An Experimental Study on the Effect of Using the Wrongly Spelled and/or Pronounced Turkish Words on Web Search Engines

Şeyda Türker

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. Yıltan Bitirim
Supervisor

Examining Committee

1. Asst. Prof. Dr. Yıltan Bitirim

2. Asst. Prof. Dr. Gürcü Öz

3. Asst. Prof. Dr. Önsen Toygar

ABSTRACT

This study investigates how the Web search engines handle Turkish words which are frequently wrongly spelled and/or pronounced with their own particular wrong form(s). First of all, the three most popular international Web search engines Google, Bing, and Yahoo were selected, and a query list consisted of a set of such words with their incorrect forms was formed. All queries were run on the Web search engines separately and, at each run, every document retrieved in the first twenty was classified as "relevant" or "non-relevant". Precision ratios and normalized recall ratios were calculated at various cut-off points. It seems that using incorrect forms affected the information retrieval effectiveness of the Web search engines in a negative way.

Keywords: Web Search Engine, Turkish Language, Evaluation, Precision Ratio, Normalized Recall Ratio.

ÖZ

Bu çalışma, Web arama motorlarının, Türkçede kendilerine özgü yanlış formlarıyla sıklıkla yanlış yazılan ve/veya yanlış telafuz edilen kelimeleri nasıl ele aldığını araştırır. İlk olarak, en popular üç uluslararası Web arama motoru, Google, Bing, ve Yahoo seçildi ve bu tür kelimelerin bir kümesini yanlış formları ile birlikte içeren bir sorgu listesi oluşturuldu. Bütün sorgular, seçilen arama motorları üzerinde ayrı ayrı çalıştırıldı ve her çalıştırmada, ilk 20' de erişilen her belge "ilgili" veya "ilgisiz" olarak sınıflandırıldı. Çeşitli kesme-noktalarında duyarlılık oranları ve normalize sıralama oranları hesaplandı. Yanlış formların kullanımının Web arama motorlarının bilgi erişim etkinliğini olumsuz yönde etkilediği görülmektedir.

Anahtar Kelimeler: Web Arama Motoru, Türkçe Dili, Değerlendirme, Duyarlılık Oranı, Normalize Sıralama Oranı.

To My Mother: Bedriye Başkale

To My Father: Atilla Türker

To My Brother: Emirhan Türker

To My Sister: İnci Türker

To My fiancé: Mohamad Yahya J. Namk

ACKNOWLEDGMENT

First and foremost, I would like to begin by articulation my gratitude and appreciation to my supervisor Asst. Prof. Dr. Yıltan Bitirim for his exemplary guidance and invaluable assistance. In addition, his great inspiration, encouragement, and motivation that contributed tremendously my ability to complete my Master of Science thesis.

My sincere thanks also goes to my mother Bedriye Başkale, my father Atilla Türker, my brother Emirhan Türker and my sister İnci Türker for their support throughout every second in my life.

Last but not least, gratitude for my lovely fiancé Mohamad Yahya J. Namk who helped me with my work and stood beside me all the time with patience and love. I owe him many things and I express my thankfulness with a pure promise, I love you and wish we are together forever.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TLA	Turkish Language Association
CFQ	Correct Form Query
IFQ	Incorrect Form Query
MPR	Mean Precision Ratio
MNRR	Mean Normalized Recall Ratio

Chapter 1

INTRODUCTION

Generally, Web search engine usage is one of the most frequently used activities of Internet users. In April 2014, there were nearly 36 million Internet users in Turkey which was about 47% of the population [1]. Hence, it could be said that Web search engine usage is popular in Turkish community. In addition, Turkish language is among the most commonly used 25 languages in the world [2] and the commonly used languages earned more significance for information retrieval [3]. These bring us to that those studies about evaluation of Web search engine(s) based on Turkish language could be important.

Some specific words in Turkish language are frequently being wrongly spelled and/or pronounced with their own particular wrong form(s). For example: the word "şemsiye" ("umbrella") could be spelled and pronounced wrongly as "şemşiye" (the letter "s" replaced with "ş"). Furthermore, the authoritative Turkish Language Association (TLA) has study for collecting such words in the online Web site [4]. In our study, we investigated how the international Web search engines Google [5], Bing [6], and Yahoo [7] handle this type of words. To do this, we used precision and normalized recall ratios.

This study could be valuable for motivating search engine providers and corresponding researchers.

In our literature survey, we encountered various studies related to language-based Web search engine evaluation such as follows:

Bar-Ilan and Gutman [8] investigated the performance of Web search engines for non-English languages. They selected four languages which are Russian, France, Hungarian, and Hebrew. For each language, they selected some local Web search engines as follows: For Russian language, Yandex, Rambler, and Aport; for French language, Voila, AOL France, and La Toile; for Hungarian language, Origa-Vizsla, Startlap, and Heureka; and for Hebrew language, Morfix and Walla were selected. Afterwards, for each language, the three international Web search engines AltaVista, AlltheWeb, and Google and the corresponding local Web search engines were evaluated. One of the findings is that the international Web search engines were not considered the characteristics of non-English languages.

Moukdad [9] evaluated the performance of international Web search engines, i.e., AltaVista, AlltheWeb, and Google and local Web search engines, i.e., Al bahhar, Ayna, and Morfix (the arabic module) based upon their ability to retrieve documents which have morphologically related Arabic terms. For Arabic document retrieval, the international Web search engines were limited. When the exact forms of Arabic words were used as search terms, the loss of high number of documents was caused.

Efthimiadis et al. [10] used the navigational queries in order to evaluate ten Web search engines, five international, i.e., A9, AltaVista, Google, MSN, and Yahoo and five Greek, i.e., Anazitisi, Ano-Kato, Phantis, Trinity, and Visto. 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”. The freshness of the Web search engine indexes were evaluated as well. One of the findings is that the Greek Web search engines showed poor performance compared to the international Web search engines.

Demirci et al. [11] evaluated the performance of the five international Web search engines Google, Yahoo, Msn, Ask, and AlltheWeb on finding Turkish documents and compared them with the local Turkish Web search engine Arabul. Google showed better performance than the others in general.

This study is an extended version of our study [12]. In this study, normalized recall ratio is additionally used. Furthermore, the relationship of mean precision ratio differences and mean normalized recall ratio differences is evaluated.

The next chapter describes the methodology, chapter 3 presents the experimental results, and the last chapter concludes the study.

Chapter 2

METHODOLOGY

First of all, the most popular three international Web search engines, namely, Google, Bing, and Yahoo, were selected from the eBizMBA's study [13].

2.1 Query Selection

Several corresponding Web sites [14-20] were used and a list of the wrongly spelled and/or pronounced Turkish words formed. All words in the list were checked one by one from the online Web site [4] of the authoritative TLA and the words included by TLA were filtered. Then, 24 of them were added into our query list. Additionally, 6 more words which are not in the list obtained from the Web sites [14-20] were added into our query list from the randomly sliding words displayed on the TLA's Web site in order to increase the variety. Finally, 30 words were being determined in the query list which would be used as queries. After that, incorrect forms of these 30 queries were also taken from the TLA's Web site and included in the query list. 30 correct form queries and their incorrect forms (total 60 queries) are shown in Table 1. Note that, from now on, "CFQ" abbreviation will be used to mention correct form query and "IFQ" abbreviation will be used to mention incorrect form query.

Table 1: List of the queries

Query #	CFQ	Query #	IFQ
1	acente (agency)	1	acenta
2	aşçı (cook)	2	ahçı
3	gazete (newspaper)	3	gaste
4	antrenman (training)	4	antreman
5	cambaz (acrobat)	5	canbaz
6	makine (machine)	6	makina
7	satranç (chess)	7	santranç
8	dinozor (dinosaur)	8	dinazor
9	dokümanter (documentary)	9	dökümanter
10	entelektüel (intellectual)	10	entellektüel
11	operasyon (operation)	11	oparasyon
12	fantezi (fantasy)	12	fantazi
13	floresan (fluorescence)	13	florasan
14	fasulye (beans)	14	fasülye
15	inkılap (revolution)	15	inkilap
16	portmanto (hallstand)	16	fortmanto
17	palyaço (clown)	17	palyanço
18	istanbul (istanbul)	18	istambul
19	poğaça (pastry)	19	poaça
20	kangren (gangrene)	20	gangren
21	stepne (spare tire)	21	istepne
22	sarımsak (garlic)	22	sarmısak
23	magma (magma)	23	mağma
24	manipülasyon (manipulation)	24	manüpülasyon
25	menopoz (menopause)	25	menapoz
26	ofsayt (offside)	26	opsayd
27	alerji (allergies)	27	allerji
28	taşikardi (tachycardia)	28	takikardi
29	şarj (charge)	29	şarz
30	şemsiye (umbrella)	30	şemşiye

2.2 Query Run

Before run the queries, some advanced settings were done for all three Web search engines as follows: Search results language feature was adjusted to Turkish Language to retrieve the documents in Turkish language only; the filter for SafeSearch was turned off to ensure that not any relevant document is eliminated in any case; and in the country/location section, necessary adjustment was done for retrieving the documents from all over the world.

The first CFQ (table 1) was run on the three Web search engines one by one and, at each run, the first 20 documents (two pages[1]) retrieved were evaluated. Then, the same was done for the first IFQ (table 1). Afterwards, for each of the other CFQ and IFQ pairs (table 1), this was done in the same manner. Thus, total 180 runs with retrieval output evaluations were done. For each retrieval output evaluation, the first 20 documents retrieved were classified as "relevant" or "non-relevant" according to the followings:

I. if a document content was related to the query topic, it was classified as "relevant";

II. if a document was repeated once or more 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 classified as "relevant" or classified as "non-relevant" depending upon the relevancy of the document content with the query topic ("T1" code will be used to represent a repeated-document in a such case);

[1] Spink and Jansen stated in their study [21] that users do not tend to look results pages after the first or second.

III. if a document was repeated once or more with the same URL address – the URL addresses appeared in the address bars were considered after the documents were displayed–, the repeated-document was classified as "relevant" or "non-relevant" depending upon the relevancy of the document content with the query topic, while the repeat-document(s) was classified as "non-relevant"("T2" code will be used to represent a repeated-document in a such case);

A document was classified as "non-relevant" if:

- a document content was not related to the query topic;
- a document consisted of images or videos which had no text details related to the query topic;
- a document content was not displayed in Turkish language;
- a document asked a login information to view the content;
- one of the following messages was displayed: "Website is under construction"; "Website is frozen"; "Temporarily unavailable, check back soon"; "Database error"; "403 error", "500 error", and "503 error";
- one of the following messages was displayed: "404 error"; "This website is not available"; "No data received". This situation was taken into account as a dead link situation.

2.3 Evaluation Metrics

Precision is one of the most widely used metric to evaluate the retrieval effectiveness of the information retrieval system [22]. In our study, we used precision ratio with the following formula [23]:

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

Precision ratios were calculated at cut-off points 5, 10, 15, and 20 for each query (CFQs and IFQs) and Web search engine pair. After that, Mean Precision Ratios (MPRs) for CFQs and MPRs for IFQs were calculated at four cut-off points for each Web search engine. Additionally, MPR difference between MPR for CFQs and MPR for IFQs was calculated at each cut-off point for each Web search engine.

Normalized recall is another metric to evaluate the information retrieval system based on performance of displaying relevant documents retrieved in higher ranks. The normalized recall is defined as follows [24]:

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

In this formula:R^{+} is the number of document pairs that relevant document displayed before non-relevant; R^{-} is the number of document pairs that non-relevant document displayed before relevant; and R^{+}_{max} is the maximal number of document pairs that relevant document displayed before non-relevant.

In our study, beside precision ratio, we used normalized recall ratio. Normalized recall ratio is the multiplication of normalized recall with 100. Normalized recall ratio were calculated at cut-off points 5, 10, 15, and 20 for each query (CFQs and IFQs) and Web search engine pair. After that, Mean Normalized Recall Ratios (MNRRs) for CFQs and MNRRs for IFQs were calculated at four cut-off points for each Web search engine. Additionally, MNRR difference between MNRR for CFQs and MRNR for IFQs was calculated at each cut-off point for each Web search engine.

