



מכון טכנולוגי חולון הפקולטה לניהול טכנולוגיה

“האם מנוע החיפוש של גוגל הוא מוטה דפדפנים?”

עבודת גמר לשם מילוי חלקי של הדרישות לקבלת תואר מוסמך במדעים

מאת:

אסף גרימברג

בהנחיית:

ד"ר אייל בריל

תקציר מנהלים

1. מבוא

אחת התופעות המעניינות במנועי חיפוש אינטרנטיים היא שוני בתוצאות החיפוש של אותו המושג בעקבות פרמטרים שונים על המשתמש. שוני בתוצאות החיפוש מפריע, לעיתים קרובות, בקידום אתרים או בעקביות התוצאות עבור משתמשים באזורים שונים בעולם.

המחקר עוסק בשוני בתוצאות החיפוש בין דפדפנים על בסיס גאוגרפי. בחנו את מנוע החיפוש המקובל ביותר כיום: גוגל. מטרת המחקר היא שעל בסיס ממצאיו יהיה ניתן להבין האם באזורים גאוגרפים מסוימים תוצאות החיפוש שונות בין דפדפנים שונים, ועל כן, נוכל לאפיין כיצד לנהוג באזורים גאוגרפים מסוימים כדי לקדם אתרים בדפדפנים שונים.

מכאן, שאלת המחקר היא: "האם קיים שוני בדרוג תוצאות החיפוש בתלות בסוג הדפדפן במקום גאוגרפי כלשהו?"

השערת המחקר הינה: "קיים שוני בתוצאות חיפוש באזורים גאוגרפים שונים על פני דפדפנים מסוגים שונים". משתני המחקר הם:

- משתנה תלוי – מרחק בין תוצאות החיפוש חיפוש
- משתנה בלתי תלוי – מילת חיפוש, אזור גאוגרפי וסוג הדפדפן

2. סקר ספרות

מנוע חיפוש אינטרנטי הוא מערכת תוכנה אשר מיועדת לבצע חיפוש באינטרנט, שפירושו לחפש באינטרנט באופן שיטתי מידע ספציפי המצוין בשאלת חיפוש טקסטואלית. תוצאות החיפוש מוצגות בדרך כלל בשורת תוצאות, המכונות לעיתים קרובות דפי תוצאות של מנועי חיפוש.

בסקר הספרות התבססנו הן על מקורות אקדמיים והן על מקורות מקצועיים הסוקרים מאפיינים שונים של מנועי חיפוש אינטרנטיים בכלל ומנוע החיפוש של גוגל בפרט. בהתבסס על מקורות אלה וניסוי שערכנו ניתן להסיק על השוני בתוצאות החיפוש של מנוע החיפוש גוגל, בין דפדפנים שונים, באזורים גאוגרפיים שונים בעולם.

באמצעות מחקרים מסקר הספרות ניתן להגיע למסקנות אודות אופן החיפוש המיטבי תוך נטרול כל הפרמטרים המשפיעים על תוצאות החיפוש. ועל כן המסקנות הן:

- נדרש לבצע חיפוש בדפדפן שאינו מחובר לאף חשבון המקושר בדרך ישירה או עקיפה לגוגל (כגון שירותי המייל של גוגל Gmail).
 - נדרש לבצע חיפוש בדפדפן נקי מהיסטוריית חיפושים.
 - נדרש לבצע חיפוש בדפדפן נקי מעוגיות (Cookies) מעקב.
- התוכנה שנכתבה למטרת המחקר הותאמה לכל הפרמטרים הרשומים מעלה.

3. מתודולוגיה

המחקר בוצע על פני 8 מדינות שונות: ארה"ב, ברזיל, פריז, הודו, יפן, אוסטרליה, בחרין, ישראל. המטרה היא לבדוק האם יש שוני בין האזורים ובין דפדפנים שונים עבור מנוע החיפוש של גוגל. האזורים נבחרו על פי אפשרויות התמיכה הגלובלית הקיימות בשירותי הענן של Amazon (שירותי AWS) וישראל. באמצעות שירותי הענן של אמזון ניתן להקים שרתים ברחבי העולם בלחיצת כפתור ולהתחבר אליהם מרחוק על ידי תוכנת RDP. כך, ניתן למעשה לייצר סימולציית חיפוש על פני דפדפנים שונים במדינות מרוחקות.

