoo shared the highest performance in “business” category,

- Both Google and Bing shared the highest performance in “sports” category,
- Google presented the highest performance in “culture” category,
- Bing presented the highest performance in the rest of the categories.
- For Turkish language:
 - Bing presented the highest performance in both “sports” and “technology” categories,
 - Google presented the highest performance in the rest of the categories.

Note that, Tables 4, 5, and 6 represent which Web search engine(s) presented the highest performance(s) based on mean precision ratio in each category for each language.

3. The greatest mean precision ratio (90.00%) reached by Bing for “health” category in English language. On the other hand, the least mean precision ratio (32.00%) reached by Yahoo for “business” category in Turkish language.

When displaying relevant documents retrieved in higher ranks (mean normalized recall ratio) is considered:

1. In “health” category, Google, Bing, and Yahoo displayed their best performance in Turkish language, Arabic language, and English language, respectively. In “travel” category, both Google and Yahoo displayed their best performance in English language, while Bing showed its best performance in Turkish language. In the rest of the categories, all three Web search engines showed their best performances in English language.
2. When the languages are considered:

- For English language:
 - Google presented the highest performance in both “culture” and “travel” categories,
 - Bing presented the highest performance only in “sports” category,
 - Yahoo presented the highest performance in the rest of the categories.
- For Arabic language:
 - Google presented the highest performance in both “culture” and “travel” categories,
 - Bing presented the highest performance in both “sports” and “technology” categories,
 - Yahoo presented the highest performance in the rest of the categories.
- For Turkish language:
 - Google presented the highest performance in all categories except “sports” and “travel”,
 - Yahoo presented the highest performance in “sports” category, and in “travel” category,
 - Bing showed the highest performance.

Note that, Tables 4, 5, and 6 represent which Web search engine(s) presented the highest performance(s) based on mean normalized recall ratio in each category for each language.

3. The highest mean normalized recall ratios (90.83%) reached by Yahoo for “health” category in English language. In the other side, the lowest mean normalized recall ratio (37.78%) reached by Yahoo for “travel” category in Turkish language.

Table 4: The best Web search engine(s) in each category for English language

Category Web Search engine	Business	Culture	Health	Sports	Technology	Travel
Google		✕✕				✕✕
Bing	✕		✕	✕✕	✕	
Yahoo	✕		✕		✕	

Notes: “✕” represents the best based on mean precision ratio; “✕” represents the best based on normalized recall ratio.

Table 5: The best Web search engine(s) in each category for Arabic language

Category Web Search engine	Business	Culture	Health	Sports	Technology	Travel
Google		✕✕		✕		✕
Bing	✕		✕	✕✕	✕✕	✕
Yahoo	✕✕		✕			

Notes: “✕” represents the best based on mean precision ratio; “✕” represents the best based on normalized recall ratio.

Table 6: The best Web search engine(s) in each category for Turkish language

Category Web Search engine	Business	Culture	Health	Sports	Technology	Travel
Google	✕✕	✕✕	✕✕		✕	✕
Bing				✕	✕	✕
Yahoo				✕		

Notes: “✕” represents the best based on mean precision ratio; “✕” represents the best based on normalized recall ratio.

When considering both the average mean precision ratios based on all the languages for Google, Bing, and Yahoo and the average mean normalized recall ratios based on all the languages for Google, Bing, and Yahoo:

- All the three Web search engines presented their highest performances in English language.

- The performance of Google in Turkish language was better than it's performances in Arabic language.
- The performance of Yahoo in Arabic language was better than it's performances in Turkish language.
- When Arabic language is considered, Bing's performance was better than the performances of both Google and Yahoo.
- When Turkish language is considered, Google's performance was better than the performances of both Bing and Yahoo.
- While the performance of Bing in Turkish language was better than it's performances in Arabic language based on the average mean precision ratios, Bing's performance in Arabic language was better than it's performance in Turkish language based on the average mean normalized recall ratios.
- When English language is considered, the performance of Bing was better than the performances of both Google and Yahoo based on the average mean precision ratios, while Yahoo's performance was better than the performances of both Google and Bing based on the average mean normalized recall ratios.

Chapter 5

CONCLUSION

The performance evaluation of the three international Web search engines Google, Bing, and Yahoo in the three languages English, Arabic, and Turkish in “business”, “culture”, “health”, “sports”, “technology”, and “travel” categories were studied.

For all categories, generally, the Web search engines performed better performances in retrieving English documents as well as showing them in higher ranks. Nevertheless, we believe that international Web search engines still require amelioration in order to retrieve more “relevant” documents and also showing “relevant” documents in higher ranks (especially in Arabic and Turkish languages) in various categories for increasing the user satisfaction on retrieval effectiveness.

REFERENCES

- [1] Gao W., Zhu Y., Jia Z., Luo C., Wang L., Li Z., Zhan J., Qi Y., He Y., Gong S., Li X., Zhang S., & Qiu B. (2013). BigDataBench: a Big Data Benchmark Suite from Web Search Engines, *The Third Workshop on Architectures and Systems for Big Data in conjunction with The 40th International Symposium on Computer Architecture*, arXiv:1307.0320v1 [cs.IR] (Accessed on January 4, 2015).
- [2] Grosjean, F. (2010). Bilingualism's Best Kept Secret, Life as a Bilingual, <http://www.psychologytoday.com/blog/life-bilingual/201011/bilingualisms-best-kept-secret> (Accessed on January 04, 2015).
- [3] Internet Live Stats, Internet users in the world as of July 2014, <http://www.internetlivestats.com/internet-users/#trend> (Accessed on January 4, 2015).
- [4] Google Inc., Google, <http://www.google.com>.
- [5] Microsoft, Bing, <http://www.bing.com>.
- [6] Yahoo! Inc., Yahoo! Search, <http://search.yahoo.com>.
- [7] Efthimiadis, E. N., Malevris, N., Kousaridas, A., Lepeniotou, A., & Loutas, N. (2009). Non-english Web search: an evaluation of indexing and searching the Greek Web, *Information Retrieval Journal*, Springer-Netherlands, Vol.12, No.3, 352-379.