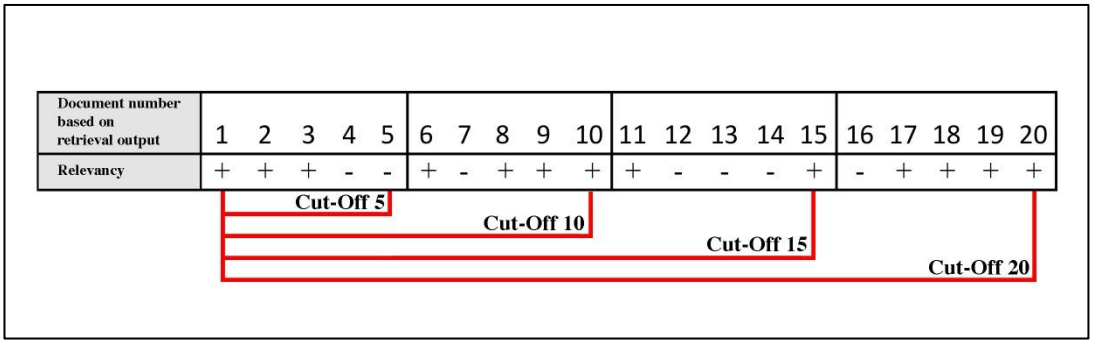

Figure 2.1: Google's retrieval output evaluation for "floresan"

For instance, Google's retrieval output for the 13th CFQ "floresan" was evaluated at cut-of points 5, 10, 15, and 20 as shown in Figure 2.1. Note that '+' is used for representing relevant documents and '-' is used for representing non-relevant documents. Afterwards, precision ratio and normalized recall ratio were calculated at cut-off points 5, 10, 15, and 20 as follows:

***At cut-off point 5*:** In the first five documents retrieved in the retrieval output, there were three '+'s and two '-'s. The number of relevant document retrieved is equal to 3. The total number of document retrieved is 5.

$$precision\ ratio = \frac{3}{5} \times 100 = 60\%$$

***At cut-off point 10*:** In the first ten documents retrieved in the retrieval output, there were seven '+'s and three '-'s. The number of relevant document retrieved is equal to 7. The total number of document retrieved is 10.

$$precision\ ratio = \frac{7}{10} \times 100 = 70\%$$

***At cut-off point 15*:** In the first fifteen documents retrieved in the retrieval output, there were nine '+'s and six '-'s. The number of relevant document retrieved is equal to 9. The total number of document retrieved is 15.

$$precision\ ratio = \frac{9}{15} \times 100 = 60\%$$

***At cut-off point 20*:** In the first twenty documents retrieved in the retrieval output, there were thirteen '+'s and seven '-'s. The number of relevant document retrieved is equal to 13. The total number of document retrieved is 20.

$$precision\ ratio = \frac{13}{20} \times 100 = 65\%$$

Note that when total number of documents retrieved was greater than or equal to cut-off point value (i.e. 5, 10, 15, or 20), total number of documents retrieved was considered the same as the cut-off point value. (The "equal" situation was not encountered.) When total number of documents retrieved was less than the cut-off point value, total number of documents retrieved was considered as it is. This case was only encountered for one query ("opsayd") at cut-off points 10, 15, and 20 which only 6 documents in total were retrieved.

***At cut-off point 5*:** For R^+, every document pair that relevant document displayed before non-relevant was counted as shown in Figure 2.2. R^+ is equal to 6. For R^-, there was no document pair that non-relevant document displayed before relevant; therefore R^- is equal 0. For R^+_{max}, first of all the retrieval output should be modified to find out the maximal number of document pairs that relevant document displayed before non-relevant. However, here, no need for modification since, the '-'s came after all '+'s, every document pair that relevant document displayed before non-

relevant was counted in the same manner of counting R^+ as shown in Figure 2.2. R^+_{max} is equal to 6.

$$Normalized\ recall\ ratio = \frac{1}{2}\left(1 + \frac{6-0}{6}\right) \times 100 = 100\%$$

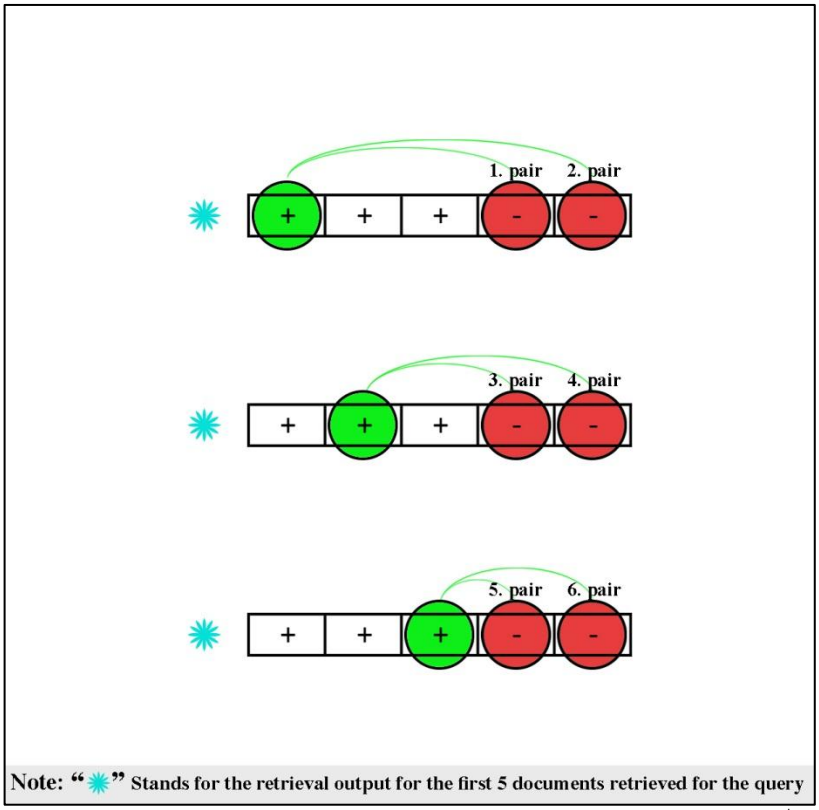

Figure 2.2: Document pair counting at cut-off point 5 for R^+

***At cut-off point 10*:** For R^+, every document pair that relevant document displayed before non-relevant was counted as shown in Figure 2.3. R^+ is equal to 10. For R^-, every document pair that non-relevant document displayed before relevant was counted as shown in Figure 2.4. R^- is equal 11. For R^+_{max}, the retrieval output was modified to find out the maximal number of document pairs that relevant document displayed before non-relevant as shown in Figure 2.5 and every document pair that relevant document displayed before non-relevant was counted as shown in Figure 2.6. R^+_{max} is equal to 21.

$$Normalized\ recall\ ratio = \frac{1}{2}\left(1 + \frac{10 - 11}{21}\right) \times 100 = 48\%$$

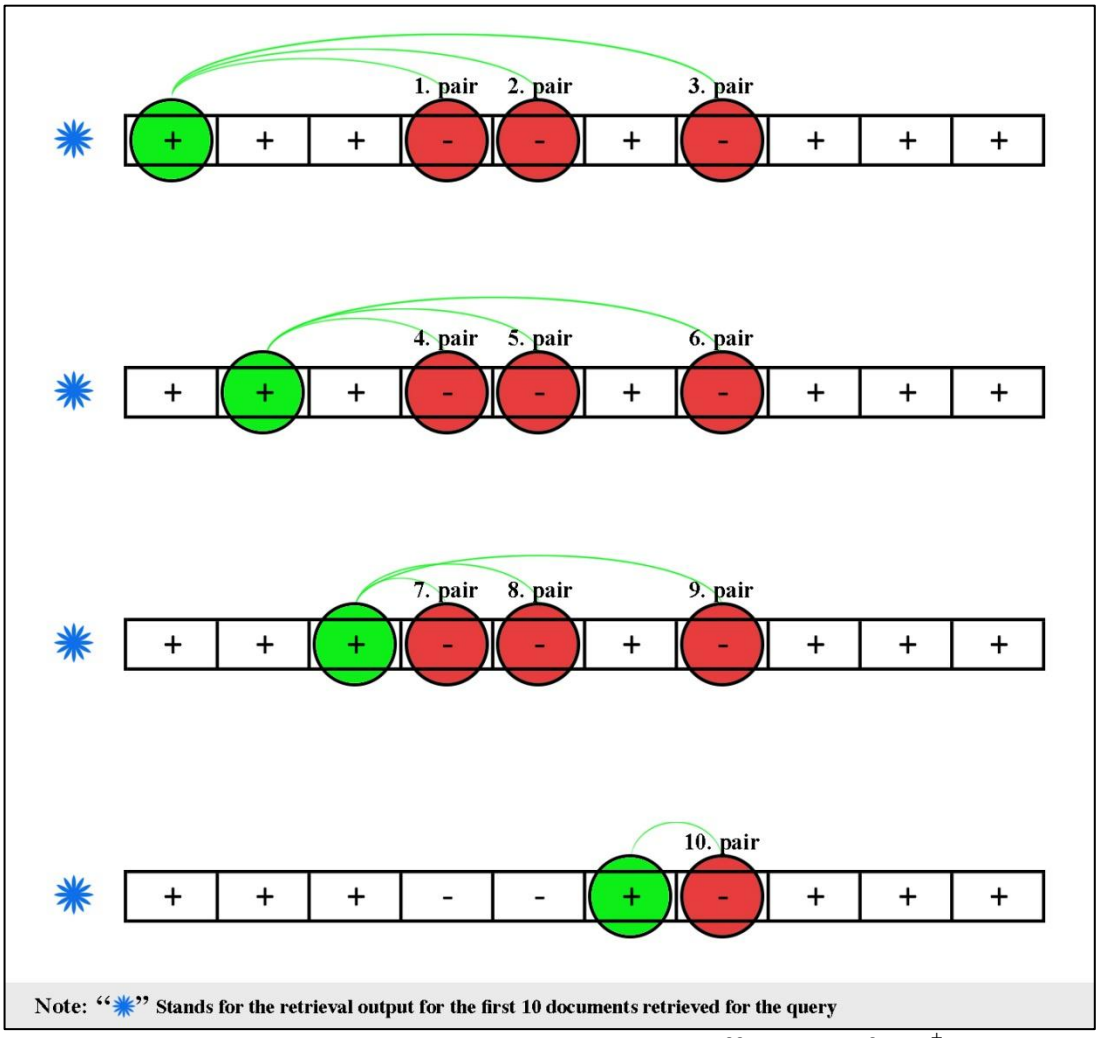

Figure 2.3: Document pair counting at cut-off point 10 for R^+

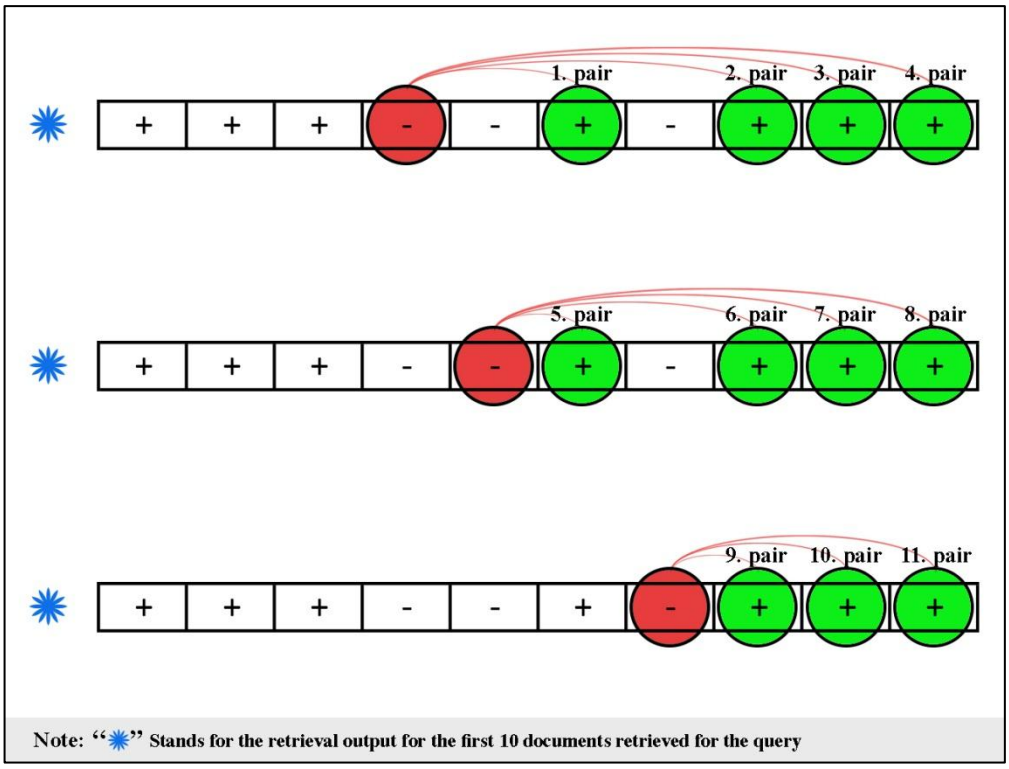

Figure 2.4: Document pair counting at cut-off point 10 for R^-

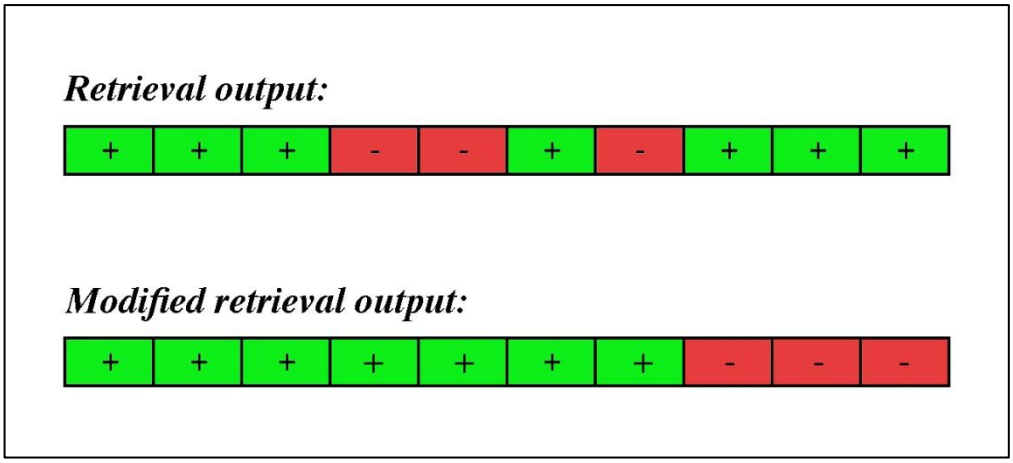

Figure 2.5: Modification of retrieval output at cut-off point 10 for $\mathbf{R^+_{max}}$