המחקר כלל את שלושת הדפדפנים הפופולריים ביותר בשוק הדפדפנים כיום:

- Chrome של חברת Google
 - Internet Explorer של חברת Microsoft
 - Firefox של חברת Mozilla
- כאשר על כל דפדפן הורצו שאילתות חיפוש דומות עם פרמטרים דומים.

לאחר מחקר גישוש מוצלח שכלל חיפוש ידני מיפן (Tokyo, ע"י תוכנת VPN המדמה מיקום) של מילות המפתח: "Facebook" ו-"New York" על גבי דפדפני Firefox ו-Chrome ובמקביל ביצוע חיפוש זהה מארה"ב (Ohio, ע"י אותה תוכנת VPN).

תוצאות מחקר הגישוש אפשרו לנו להמשיך בניסוי והראו כי קיים שוני בין אזורים גיאוגרפיים ואפילו בין דפדפנים שונים, כאשר ביפן ישנה תוצאה חסרה בדפדפן Chrome הקיימת ב Firefox (ולהפך) עבור מילת החיפוש "New York".

מילות החיפוש שנבחרו היו סביב קטגוריה ספציפית אחת: "תינוקות" והיו בעלות מדגם רחב יחסית הכולל: newborn, infant formula, colic, bassinet, labor, baby dummy, teether, baby bottle, playpen, porridge, baby chair, baby car seat, stroller, bib, diaper rash, nanny, reflux, baby romper, baby carrier.

מהלך הניסוי:

- יצירת חשבון בשירות AWS של Amazon.
- הקמת 8 שרתי Windows Server 10 בכל אחד מהאזורים: אוהיו (ארה"ב), פריז (צרפת), מומביי (הודו), סידני (אוסטרליה), טוקיו (יפן), סאו פאולו (ברזיל), בחרין.
- המכונות היו מסוג t2.micro משיקולי תקציב (חינם למשך שנה ביצירת חשבון AWS חדש).

- שימוש במחשב אישי כדי להריץ את התוכנה בישראל. על המחשב האישי הותקנה מכונה וירטואלית חדשה, גם היא מעל מערכת ההפעלה Windows 10 (הפופולארית בשוק המחשבים כיום), זאת כדי שהחיפוש יבוצע ממחשב "חדש" ללא פרטים אישיים או מידע אישי כלשהו.
- כתיבת קוד בטכנולוגיית NodeJS אשר משתמש בכלי Open Source הנקרא Selenium המאפשר לדמות התנהגות של בנאדם על כל סוגי הדפדפנים (לחיצה, הקלדה ועוד).
- הרצת התוכנה על כל השרתים המוזכרים לעיל פעם ביום במשך שלושה ימים.
- לא היה שוני בין ההרצות עבור אוסף של הגדרות מסוים (דפדפן, מילת חיפוש, אזור גאוגרפי).
- בכל הרצה המחשב עלה מחדש, על כל הגדרותיו, ללא כל מידע קודם או היסטוריית חיפוש.
- בכל חיפוש הדפדפן נוקה מחדש כך שלא תהיה היסטוריית חיפוש או מידע מקדים על המשתמש.
- כלל התוצאות הגולמיות נשמרו בקבצי CSV עליהם בהמשך הורצו אלגוריתמים נוספים.

מרחב המדגם:

N = סוגי דפדפנים

M = מדינות

K = מילות חיפוש

$$\Omega = N * M * K = 3 * 8 * 20 = 480$$

חישוב המרחק בין תוצאות, תיאור האלגוריתם:

1. לכל מיקום גאוגרפי נאספו כ 10 תוצאות חיפוש עבור מילת חיפוש כלשהי וסוגי דפדפנים שונים.
2. עבור דפדפן מסוג כלשהו נבדק מה המרחק של תוצאות החיפוש, באותה המדינה, כנגד תוצאות החיפוש של דפדפן אחר.
3. הנ"ל נעשה על ידי הרצת אלגוריתם מרחק הבודק את דירוג התוצאות בין שני דפדפנים שונים.
4. אלגוריתם המרחק: ניתן ניקוד (מרחק) לדירוג תוצאות שונות על ידי הפרש אבסולוטי בין המיקומים של תוצאות החיפוש (לדוגמא: בעת חיפוש המילה "reflux", אם תוצאת חיפוש מסוימת, המיוצגת על ידי קישור לאתר מסוים, דורגה במיקום 3 בדפדפן Firefox אך במיקום 5 בדפדפן Chrome על ידי מנוע החיפוש של גוגל, הדבר היווה $5 - 3 = 2$ נקודות, כלומר המרחק הוא 2)
5. תוצאת חיפוש שאינה נמצאת בדפדפן אחד אך נמצאת באחר זיכתה ב 10 נקודות, כלומר המרחק הוא 10.
6. הפלט של האלגוריתם הנ"ל גם הוא נכתב לקובץ CSV.