- [8] Çakır, E., Bahçeci, H., & Bitirim, Y. (2008). An Evaluation of Major Image Search Engines on Various Query Topics, *3rd IEEE Conf. on Internet Monitoring and Protection*, 161-165.
- [9] Goel, Sh., & Yadav, S. (2013). Search Engine Evaluation Based on Page Level Keywords, *The 3rd IEEE International Advanced Computing Conference (IACC)*, 870 – 876.
- [10] Moukdad, H. (2004). Lost in cyberspace: How Do Search Engines Handle Arabic Queries?, *In proceedings of the 32nd annual conference of the Canadian Association for Information Science*, Winnipeg, 1-7. www.caisacsi.ca/proceedings/2004/moukdad_2004.pdf (Accessed on January 02, 2015).
- [11] Namk, M. Y. J., & Bitirim, Y. (2015). An Evaluation of International Web Search Engines' Performances Based on English, Arabic, and Turkish Languages in Various Categories, *International Journal of Scientific Knowledge, Computing and Information Technology*, Vol. 5, No. 8. 14-20.
- [12] Wikimedia Foundation, Inc., Web search engine, http://en.wikipedia.org/wiki/Web_search_engine (Accessed on December 27, 2014).
- [13] Wikimedia Foundation, Inc., Yahoo!, <http://en.wikipedia.org/wiki/Yahoo!> (Accessed on December 27, 2014).

[14] Google Inc., Our history in depth, <http://www.google.com/about/company/history/> (Accessed on December 27, 2014).

[15] Wikimedia Foundation, Inc., Bing, <http://en.wikipedia.org/wiki/Bing> (Accessed on December 27, 2014).

[16] eBizMBA Inc., Top 15 Most Popular Search Engines as of January 2015, <http://www.ebizmba.com/articles/search-engines> (Accessed on January 02, 2015).

[17] Q-Success, W3Techs-Web technology surveys, Usage of content languages for websites as of January 2015. http://w3techs.com/technologies/overview/content_language/all (Accessed on January 02, 2015).

[18] Lewis, M. Paul, Gary F. Simons, & Charles D. Fennig (eds.). (2014). *Ethnologue: Languages of the World, Seventeenth edition*. Dallas, Texas: SIL International. <http://www.ethnologue.com/statistics/size> (Accessed on January 5, 2015).

[19] Cable News Network. Turner Broadcasting System, Inc., CNN, <http://edition.cnn.com>.

[20] euronews, <http://www.euronews.com>.

[21] FOX News Network, LLC., FOX News, <http://www.foxnews.com>.

[22] TheHuffingtonPost.com, Inc., The Huffington post,
<http://www.huffingtonpost.com>.

[23] Yates, R., & Neto, B. (2011). *Modern Information Retrieval*, The Concepts and Technology behind Search, 2nd Edition, ACM Press Books.

[24] Kent, A., Lancour, H., & Daily, J. E. (1978). *Encyclopedia of Library and Information Science*, Libraries and Information Centers in to Printers and Printing (Library and Information Science Encyclopedia), CRC Press, Vol. 23 - Poland, 170-180.

[25] Bollmann, P., Jochum, R., Reiner, U., Weissmann, V., & Zuse, H. (1986). *Planung und durchführung der retrieval tests*. In Schneider, H. (Ed.), *Leistungsbewertung von Information Retrieval Verfahren (LIVE)*, 183-212.

APPENDICES

Appendix A: Mean Precision Ratios

Google Web search engine				Bing Web search engine				Yahoo Web search engine			
Category	English	Arabic	Turkish	Category	English	Arabic	Turkish	Category	English	Arabic	Turkish
Business	74.00	52.00	64.00	Business	78.00	54.00	44.00	Business	76.00	54.00	32.00
Culture	88.00	64.00	66.00	Culture	74.00	42.00	64.00	Culture	84.00	44.00	38.00
Health	84.00	84.00	82.00	Health	90.00	88.00	70.00	Health	82.00	82.00	56.00
Sports	72.00	54.00	64.00	Sports	88.00	54.00	78.00	Sports	86.00	42.00	58.00
Technology	84.00	74.00	70.00	Technology	86.00	78.00	74.00	Technology	78.00	74.00	72.00
Travel	80.00	52.00	86.00	Travel	72.00	66.00	62.00	Travel	64.00	52.00	50.00

Appendix B: Mean Normalized Recall Ratios

Google Web search engine				Bing Web search engine				Yahoo Web search engine			
Category	English	Arabic	Turkish	Category	English	Arabic	Turkish	Category	English	Arabic	Turkish
Business	71.43	41.85	70.11	Business	61.88	52.92	45.34	Business	75.00	68.99	47.92
Culture	80.83	72.52	75.60	Culture	73.65	60.04	59.35	Culture	77.78	53.86	45.75
Health	76.03	69.60	76.11	Health	59.17	76.25	53.65	Health	90.83	79.72	58.10
Sports	73.63	52.96	49.88	Sports	83.25	58.71	58.87	Sports	81.11	38.81	61.38
Technology	79.35	61.67	70.00	Technology	77.50	62.20	60.14	Technology	83.91	60.95	51.81
Travel	87.62	54.45	71.94	Travel	63.33	48.27	79.68	Travel	68.47	48.13	37.78