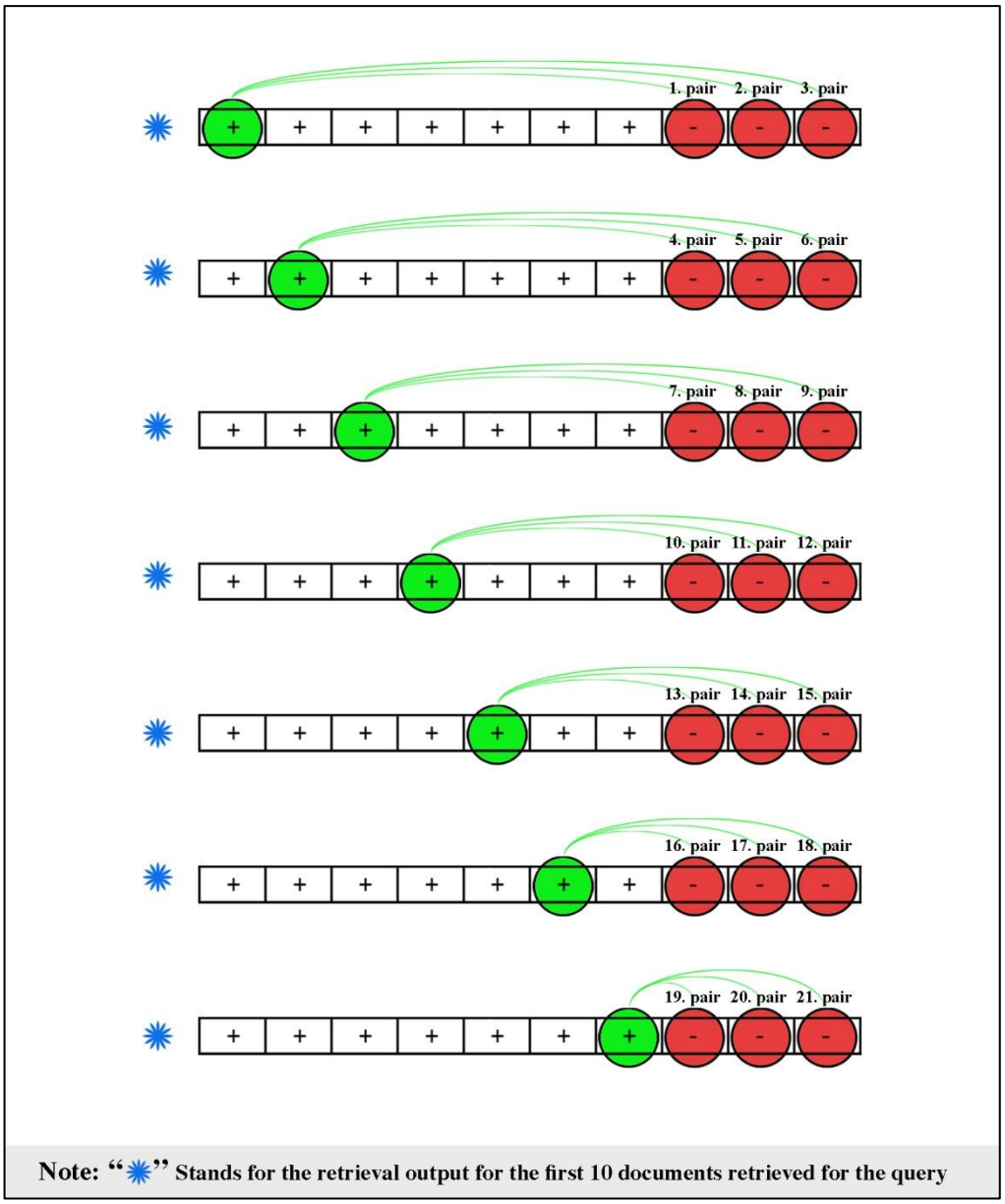

Figure 2.6: Document pair counting at cut-off point 10 for $\mathbf{R^{+}_{max}}$

***At cut-off point 15*:** For R^{+}, every document pair that relevant document displayed before non-relevant was counted as shown in Figure 2.7. R^{+} is equal to 34. For R^{-}, every document pair that non-relevant document displayed before relevant was counted as shown in Figure 2.8. R^{-} is equal 20. For R^{+}_{max}, the retrieval output was modified to find out the maximal number of document pairs that relevant document displayed before non-relevant as shown in Figure 2.9 and every document pair that

relevant document displayed before non-relevant was counted as shown in Figure 2.10. R^{+}_{max} is equal to 54.

$$Normalized\ recall\ ratio = \frac{1}{2}\left(1 + \frac{34 - 20}{54}\right) \times 100 = 63\%$$

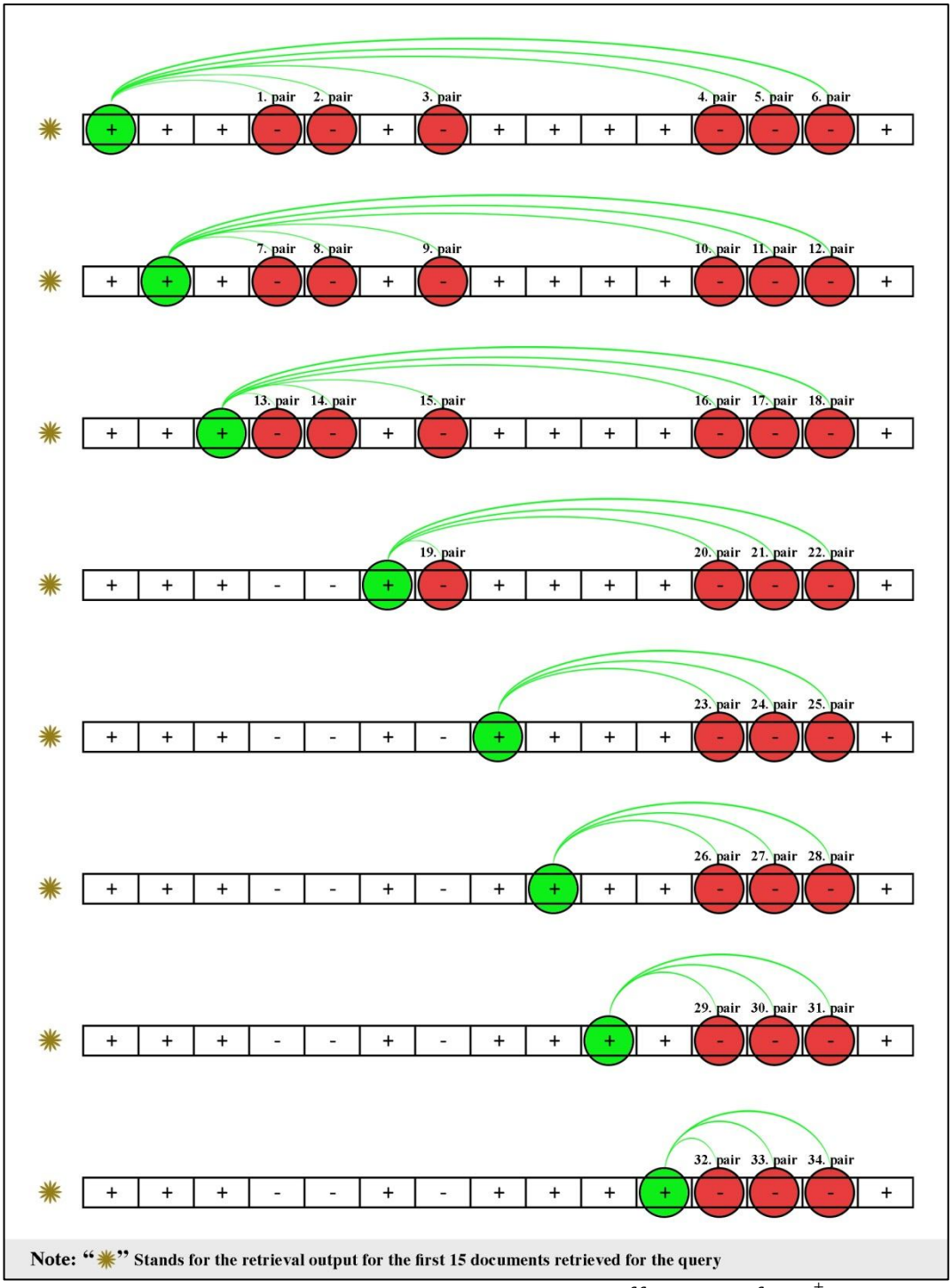

Figure 2.7: Document pair counting at cut-off point 15 for R^{+}

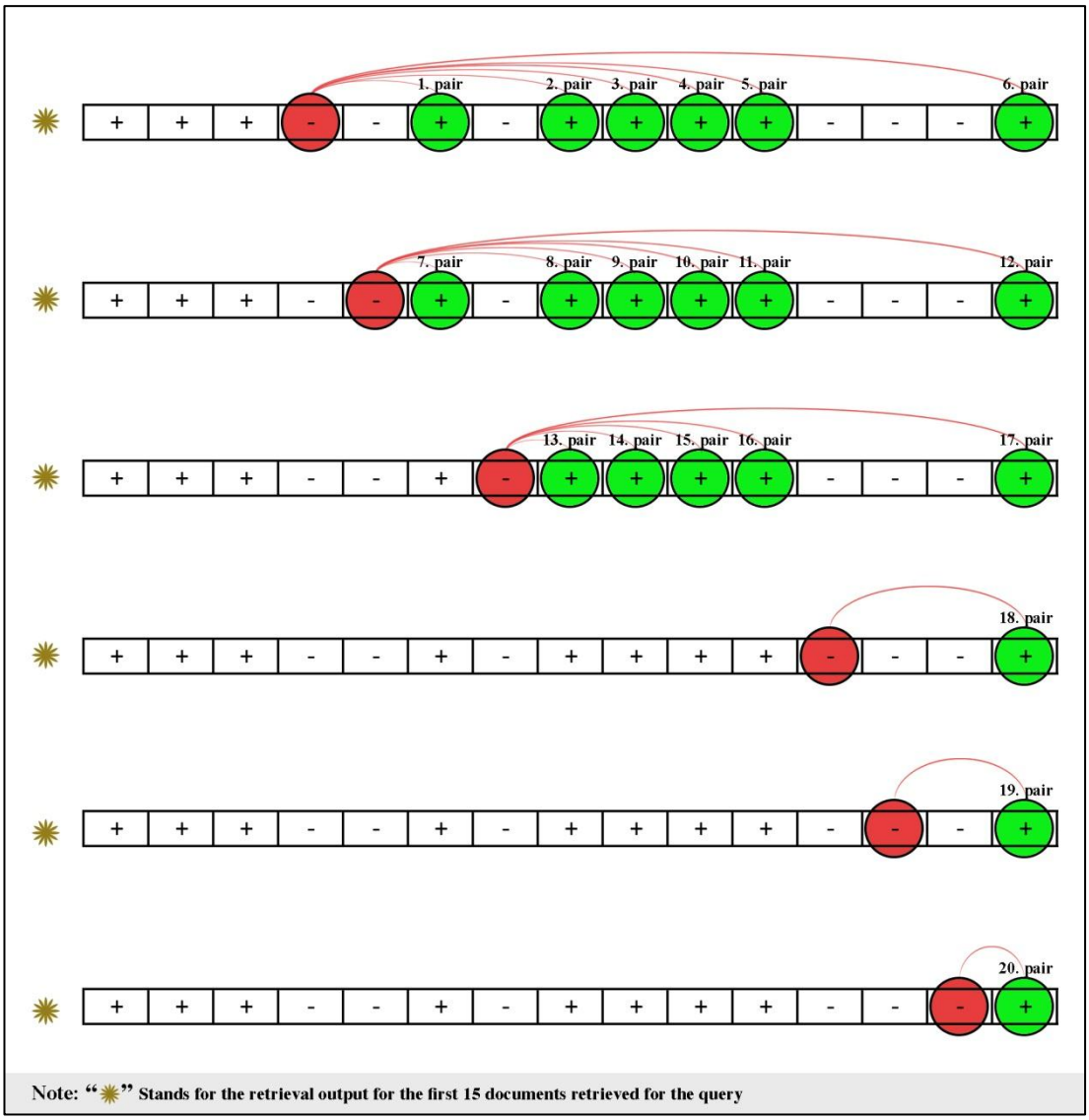

Figure 2.8: Document pair counting at cut-off point 15 for R^-

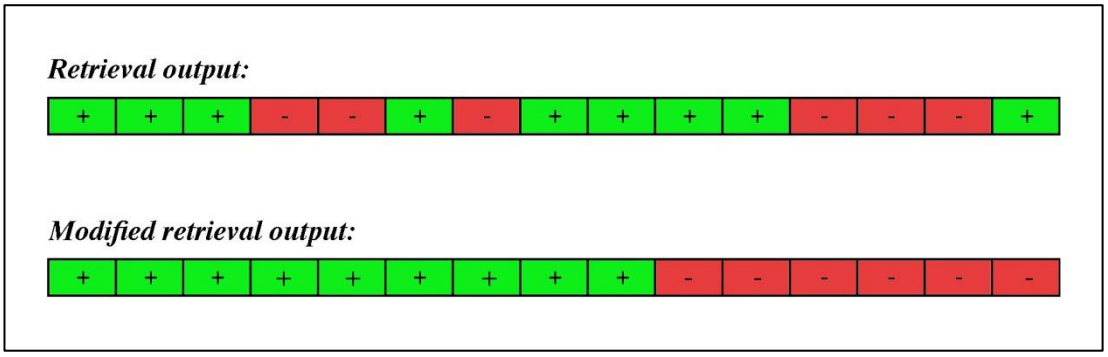

Figure 2.9: Modification of retrieval output at cut-off point 15 for $\mathbf{R^+_{max}}$