קטגוריזציה של מרחקים ומילות חיפוש:

1. נעשה שימוש בתוכנת Weka על מנת למצוא את פיזור המרחקים על גבי כלל התוצאות.

2. התקבל הפיזור הבא:

i. 268 תוצאות קיבלו מרחקים: 0 – 13.1

ii. 127 תוצאות קיבלו מרחקים: 13.1 – 52.4

iii. 75 תוצאות קיבלו מרחקים הגדולים מ 52.4

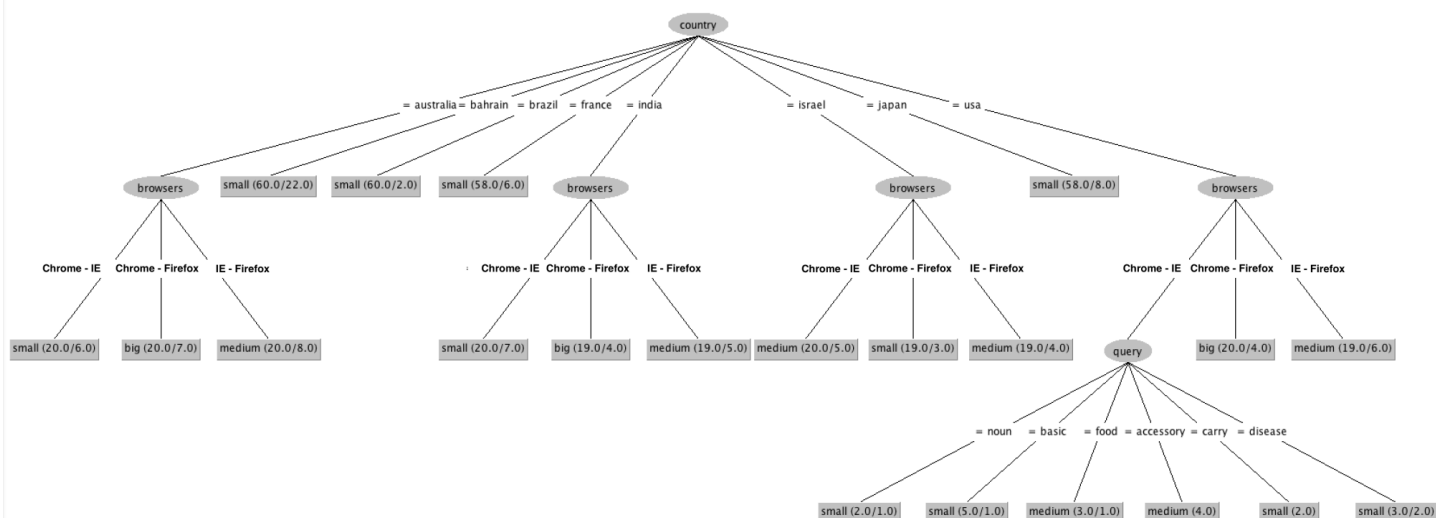
על כן, הוחלט לבצע קטגוריזציה של התוצאות על פי הפיזור למעלה, כך שכל קטגוריה מקבלת אות A,B,C בהתאמה.

בנוסף, מכיוון שכמות מילות החיפוש גדולה, גם אותה חילקנו ל 6 קטגוריות: מזון (B), מוצרי בסיס (C), אביזרים נלווים (E), מחלות ילדים (D), מוצרי הנגשה (F) ושמות עצם כלליים (A).

4. תוצאות

נעשה שימוש ברגרסיה לינארית על מנת להראות את נכונות המשתנים המסבירים (מילת חיפוש, סוג דפדפן, מדינה) עבור המשתנה המוסבר (מרחק). אך תוצאות הרגרסיה כללו R Squared נמוך מידי, כלומר לא ניתן להסביר שינוי של המשתנה המוסבר על ידי שינוי של המשתנים המסבירים. בנוסף, באופן ויזואלי הראנו את השפעת המשתנים על המרחק, ולכן בחרנו באלגוריתם עץ החלטות J48 כדי להציג עץ של המשתנים המסבירים (מדינה, מילת חיפוש, סוג דפדפן) והמשתנה המוסבר (מרחק). באופן זה, נוכל לקבל עבור כל מדינה את רמת ההשפעה של המשתנים המסבירים ואת ההסתברות לקבל מרחק גדול (שוני רב) בין תוצאות חיפוש. לבסוף, האלגוריתם חזה ב 89% את התוצאות במדינות בהן היה מרחק קטן (0-13) בין הדפדפנים, ב 55% את התוצאות במדינות בהן היה מרחק בינוני (13-52) בין הדפדפנים וב 58% את התוצאות במדינות בהן היה מרחק גדול (52 ומעלה) בין הדפדפנים.