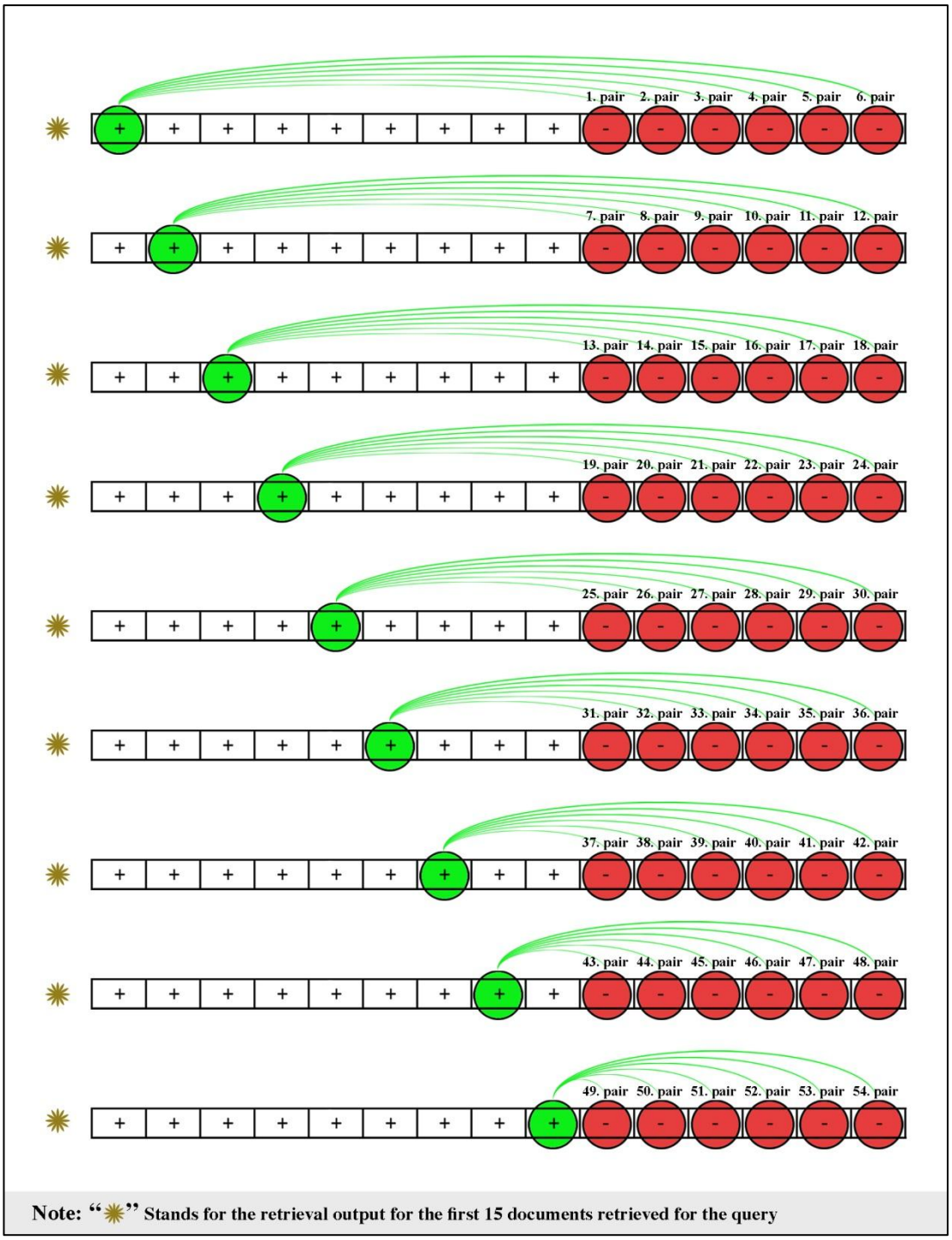

Figure 2.10: Document pair counting at cut-off point 15 for $\mathbf{R^{+}_{max}}$

***At cut-off point 20*:** For R^{+}, every document pair that relevant document displayed before non-relevant was counted as shown in Figure 2.11. R^{+} is equal to 43. For R^{-}, every document pair that non-relevant document displayed before relevant was counted as shown in Figure 2.12. R^{-} is equal 48. For R^{+}_{max}, the retrieval output was

modified to find out the maximal number of document pairs that relevant document displayed before non-relevant as shown in Figure 2.13 and every document pair that relevant document displayed before non-relevant was counted as shown in Figure 2.14. R^{+}_{max} is equal to 91.

$$Normalized\ recall\ ratio = \frac{1}{2}\left(1 + \frac{43 - 48}{91}\right) \times 100 = 47\%$$

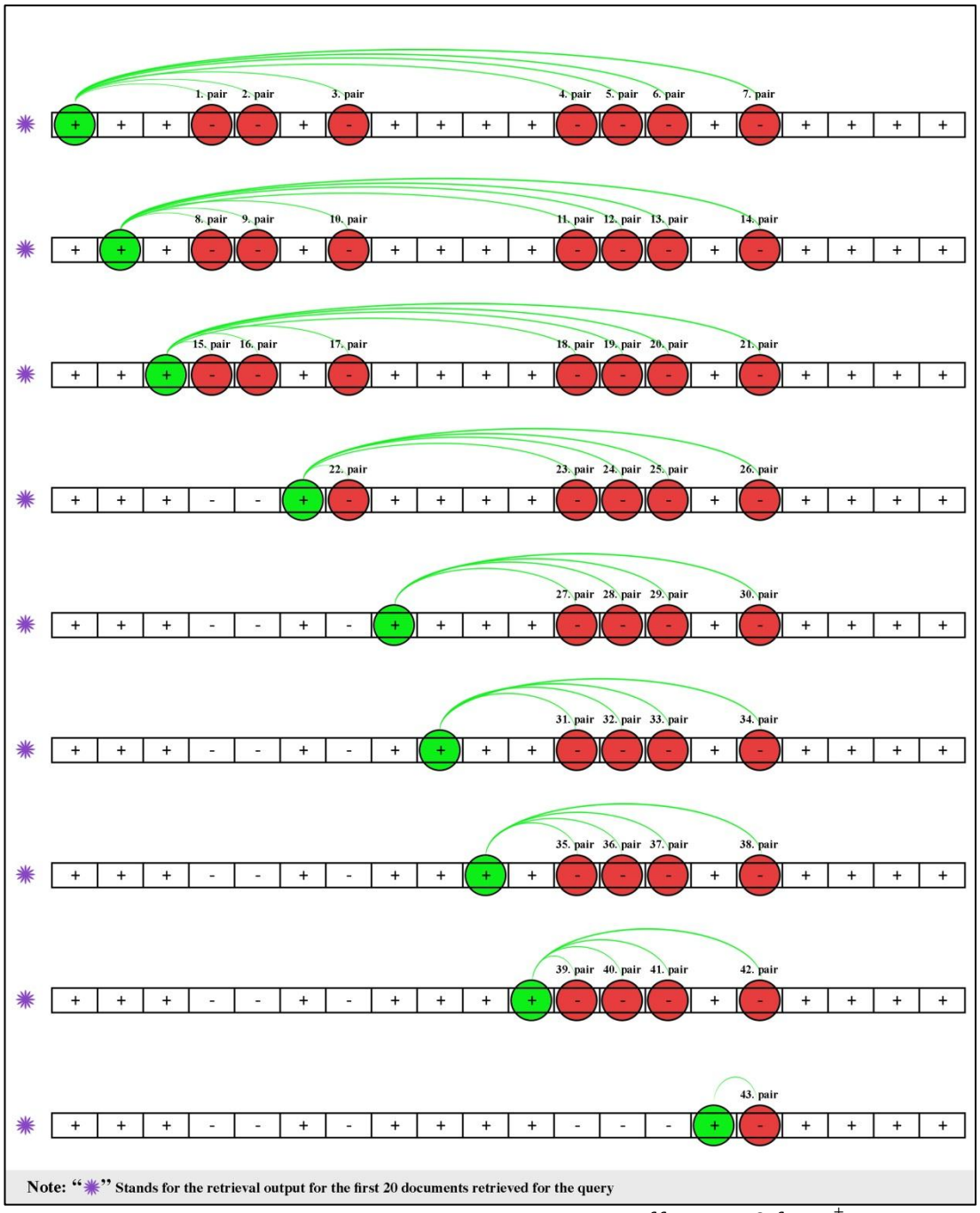

Figure 2.11: Document pair counting at cut-off point 20 for R^{+}

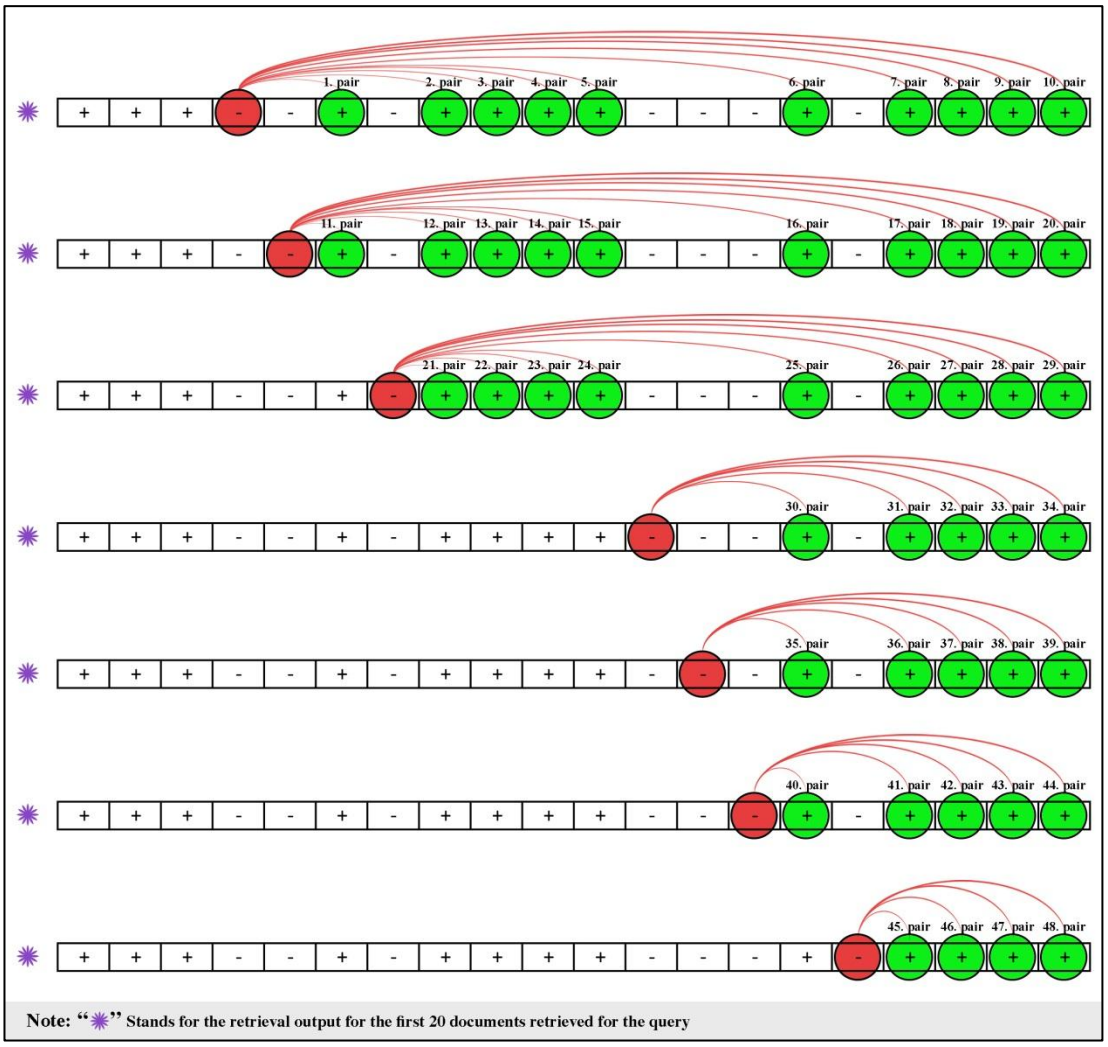

Figure 2.12: Document pair counting at cut-off point 20 for R^-

Figure 2.13: Modification of retrieval output at cut-off point 20 for $\mathbf{R^+_{max}}$