לאחר שימוש באלגוריתם J48 בתוכנת Weka התקבל העץ הבא:



ניתן לראות כי:

- קיימות מדינות בהן יש הבדל בתוצאות החיפוש בין דפדפנים ללא קשר למילת החיפוש שנבחרה, והן: אוסטרליה, הודו וישראל.
 - קיימות מדינות בהן אין הבדל בתוצאות החיפוש בין סוגי דפדפנים, והן: בחריין, ברזיל, צרפת ויפן.
 - בארה"ב יש הבדל בין תוצאות החיפוש בין דפדפנים מסוגים שונים, כאשר עבור ההבדלים בין דפדפן Chrome ודפדפן Internet Explorer מושפעים גם ממילת החיפוש שנדגמה.
 - באוסטרליה, ארה"ב ובהודו ההבדל הגדול ביותר בין תוצאות החיפוש יהיה בין דפדפן Chrome לדפדפן Firefox, ואילו בישראל ההפך הוא הנכון.
- האלגוריתם הצליח לנבא נכון 75% מההחלטות שבקובץ נתונים. בנוסף, ניתן לראות שהמודל חוזה בקירוב של 89% כאשר המרחק בין תוצאות החיפוש הוא קטן, 55% כאשר המרחק הוא בינוני ובקירוב של 58% כאשר המרחק הוא גדול בין תוצאות החיפוש.

=== Summary ===

Correctly Classified Instances	353	75.1064 %
Incorrectly Classified Instances	117	24.8936 %
Kappa statistic	0.5481	
Mean absolute error	0.2396	
Root mean squared error	0.3601	
Relative absolute error	62.2783 %	
Root relative squared error	82.1186 %	
Total Number of Instances	470	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.892	0.312	0.791	0.892	0.839	0.599	0.843	0.854	small
	0.551	0.114	0.642	0.551	0.593	0.460	0.733	0.522	medium
	0.587	0.038	0.746	0.587	0.657	0.606	0.853	0.572	big
Weighted Avg.	0.751	0.215	0.744	0.751	0.743	0.563	0.815	0.719	

=== Confusion Matrix ===

a	b	c	← classified as
239	22	7	a = small
49	70	8	b = medium
14	17	44	c = big

5. סיכום

במחקר זה נבדק האם קיים שוני בדרוג תוצאות החיפוש בין דפדפנים מסוגים שונים במקום גאוגרפי כלשהו. נדגמו 8 מדינות ברחבי העולם (אוסטרליה, ברזיל, ברייטן, ישראל, יפן, צרפת, ארה"ב, הודו), על גבי מנוע החיפוש של גוגל. בכל מדינה, הורצה תוכנת חיפוש אוטומטית ללא היסטוריה או מידע מקדים על המשתמש לפי 20 מילות חיפוש בנושא תינוקות. נבדקו 3 הדפדפנים הפופולאריים ביותר בשוק כיום. מתוצאות המחקר ניתן להסיק כי קיים שוני בדרוג תוצאות החיפוש בתלות בסוג הדפדפן במדינות אוסטרליה, הודו, ישראל וארה"ב. בארה"ב, בניגוד לשאר המדינות בהם יש שוני בדירוג תוצאות החיפוש, ראינו כי השוני בתוצאות החיפוש בין דפדפן Chrome ל Internet Explorer מוסבר גם באמצעות מילת החיפוש.

ראינו בתוצאות המחקר כי אין שוני בדרוג תוצאות החיפוש במדינות יפן, ברייטן, ברזיל וצרפת בין סוגי הדפדפנים השונים. הדבר מעיד על כך שבמדינות מסוימות בעולם דירוג תוצאות החיפוש בין דפדפנים יהיה זהה לעומת מדינות אחרות בהן הדירוג יהיה שונה.

בנוסף, רואים בתוצאות המחקר כי ההבדל הגדול ביותר בין תוצאות החיפוש הוא בין דפדפן Chrome לדפדפן Firefox באוסטרליה, ארה"ב והודו. בניגוד לכך, בישראל, ההפך הוא הנכון (ההבדל בין תוצאות החיפוש בין דפדפן Chrome לדפדפן Firefox הוא הקטן ביותר).

המסקנה למטרת קידום אתרים היא כי יש שוני בדרוג תוצאות החיפוש בין דפדפנים במדינות מסוימות בעולם, ולכן, כאשר עושים אופטימיזציה למנועי חיפוש, יש לבדוק כי האופטימיזציה מצליחה על כל סוגי הדפדפנים הפופולאריים בשוק עבור כל מדינה שרלוונטית לקידום האתר. הוכחנו כי קיים הבדל בין תוצאות החיפוש במדינות מסוימות על פני סוגי דפדפנים שונים ולכן מנוע החיפוש של גוגל הוא מוטה דפדפנים.

המלצות להמשך:

1. הרצת תוכנת החיפוש האוטומטית על מרחב מדגם גדול יותר:
 - i. כל המדינות בעולם (יש 193 מדינות בעולם).
 - ii. מילות חיפוש מתחומים נוספים הרלוונטיים למדגם רחב של אתרי אונליין כגון: ספורט, בגדים, מוצרי אלקטרוניקה ועוד.
 - iii. הוספת דפדפן Safari לתוכנת החיפוש באמצעות שימוש במחשבי Mac של חברת Apple.
2. מומלץ לבדוק האם מערכת ההפעלה, עליה רצה התוכנה, משפיעה על תוצאות המחקר. במחקר זה בדקנו את מערכת ההפעלה Windows 10 בלבד (הפופולארית ביותר בשוק מערכות ההפעלה כיום).

תוכן עניינים

נושאים

11	1. מבוא
11	1.1 רציונל
11	1.2 מטרת המחקר
12	1.3 שאלת המחקר
12	1.4 משתנים תלויים ובלתי תלויים, השערות המחקר
13	1.5 חשיבות המחקר במישור התיאורטי ובמישור הפרקטי
14	2. סקר ספרות
14	2.1 הקדמה
14	2.2 הגדרת מנוע חיפוש אינטרנטי
16	2.3 אלגוריתמים של מנועי חיפוש והשפעתם על תוצאות החיפוש
17	2.4 הבדלים בין תוצאות חיפוש על גבי דפדפנים שונים
18	2.5 הבדלים בין תוצאות חיפוש על בסיס גאוגרפי
19	2.6 סיכום
20	3. מתודולוגיה
20	3.1 מערך המחקר, אוכלוסיית המחקר
22	3.2 הפרדיגמה המחקרית – מחקר כמותי
23	3.3 תכנון מחקר גישוש
25	3.4 המדגם
25	3.5 שיטת המחקר
27	3.6 כלי איסוף הנתונים
28	3.7 ניתוח הנתונים
31	4. תוצאות
31	4.1 ניתוח תוצאות על ידי רגרסיה לינארית
40	4.2 ניתוח תוצאות על ידי אלגוריתם עץ J48
44	5. סיכום ודיון
45	6. רשימת מקורות
47	7. נספחים

טבלאות

20	טבלה 1: אזורים נתמכים על ידי AWS
22	טבלה 2: נתח שוק של דפדפנים
23	טבלה 3: תוצאות מחקר גישוש
31	טבלה 4: דוגמא לפלט אלגוריתם הקטגוריזציה
33	טבלה 5: תוצאות מתאם פירסון
34	טבלה 6: פלט תוכנת R - תוצאות מבחן T שוניות שוות
34	טבלה 7: פלט תוכנת R - תוצאות מבחן T שוניות שונות
34	טבלה 8: תוצאות מבחן T מאוחדות
36	טבלה 9: תוצאות רגרסיה לינארית

גרפים

26	גרף 1: נתח שוק מערכות הפעלה
31	גרף 2: פיזור התוצאות למרחקים
32	גרף 3: פיזור מרחק לאחר קטגוריזציה
32	גרף 4: פיזור מילות חיפוש לאחר קטגוריזציה

איורים

14	איור 1: דוגמא לדף תוצאות SERP של מנוע החיפוש גוגל
23	איור 2: ניקוי נתוני משתמש בדפדפן Chrome
30	איור 3: דוגמא לעותק מקובץ הנתונים עבור תוכנת R
30	איור 4: דוגמא לעותק מקובץ הנתונים עבור תוכנת Weka
40	איור 5: תוצאות הרצת אלגוריתם J48
42	איור 6: עץ ההחלטות של אלגוריתם J48