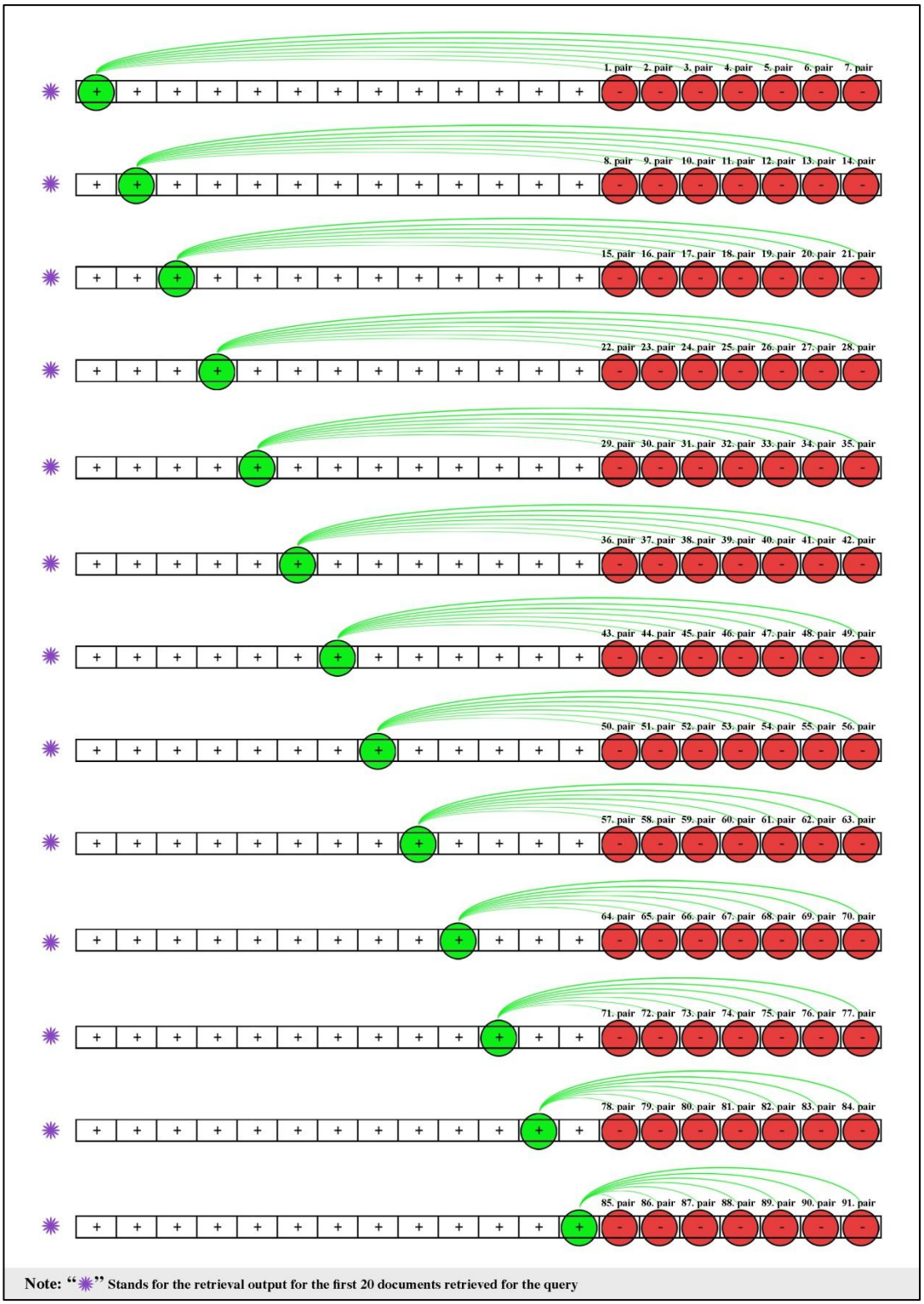

Figure 2.14: Document pair counting at cut-off point 20 for $\mathbf{R^{+}_{max}}$

Note that when total number of relevant documents retrieved was less than cut-off point value (i.e. 5, 10, 15, or 20), the rest till the cut-off point were considered as neutral. Thus, for R^+ calculation, each neutral was considered as '-'; for R^- calculation, each neutral was considered as '+'; and for R^+_{max} calculation, in the modification of the retrieval output, the neutrals came after all '+'s and before '-'s and each was considered as '+'. This case was only encountered for one query ("opsayd") at cut-off points 10, 15, and 20 which only 6 documents in total were retrieved.

Chapter 3

EXPERIMENTAL RESULTS

All queries were run on each Web search engine, and then in total 3586 documents were evaluated one by one. Due to the Web search engine Bing retrieved only 6 documents in total for 26th IFQ, the total number of documents did not result as 3600 (60 queries × the first 20 documents retrieved × 3 search engines). When considering all retrieval outputs in the experiment, total 41 dead link situations, 50 T1s, and 35 T2s were encountered.

Google, Bing, and Yahoo retrieved at least one relevant document at cut-off points 10, 15, and 20 for all CFQs, except for one query (CFQ 2) of Bing at cut-off point 10 as it is seen in Table 2. Each of the Web search engines retrieved zero relevant documents at all cut-off points for at least 2 IFQs. At all cut-off points: Google could not retrieve any relevant document for IFQ 2 and IFQ 29; Bing could not retrieve any relevant document for IFQ 1, IFQ 10, and IFQ 26; and Yahoo could not retrieve any relevant document for IFQ 8 and IFQ 26. When considering total number of relevant documents of all CFQs at cut-off point 20 for each Web search engine, Google, Bing, and Yahoo retrieved relevant documents 45%, 44%, and 42%, respectively; likewise, for IFQs, these percentages were 31%, 26%, and 29%, respectively.

Table 2: Number of relevant documents retrieved

CFQ #	Google				Bing				Yahoo			
	cut-off				cut-off				cut-off			
	5	10	15	20	5	10	15	20	5	10	15	20
1	3	3	3	4	0	2	2	3	1	2	3	3
2	1	1	1	1	0	0	1	3	1	1	4	4
3	5	9	13	18	5	9	12	17	4	8	13	18
4	4	6	8	8	4	5	5	7	3	4	6	10
5	2	3	4	4	0	1	2	2	0	1	1	1
6	1	2	2	2	1	4	5	5	2	2	2	4
7	2	4	8	9	4	5	5	7	1	3	5	6
8	1	2	4	5	3	3	5	5	4	4	4	4
9	0	2	4	5	3	5	8	12	3	6	9	10
10	4	6	7	7	1	3	4	6	2	4	6	6
11	5	8	9	9	2	4	7	9	0	1	3	3
12	3	6	8	8	2	2	3	3	1	2	2	3
13	3	7	9	13	0	2	3	4	3	4	5	7
14	3	5	9	12	3	6	8	10	4	6	9	9
15	1	3	4	4	2	5	7	12	3	7	10	10
16	1	3	7	8	1	5	7	8	2	6	9	12
17	1	4	8	11	4	9	12	15	4	6	9	11
18	2	3	4	5	3	3	4	4	2	3	5	5
19	4	7	11	14	4	7	12	15	5	9	13	18
20	3	6	7	7	4	6	10	13	5	8	13	15
21	4	7	8	11	3	7	8	11	1	2	3	6
22	3	5	7	8	4	7	9	13	4	6	10	12
23	1	2	2	2	1	1	1	2	1	2	2	2
24	4	7	8	8	3	5	6	8	1	2	3	4
25	4	9	14	18	5	10	14	18	4	9	12	17
26	5	5	5	6	2	4	5	6	1	1	1	1
27	5	10	14	16	5	6	9	12	4	5	9	10
28	4	8	13	17	5	8	11	14	5	9	13	17
29	2	7	11	16	2	6	9	11	4	6	8	11
30	4	6	7	11	3	7	10	11	5	8	9	12
Total	85	156	219	267	79	147	204	266	80	137	201	251

IFQ #	Google				Bing				Yahoo			
	cut-off				cut-off				cut-off			
	5	10	15	20	5	10	15	20	5	10	15	20
1	1	2	2	2	0	0	0	0	0	0	1	1
2	0	0	0	0	1	3	3	4	0	0	1	2
3	1	3	3	6	1	2	3	4	2	2	4	6
4	2	3	6	7	3	4	4	5	2	3	5	7
5	1	1	1	1	0	0	0	1	0	0	1	2
6	0	3	5	8	2	4	4	4	3	3	4	4
7	2	4	8	9	1	1	2	3	1	2	2	3
8	1	2	2	2	1	1	1	1	0	0	0	0
9	0	2	2	2	0	0	0	1	1	2	3	3
10	3	4	5	5	0	0	0	0	3	3	5	6
11	1	2	4	6	3	4	6	8	2	2	2	3
12	1	1	1	1	1	2	2	3	0	0	2	4
13	2	4	4	5	0	2	2	2	3	6	7	8
14	3	5	9	12	0	1	4	5	4	5	6	9
15	1	2	2	2	0	1	1	1	1	5	6	7
16	1	1	1	2	1	5	8	9	2	5	9	10
17	1	2	3	3	1	2	3	3	0	0	2	2
18	2	3	4	5	3	3	5	5	2	3	3	6
19	4	8	12	16	2	6	10	13	4	9	14	17
20	3	7	11	12	3	7	10	14	3	7	11	14
21	2	3	5	6	2	2	4	7	1	3	4	4
22	3	5	7	8	4	6	10	13	4	7	7	9
23	2	2	2	3	3	4	5	7	3	5	6	9
24	2	2	4	5	1	2	2	2	1	1	2	2
25	4	9	14	18	4	9	13	17	4	8	12	16
26	1	1	1	1	0	0	0	0	0	0	0	0
27	5	7	8	10	5	8	11	12	4	7	11	14
28	4	8	13	17	0	1	1	1	1	1	1	1
29	0	0	0	0	0	1	2	4	2	2	3	3
30	5	7	8	11	3	5	7	8	2	2	2	3
Total	58	103	147	185	45	86	123	157	55	93	136	175

3.1 Mean Precision Ratio

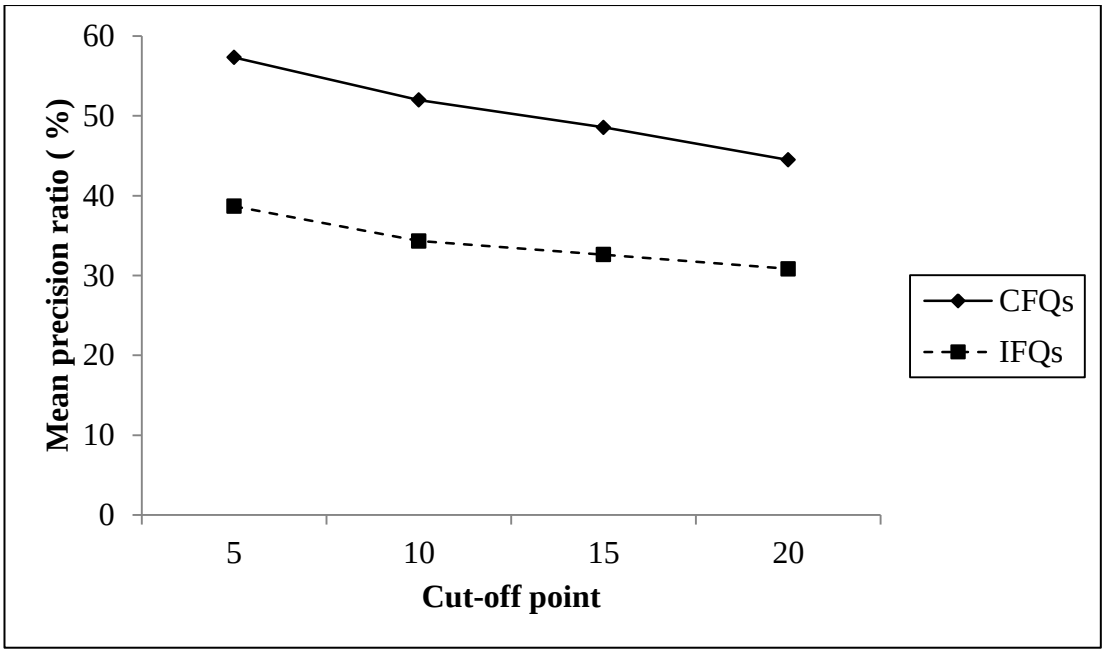

Figure 3.1: Mean precision ratios of Google search engine

Figure 3.1 presents MPRs of Google at four cut-off points for CFQs and IFQs. Google has the highest MPR at cut-off point 5 for both CFQs and IFQs; however, when the cut-off point was increased, MPR decreased till cut-off point 20. Thus, for CFQs as well as IFQs, the best performance of Google was at cut-off point 5 and the worst was at cut-off point 20. MPR for both CFQs and IFQs decreased 5% from cut-off point 5 to 10. For CFQs, it decreased 3% from cut-off point 10 to 15 and 4% from cut-off point 15 to 20. Besides, MPR for IFQs decreased 2% both from cut-off point 10 to 15 and from cut-off point 15 to 20. The MPR range for CFQs was from 45% to 57% and for IFQs was from 31% to 39%. At all cut-off points, Google has better MPRs for CFQs. The MPR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 were 19%, 18%, 16%, and 14%, respectively.

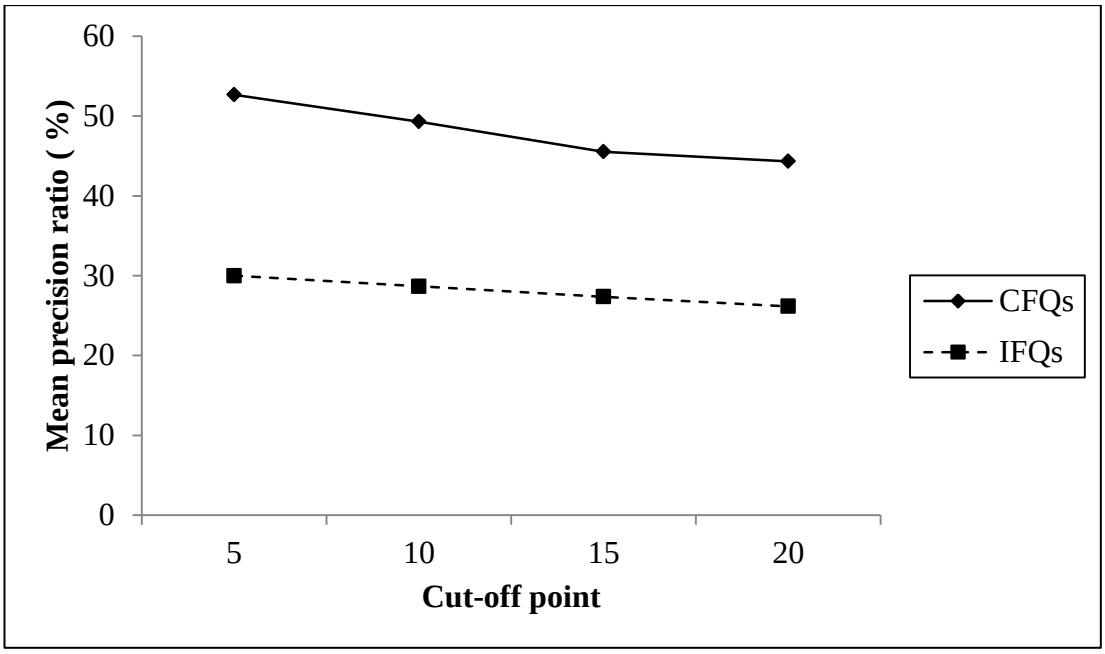

Figure 3.2: Mean precision ratios of Bing search engine

Figure 3.2 shows MPRs of Bing at four cut-off points for CFQs and IFQs. Bing has the highest MPR at cut-off point 5 for both CFQs and IFQs; however, when the cut-off point was increased, MPR decreased till cut-off point 20. So, for CFQs as well as IFQs, the best performance of Bing was at cut-off point 5 and the worst was at cut-off point 20. MPR for CFQs decreased 4% from cut-off point 5 to 10, 3% from cut-off point 10 to 15, and 2% from cut-off point 15 to 20. Besides, MPR for IFQs decreased 1% from cut-off point 5 to 10, 2% from cut-off point 10 to 15, and 1% from cut-off point 15 to 20. The MPR range for CFQs was from 44% to 53%, and for IFQs was from 26% to 30%. At all cut-off points, Bing has better MPRs for CFQs. The MPR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 were 23%, 21%, 18%, and 18%, respectively.

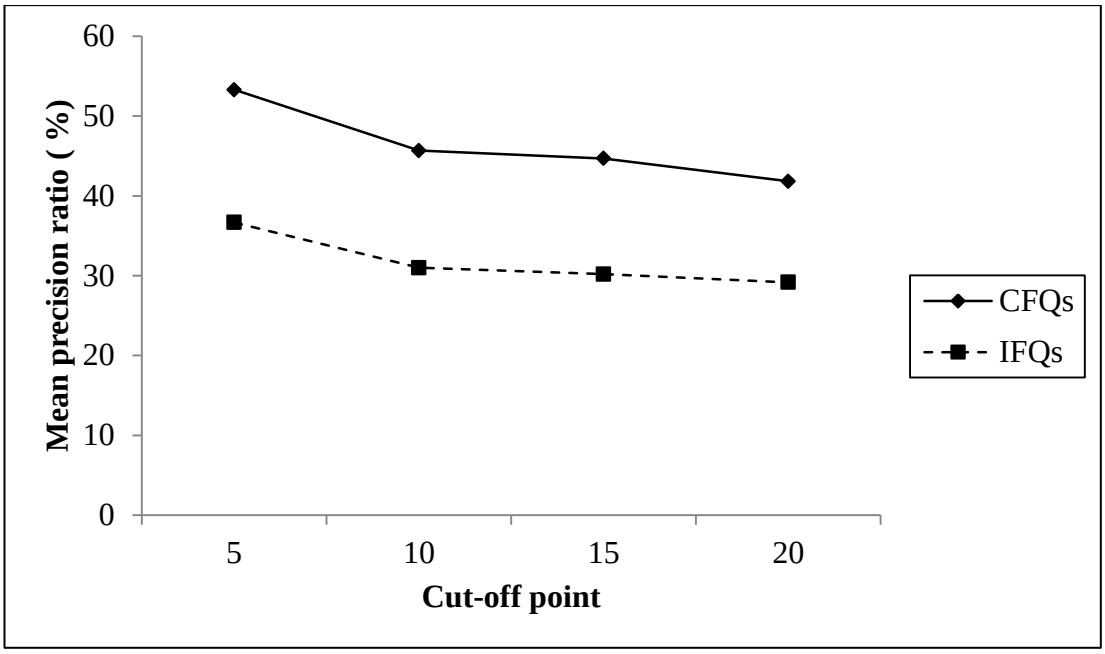

Figure 3.3: Mean precision ratios of Yahoo search engine

Figure 3.3 demonstrates MPRs of Yahoo at four cut-off points for CFQs and IFQs. Yahoo has the highest MPR at cut-off point 5 for both CFQs and IFQs; however, when the cut-off point was increased, MPR decreased till cut-off point 20. Hence, for CFQs as well as IFQs, the best performance of Yahoo was at cut-off point 5 and the worst was at cut-off point 20. MPR for CFQs decreased 8% from cut-off point 5 to 10 and 3% from cut-off point 15 to 20. Besides, MPR for IFQs decreased 6% form cut-off point 5 to 10 and 1% from cut-off point 15 to 20. MPR for both CFQs and IFQs decreased 1% from cut-off point 10 to 15. The MPR range for CFQs was from 42% to 53% and for IFQs was from 29% to 37%. At all cut-off points, Yahoo has better MPRs for CFQs. The MPR differences between CFQs and IFQs at cut-off points 5, 10, 15 and 20 were 17%, 15%, 15%, and 13%, respectively.

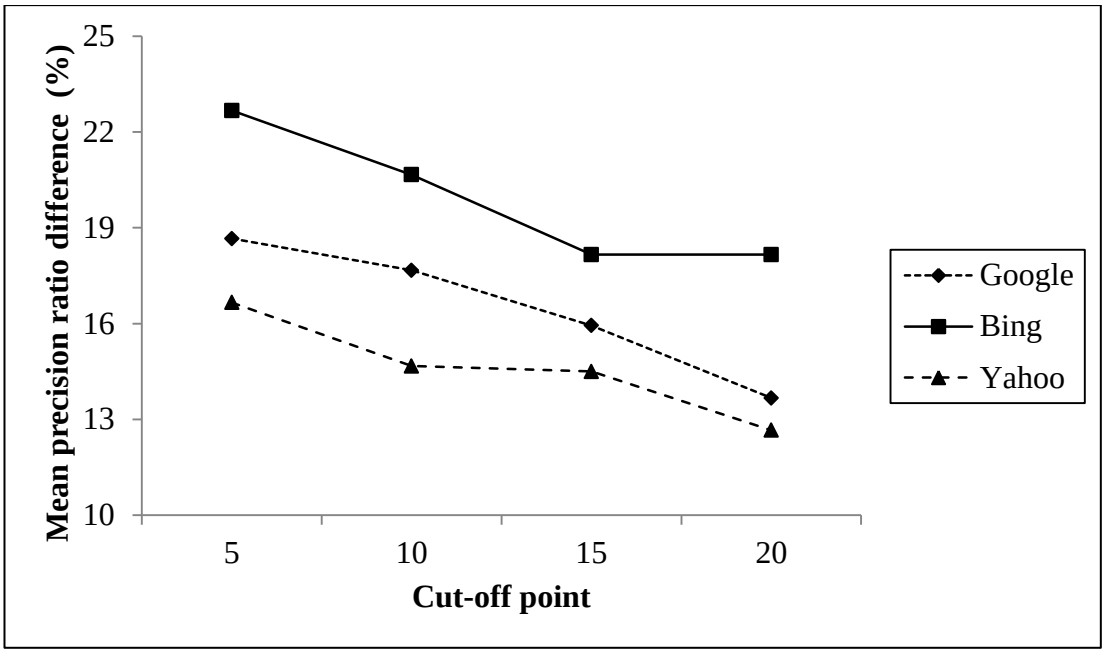

Figure 3.4: MPR differences between CFQs and IFQs at four cut-off points for each search engines

Figure 3.4 presents MPR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 for Google, Bing, and Yahoo. Each of the Web search engine has its highest MPR difference at cut-off point 5; however, as the cut-off point was increased, MPR difference generally decreased. The difference range for Bing was from 18 % to 23%, for Google it was from 14% to 19%, and for Yahoo it was from 13% to 17%. Bing has the highest MPR difference at all cut-off points. Yahoo achieved better than the other Web search engines (Google and Bing), since Yahoo has the lowest MPR difference at all cut-off points.

3.2 Mean Normalized Recall Ratio

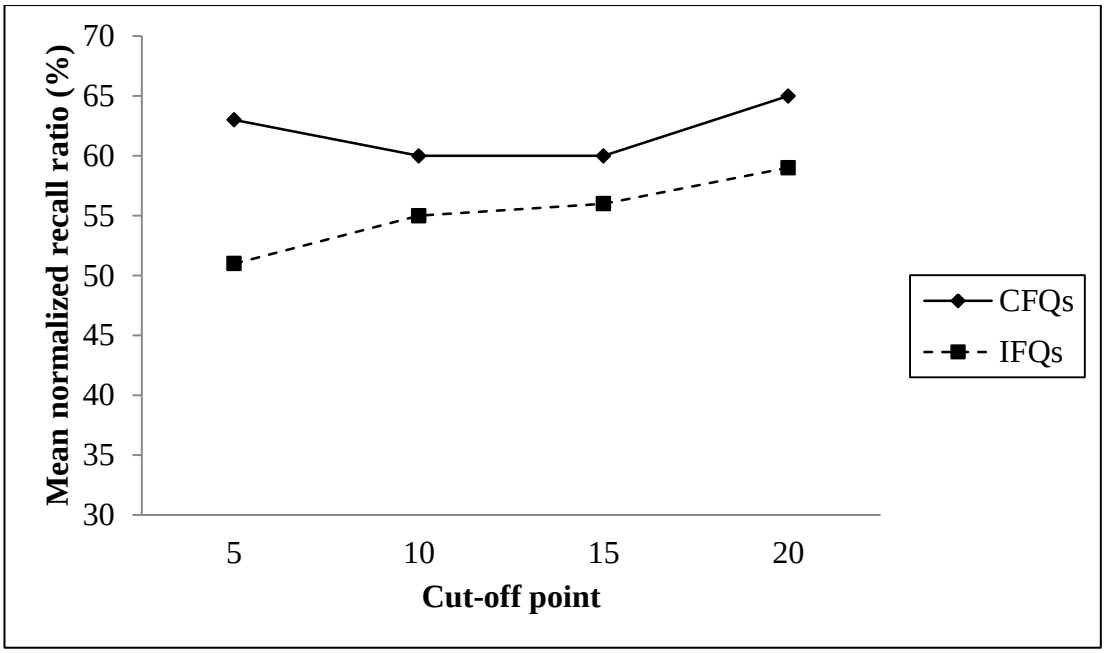

Figure 3.5: Mean normalized recall ratios of Google search engine

Figure 3.5 presents MNRRs of Google at four cut-off points for CFQs and IFQs. Google has the highest MNRR at cut-off point 20 for CFQs and it has the lowest MRPN at cut-off points 10 and 15. Google has the lowest MNRR at cut-off point 5 for IFQs; however, when the cut-off point was increased, MNRR also increased till cut-off point 20. For CFQs as well as IFQs, the best performance of Google was at cut-off point 20. MNRR for CFQs decreased 3% from cut-off point 5 to 10, not changed from cut-off point 10 to 15, and increased 5% from cut-off point 15 to 20. Besides, MNRR for IFQs increased 4% from cut-off point 5 to 10, 1% from cut-off point 10 to 15, and 3% from cut-off point 15 to 20. The MNRR range for CFQs was from 60% to 65%, and for IFQs was from 51% to 59%. At all cut-off points, Google has better MNRRs for CFQs. The MNRR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 were 12%, 5%, 4%, and 6%, respectively.