רשימת קיצורים ומילון מושגים

קיצורים:

SERP	Search Engine Results Page
VPN	Virtual Private Network
AWS	Amazon Web Services
SEO	Search Engine Optimization
CSV	Comma Separated Values
RDP	Remote Desktop Protocol

מושגים:

1. **Open Source** - קוד פתוח משמש בעולם התוכנה לציון תוכנה שקוד המקור שלה פתוח ונגיש לכל מי שחפץ בו והוא חופשי לשימוש, לצפייה, לעריכת שינויים ולהפצה מחודשת לכל אחד ואחת.
2. **Page Ranking** – אלגוריתם בו משתמשת גוגל כדי לדרג אתרים בתוצאות החיפוש של מנוע החיפוש שלה.
3. **NodeJS** – סביבת קוד פתוח המריצה קוד JavaScript על כל פלטפורמה שאינה דפדפן אינטרנט.
4. **Confusion Matrix** - בתחום למידת מכונה ובמיוחד בעיית הסיווג הסטטיסטי, מטריצת שגיאות, היא פריסת טבלה ספציפית המאפשרת הדמיית ביצועי אלגוריתם, בדרך כלל במערך למידה מפוקח. כל שורה של המטריצה מייצגת את המופעים בסיווג החזוי ואילו כל עמודה מייצגת את המופעים בסיווג בפועל.

1. מבוא

1.1 רציונל

מנועי חיפוש אינטרנטיים, כגון מנוע החיפוש של גוגל, משתמשים באלגוריתמים מורכבים ומסווגים על מנת למצוא תוצאות הקרובות ביותר למושג אותו מחפש המשתמש. אלגוריתמים אלה משתמשים במגוון פרמטרים אודות המשתמש שלא בהכרח ידועים או חשופים. פרמטרים אלה כוללים בין היתר את מיקום משתמש, היסטוריית חיפוש, גיל, מין, מידע על אינטרסים ועוד. אחת התופעות המעניינות היא שוני בתוצאות החיפוש של אותו המושג בעקבות פרמטרים שונים על המשתמש. שוני בתוצאות החיפוש הוא מאפיין נדרש המאפשר התאמה אישית של תוצאות החיפוש למשתמש הדורש אותן. הדבר מאפשר גמישות מרבית והצגת נתונים רלוונטיים עבור כל משתמש. אך, למרות היתרונות הברורים של תוצאות חיפוש מותאמות אישית, שוני בתוצאות החיפוש על בסיס סוג הדפדפן מפריע לעיתים בקידום אתרים או בעקביות התוצאות עבור משתמשים באזורים שונים בעולם. משתמש המחפש מוצר כלשהו באינטרנט בדפדפן כלשהו יקבל תוצאות אשר פחות רלוונטיות מאותו חיפוש בדפדפן אחר. כך גם אתרים רלוונטיים יופיעו בחיפוש בדפדפן אחד אך לא בדפדפן אחר, דבר היכול להשפיע לרעה על קידום אתרים אלה.

1.2 מטרת המחקר

המחקר עוסק בשוני בתוצאות החיפוש בין דפדפנים על בסיס גאוגרפי. בחנו האם מנוע החיפוש של גוגל מבדיל בין סוגי דפדפנים שונים כאשר מציג את תוצאותיו עבור שאילתת חיפוש כלשהי במיקום גאוגרפי מסוים. בדקנו האם במדינות מסוימות בעולם יש שוני בין תוצאות החיפוש כאשר אותו החיפוש מבוצע מכמה דפדפנים שונים. על בסיס מחקר זה יהיה ניתן להבין האם באזורים גאוגרפים מסוימים תוצאות החיפוש שונות בין דפדפנים שונים ועל כן נוכל לאפיין כיצד לנהוג באזורים גאוגרפים מסוימים כדי לקדם אתרים בדפדפנים שונים. חברות העוסקות בקידום אתרים מבצעות אופטימיזציה של תוצאות החיפוש על ידי מידע מקדים על אופן הפעולה של מנועי החיפוש. למשל, מנוע החיפוש של גוגל מדרג אתרים בתוצאות החיפוש על פי מודל הנקרא Page Ranking. המודל למעשה סופר את כמות ההפניות שיש לאתר אינטרנט מסוים מתוך אתרי אינטרנט אחרים. ככל שלאחר אינטרנט יש יותר הפניות מאתרים אחרים סיכויי להופיע ראשון בתוצאות החיפוש הם גבוהים יותר. ולכן, כדי לקדם אתר באופן מיטבי נדרש להפנות כמה שיותר אתרים אל האתר אותו רוצים לקדם. החברות הנ"ל עושות זאת על ידי הפצה של תכנים, באתרים שונים ומגוונים, המפנים לאתר המבוקש. תכנים אלה יכולים להיות כתבות נושא באתרי תוכן וחדשות או "פוסטים" ברשתות חברתיות כגון פייסבוק, אינסטגרם ועוד. שוני בין תוצאות החיפוש על פני דפדפנים שונים הוא מידע מקדים בו נדרש להתחשב בתהליך קידום אתרים.