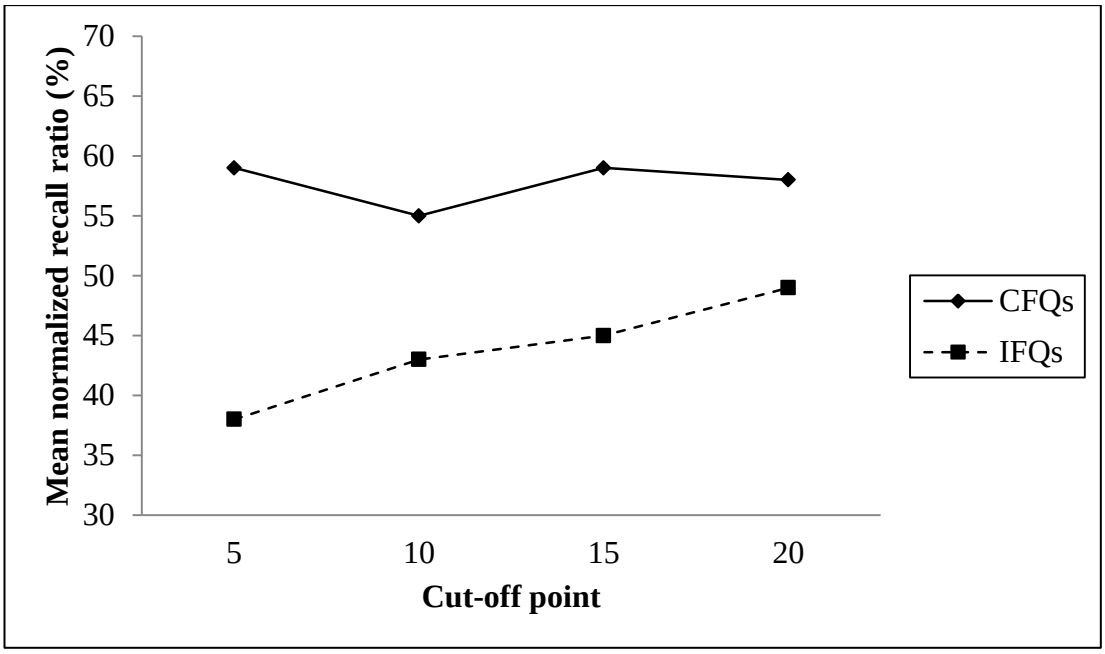

Figure 3.6: Mean normalized recall ratios of Bing search engine

Figure 3.6 shows MNRRs of Bing at four cut-off points for CFQs and IFQs. Bing has the highest MNRR both at cut-off point 5 and at cut-off point 15 for CFQs and it has the lowest MNRR at cut-off points 10. Bing has the lowest MNRR at cut-off point 5 for IFQs; however, when the cut-off point was increased, MNRR also increased till cut-off point 20. MNRR for CFQs decreased 4% from cut-off point 5 to 10, increased 4% from cut-off point 10 to 15, and decreased 1% from cut-off point 15 to 20. Besides, MNRR for IFQs increased 5% from cut-off point 5 to 10, 2% from cut-off point 10 to 15, and 4% from cut-off point 15 to 20. The MNRR range for CFQs was from 59% to 55%, and for IFQs was from 38% to 49%. At all cut-off points, Bing has better MNRRs for CFQs. The MNRR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 were 21%, 12%, 14%, and 9%, respectively.

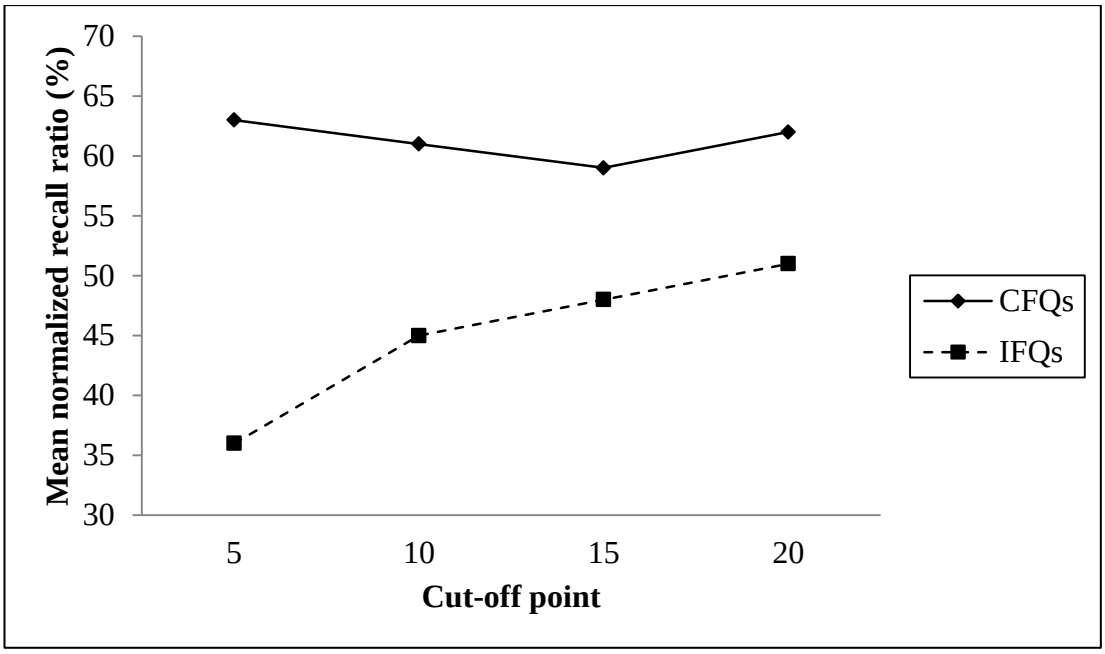

Figure 3.7: Mean normalized recall ratios of Yahoo search engine

Figure 3.7 shows MNRRs of Yahoo at four cut-off points for CFQs and IFQs. Yahoo has the highest MNRR at cut-off point 5 for CFQs and it has the lowest MNRR at cut-off points 15. Yahoo has the lowest MNRR at cut-off point 5 for IFQs; however, when the cut-off point was increased, MNRR also increased till cut-off point 20. MNRR for CFQs decreased 2% both from cut-off point 5 to 10, and from cut-off point 10 to 15, and increased 3% from cut-off point 15 to 20. Besides, MNRR for IFQs increased 9% from cut-off point 5 to 10, and 3% both from cut-off point 10 to 15 and from cut-off point 15 to 20. The MNRR range for CFQs was from 59% to 63%, and for IFQs was from 36% to 51%. At all cut-off points, Yahoo has better MNRRs for CFQs. The MNRR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 were 27%, 16%, 11%, and 11%, respectively.

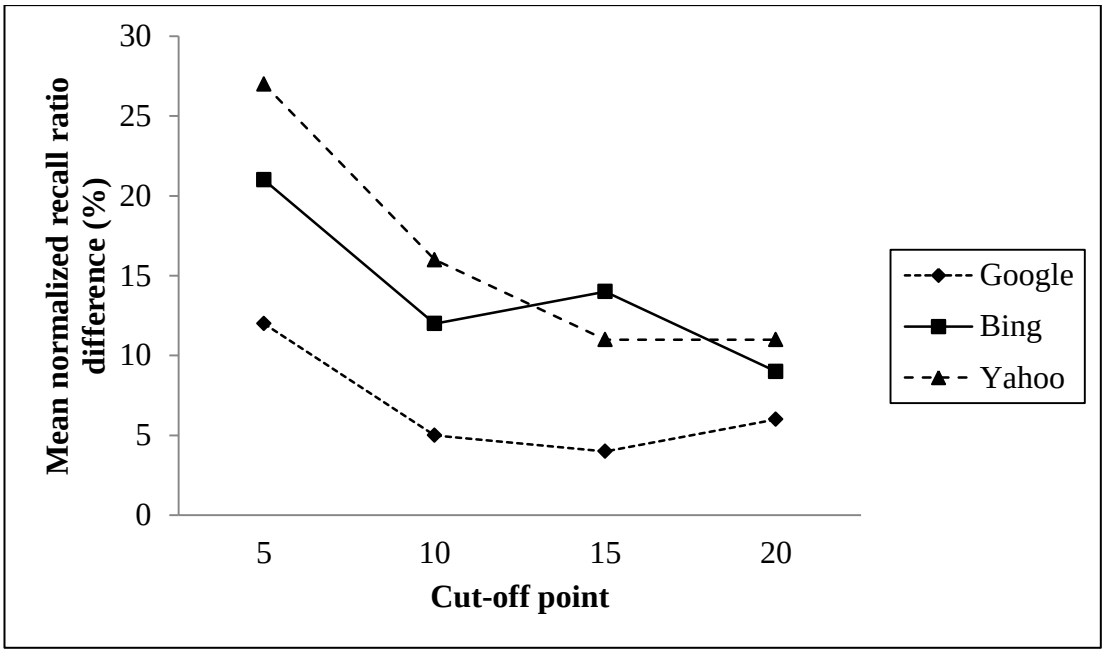

Figure 3.8: MNRR differences between CFQs and IFQs at four cut-off points for each search engines

Figure 3.8 presents MNRR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 for Google, Bing, and Yahoo. Each of the Web search engine has its highest MNRR difference at cut-off point 5; however, as the cut-off point was increased, MNRR difference generally decreased. The difference range for Bing was from 9 % to 21%, for Google it was from 4% to 12%, and for Yahoo it was from 11% to 27%. Yahoo has the highest MNRR difference at cut-off points 5, 10, and 20. At cut-off point 15 Bing has the highest MNRR differences. Google achieved better than the other Web search engines (Bing and Yahoo), since Google has the lowest MNRR difference at all cut-off points.

3.3 Relation of the MPR Differences and the MNRR Differences

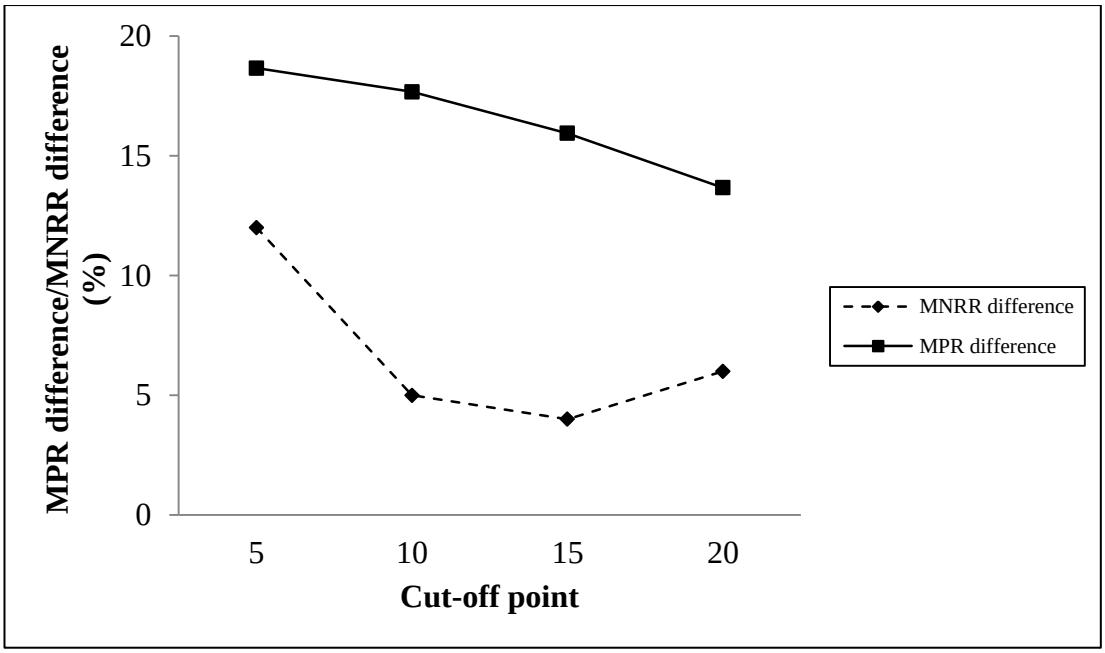

Figure 3.9: MPR differences and MNRR differences between CFQs and IFQs at four cut-off points for Google

Figure 3.9 presents MPR differences as well as MNRR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 for Google Web search engine. Google presents better performance on MNRR, since Google has the lowest MNRR difference compared to its MPR difference at all cut-off points. Google presented the highest MPR difference and MNRR difference at cut-off point 5. While Google displayed the lowest MPR difference at cut-off point 20, the lowest MNRR difference was at cut-off point 15. While MPR difference decreased from cut-off point 5 till cut-off point 20, MNRR difference was decreased as well as increased between cut-off point 5 and 20.