1.3 שאלת המחקר

האם קיים שוני בין תוצאות החיפוש של מנוע החיפוש גוגל על פני דפדפנים שונים במדינות שונות בעולם?
נבחן את השוני על ידי מידול "מרחק" בין תוצאות אשר יפורט בהמשך.

1.4 משתנים תלויים ובלתי תלויים

1. משתנה תלוי – מרחק בין תוצאות החיפוש חיפוש
 2. משתנה בלתי תלוי – מילת חיפוש, אזור גיאוגרפי וסוג הדפדפן
- מרחק בין תוצאות חיפוש מייצג את השוני בדרוג של תוצאות חיפוש בין דפדפן אחד למשנהו. בהמשך נגדיר כיצד חושב המרחק הנ"ל.

1.5 השערת המחקר

ההנחה היא שסוג הדפדפן מכיל שני נתונים עיקריים אודות המשתמש אותם מנועי החיפוש לוקחים בחשבון באלגוריתם החיפוש שלהם ובהצגת תוצאות החיפוש:

1. סוג הדפדפן מייצג את קבוצת המשתמשים המבצעים בו חיפוש. כך לדוגמא ניתן להניח כי משתמשים טכנולוגיים ישתמשו בדפדפן פיירפוקס או אופרה, הדורשים התקנה נפרדת ואינם מגיעים כברירת מחדל במחשבים האישיים, לעומתם סביר להניח כי משתמשים שאינם טכנולוגיים ישתמשו בדפדפן אשר הותקן להם מבעוד מועד על המחשב (אינטרנט אקספלורר במחשבי ווינדוס של מיקרוסופט או דפדפן ספארי במחשבי מק של אפל). תיוג של המשתמשים הנ"ל מאפשר להציג להם נתונים שונים רק על פי סוג הדפדפן בו הם משתמשים.
2. בהסתמך על הנאמר לעיל, בסוגי דפדפן שונים מתבצעים אומנם חיפושים זהים אך הרלוונטיות של תוצאות החיפוש עבור המשתמשים היא שונה. כך ניתן לאסוף מידע אודות התעניינות בתוצאות חיפוש מסוימות של משתמשים שונים, על פני סוגי דפדפנים שונים. על פי מידע היסטורי זה, ניתן לדרג באופן שונה את תוצאות החיפוש על פי סוג הדפדפן בלבד, ובכך לקבל תוצאות שונות בין סוגי דפדפנים שונים.

ומכאן השערת המחקר היא: קיים שוני בין תוצאות החיפוש במדינות שונות על פני דפדפנים מסוגים שונים.

1.6 חשיבות המחקר במישור התיאורטי ובמישור הפרקטי

היכולת לאפיין כי קיים שוני בתוצאות חיפוש בין דפדפנים עבור אזורים גאוגרפים שונים יכולה לאפשר בעתיד ידע מקדים אותו ניתן לנצל בעת קידום אתרים במנועי חיפוש, כגון מנוע החיפוש של גוגל.

חברות העוסקות בקידום אתרים, יכולות על ידי מחקר זה ומחקרי המשך להבין באילו מדינות קיים שוני בתוצאות חיפוש על פני סוגי דפדפנים שונים. במדינות אלה נדרש לבצע בדיקה עבור כל דפדפן בנפרד ולהבחין האם יש לבצע אופטימיזציה שונה או מותאמת לדפדפנים השונים. באמצעות שינויי פרמטרים של האתר המקודם ניתן להבין כיצד לקדם אתרים על פני כל סוגי הדפדפנים הפופולאריים כיום בשוק.