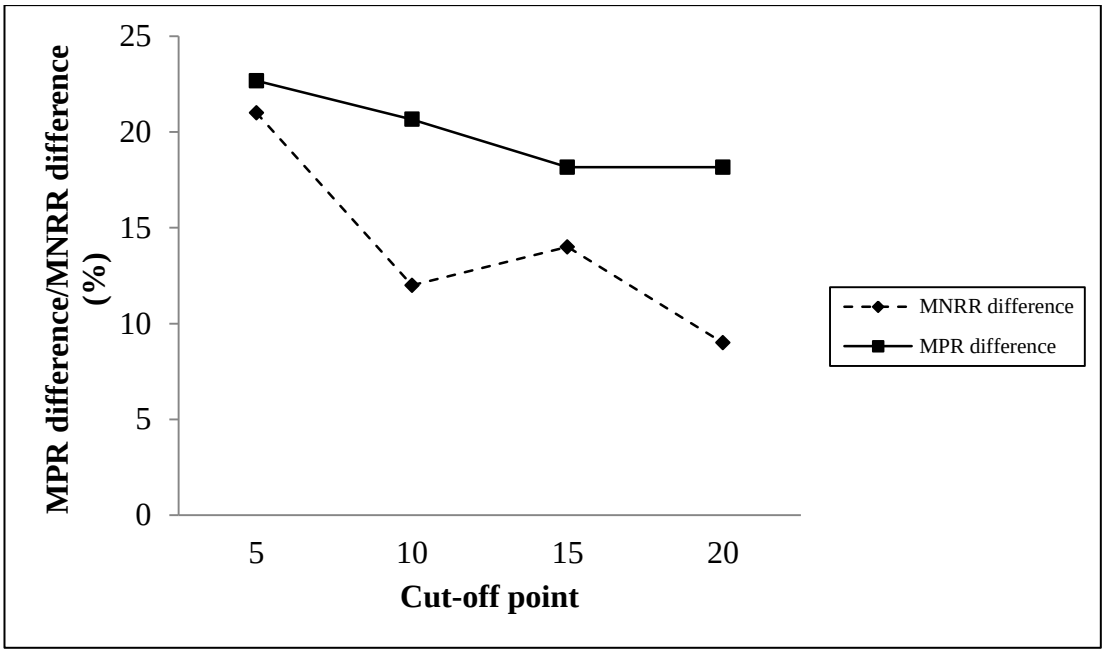

Figure 3.10: MPR differences and MNRR differences between CFQs and IFQs at four cut-off points for Bing

Figure 3.10 presents MPR differences as well as MNRR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 for Bing Web search engine. Bing presents better performance on MNRR, since Bing has the lowest MNRR difference compared to its MPR difference at all cut-off points. Bing presented the highest MPR difference and MNRR difference at cut-off point 5. While Bing displayed the lowest MPR difference at both cut-off point 15 and 20, the lowest MNRR difference was at cut-off point 20. While MPR difference decreased in general from cut-off point 5 till cut-off point 20, MNRR difference was decreased as well as increased between cut-off point 5 and 20.

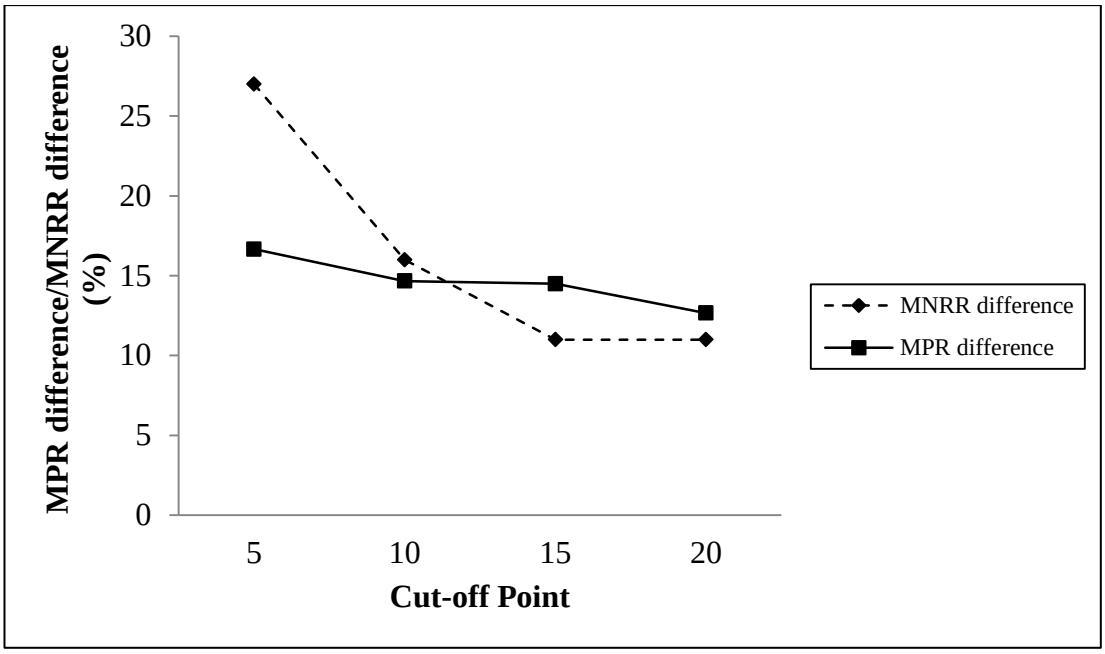

Figure 3.11: MPR differences and MNRR differences between CFQs and IFQs at four cut-off points for Yahoo

Figure 3.11 presents MPR differences and MNRR differences between CFQs and IFQs at cut-off points 5, 10, 15, and 20 for Yahoo Web search engine. Yahoo presents less performance on MNRR at cut-off points 5 and 10, since Yahoo has the highest MNRR difference compared to its MPR difference at cut-off points 5 and 10. Besides, at cut-off points 15 and 20, the engine was presented better performance on MNRR, since Yahoo has the lowest MNRR difference compared to its MPR difference at those cut-off points. Yahoo presented the highest MPR difference and MNRR difference at cut-off point 5. While Yahoo displayed the lowest MPR difference at cut-off point 20, the lowest MNRR difference was at both cut-off point 15 and 20. While MPR difference was the same at cut-off points 10 and 15 and MNRR difference was the same at cut-off points 15 and 20, for the rest, both MPR difference and MNRR difference were decreased.

3.4 Summary and Discussion

As shown in Figures 3.1 to 3.3, Google, Bing, and Yahoo retrieved their highest MPRs at cut-off point 5 for both CFQs and IFQs; however, when the cut-off point was increased, their MPRs decreased till cut-off point 20. Therefore, for CFQs as well as IFQs, the Web search engines retrieved their lowest MPRs at cut-off point 20. At all cut-off points, all the Web search engines have better MPRs for CFQs compared to their MPRs for IFQs. However, when the cut-off point was increased, for all the engines, the MPR difference between CFQs and IFQs decreased in general (Figure 3.4). Furthermore, Yahoo has the lowest MPR differences at all cut-off points.

As shown in Figures 3.5 to 3.7, Google, Bing, and Yahoo retrieved their lowest MNRRs at cut-off point 5 for IFQs; however, when the cut-off point was increased, their MNRRs also increased till cut-off point 20. Therefore, for IFQs, the Web search engines retrieved their highest MNRRs at cut-off point 20. Besides, for CFQs: Bing retrieved its lowest MNRR at cut-off point 10, Yahoo retrieved its lowest MNRR at cut-off point 15, and Google retrieved its lowest MNRR at cut-off points 10 and 15; and Google retrieved its highest MNRR at cut-off point 20, Yahoo retrieved its highest MNRR at cut-off point 5, and Bing retrieved its highest MNRR at cut-off points 5 and 15. At all cut-off points, all the Web search engines have better MNRRs for CFQs compared to their MNRRs for IFQs. However, when the cut-off point was increased, for all the engines, the MNRR difference between CFQs and IFQs decreased in general (Figure 3.8). Furthermore, Google has the lowest MNRR differences at all cut-off points.

As shown in Figures 3.9 to 3.11, while the search engines displayed their highest MPR differences and MNRR differences at cut-off point 5, their lowest MPR differences and MNRR differences were at cut-off point(s) 15 and/or 20. At all cut-off points, all the Web search engines have the lowest MNRR differences compared to their own MPR differences. While Google's and Bing's MPR differences generally decreased from cut-off point 5 till cut-off point 20, MNRR differences were decreased as well as increased between cut-off point 5 and 20. Besides, Yahoo's MPR and MNRR differences were generally decreased from cut-off point 5 till cut-off point 20.

Chapter 4

CONCLUSION

In this study, we investigated how the Web search engines handle Turkish words which are frequently wrongly spelled and/or pronounced with their own particular wrong form(s).

All Web search engines have better performance on retrieving relevant documents for CFQs at all cut-off points. Furthermore, the least negative effectiveness of using IFQs on retrieving relevant documents was presented by Yahoo, followed by Google, and then, Bing.

All Web search engines have better performance on showing relevant documents retrieved for CFQs in the top ranks at all cut-off points. Furthermore, the least negative effectiveness of using IFQs on retrieving relevant documents was presented by Google, followed by Bing, and then, Yahoo.

When considering the effect of using IFQs on each of the Web search engines, in general, Google, Bing, and Yahoo were presented better performance on displaying relevant documents retrieved in higher ranks than retrieving relevant documents.

It seems that using IFQs affected the retrieval effectiveness of the Web search engines badly. Web search engines are needed to be improved for handling the

wrongly spelled and/or pronounced Turkish words and during this period, using CFQs should be preferred by the users.

REFERENCES

[1] Internet Live Stats, Internet Users by Country as of July 2014, http://www.internetlivestats.com/internet-users-by-country/ (Accessed on April 2014).

[2] Wikimedia Foundation, Inc., List of languages by number of native speakers, http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers (Accessed on April 2014).

[3] Can, F., Koçberber, S., Balçık, E., Kaynak, C., Öcalan, H., & Vursavaş, O. (2008). Information Retrieval on Turkish Texts, *Journal of the American Society for Information Science and Technology*, Vol. 59, No. 3, 407-421.

[4] Turkish Language Association, Sıkça Yapılan Yanlışlara Doğrular Kılavuzu as of July 2010, http://www.tdk.gov.tr/index.php?option=com_yanlis (Accessed on 27 Oct 2013).

[5] Google Inc., http://www.google.com

[6] Microsoft, http://www.bing.com

[7] Yahoo Inc., http://search.yahoo.com

[8] Bar-Ilan, J., & Gutman,T. (2005). How do search engines respond to some non-English queries?, *Journal of Information Science*, vol. 31, no. 1, 13-28.

[9] 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. Available at: www.cais-acsi.ca/proceedings/2004/moukdad_2004.pdf.

[10] 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*, Springer, Vol. 12, No. 3, 352-379.

[11] Demirci, R., G., Kişmir, V., & Bitirim, Y. (2007). An Evaluation of Popular Search Engines on Finding Turkish Documents, *International Conference on Internet and Web Applications and Services (ICIW)*, 61, IEEE Computer Society, Morne, Mauritius.

[12] Türker, Ş., & Bitirim Y. (2015). A Study on the Effect of Using the Wrongly Spelled and/or Pronounced Turkish Words on Web Search Engines, *International Journals of Scientific Knowledge, Computing and Information Technology*, Vol. 5, No. 8, 1-8.

[13] eBizMBA Inc., Top 15 Most Popular Search Engines as of January 2015, http://www.ebizmba.com/articles/search-engines. (Accessed on Nov 2013)

[14] Wikimedia Foundation, Inc., Türkçede sık yapılan hatalar, http://tr.wikipedia.org/wiki/Vikipedi:T%C3%BCrk%C3%A7ede_s%C4%B1k_yap%C4%B1lan_hatalar (Accessed on Nov 2013).

[15] Bilgiustam, Türkçe de Sıklıkla Yanlış Kullandığımız Kelimeler, http://www.bilgiustam.com/turkcede-siklikla-yanlis-kullandigimiz-kelimeler/ (Accessed on Nov 2013)

[16] Msxlabs, Türkçede en çok yazım yanlışı yapılan sözcükler ve bunların doğru yazılımları nedir, http://www.msxlabs.org/forum/soru-cevap/252964-turkcede-en-cok-yazim-yanlisi-yapilan-sozcukler-ve-bunlarin-dogru-yazimlari-nedir.html (Accessed on Nov 2013).

[17] Milliyet Blog, Doğru ve Yanlış Kelimeler. (2012). http://blog.milliyet.com.tr/dogru-ve-yanlis-kelimeler--sozcukler--/Blog/?BlogNo=352240 (Accessed on Nov 2013).

[18] Blogcu, şiirlopedi, Yanlış Doğru Kelimeler. (2009). http://siirlopedi.blogcu.com/yanlis-dogru-kelimeler/4383920 (Accessed on Nov 2013).

[19] Eokul-Meb, Sıkça Yazımı Karıştırılan Sözcükler, http://www.eokul-meb.com/sikca-yazimi-karistirilan-sozcukler-tdk-34754/ (Accessed on Nov 2013).

[20] Uğurlu, A., Yanlış Yazılan Kelimeler Sözlüğü, http://www.allaturkaa.de/forum/index.php?page=Thread&threadID=29083 (Accessed on Nov 2013).

[21] Spink, A. & Jansen, B. J. (2004), *Web Search: Public Searching of the Web*, Berlin, Germany: Springer-Verlag.

[22] Yates, R. & Neto, B. (2011). *Modern Information Retrieval*, The Concepts and Technology behind Search, 2nd Edition, ACM Press Books, 75.

[23] 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.

[24] Bollmann, P., Jochum, R., Reiner, U., Weissmann, V., & Zuse, H. (1986). *Planung und durchf¨uhrung der retrievaltests*. In Schneider, H. (Ed.), Leistungsbewertung von Information Retrieval Verfahren (LIVE), 183-212.