2. סקר ספרות

2.1 הקדמה

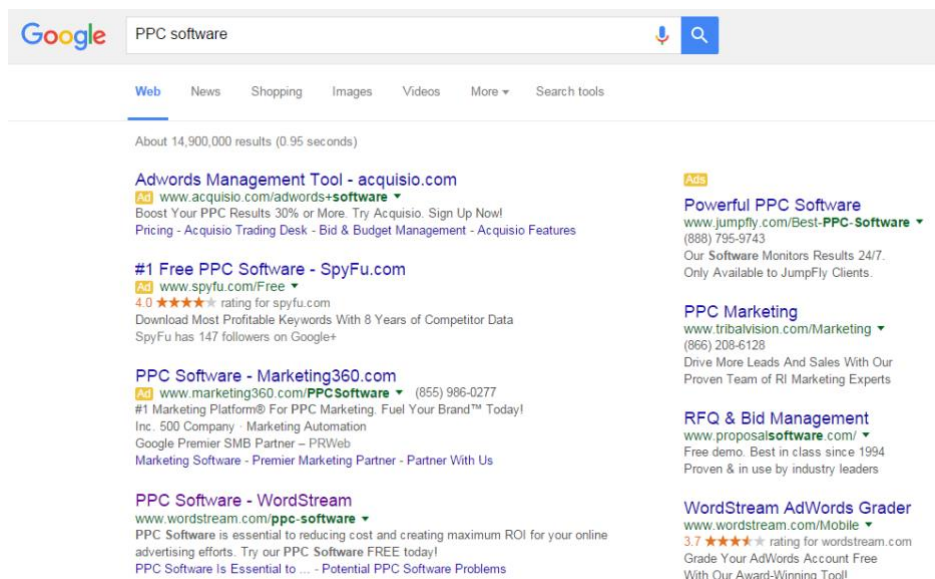
בפרק זה סקרנו ספרות מקיפה המתבססת הן על מקורות אקדמיים והן על מקורות מקצועיים הסוקרים מאפיינים שונים של מנועי חיפוש אינטרנטיים בכלל ומנוע החיפוש של גוגל בפרט. בהתבסס על מקורות אלה וניסוי שערכנו ניתן יהיה להסיק על השוני בתוצאות החיפוש של מנוע החיפוש גוגל, בין דפדפנים שונים, באזורים גיאוגרפיים שונים בעולם. סקרנו כמה נושאים מרכזיים:

1. הגדרת מנוע חיפוש אינטרנטי
2. אלגוריתמים של מנועי חיפוש והשפעתם על תוצאות החיפוש.
3. הבדלים בין תוצאות חיפוש על גבי דפדפנים שונים.
4. הבדלים בין תוצאות חיפוש על בסיס גאוגרפי.
5. סיכום

2.2 הגדרת מנוע חיפוש אינטרנטי

מנוע חיפוש אינטרנטי הוא מערכת תוכנה אשר מיועדת לבצע חיפוש באינטרנט, באופן שיטתי, מידע ספציפי המצוין בשאלת חיפוש טקסטואלית. תוצאות החיפוש מוצגות בדרך כלל בשורת תוצאות, המכונות לעתים קרובות דפי תוצאות של מנועי חיפוש (SERP).

איור 1: דוגמא לדף תוצאות SERP של מנוע החיפוש גוגל



מקור: <https://www.wordstream.com/serp>

בראשיתו של האינטרנט היה קושי עצום במציאת המידע הרלוונטי מתוך "ים" המידע הקיים בו. עולם האינטרנט היווה ממשק ידידותי למשתמש אשר "סחף" אליו משתמשים רבים וספקי מידע (Chu and Rosenthal, 1996) כדי לפתור בעיה זו, קמו עשרות חברות (לדוגמה: חברת "גוגל") אשר פיתחו מנועי חיפוש אינטרנטיים המאפשרים להכניס אליהם כקלט מילות חיפוש והן יציגו תוצאות רלוונטיות למשתמש כגון: אתרים, מאמרים, כתבות ועוד. מנועי חיפוש אלה מאפשרים חיפוש מידע באינטרנט ע"י תהליך מקדים הנקרא "Crawling". התהליך מתבצע באמצעות תוכנה אשר קוראת תוכן של עמודי אינטרנט ו"מאנדקסט" אותם (Indexing, שומרת אותם במבנה שיהיה קל לחפשם מאוחר יותר). התוכנה עוברת בין עמודי אינטרנט ידועים מראש או בין עמודים אשר ניתן לגשת אליהם מהפניות (קישורים) בעמודים אחרים. בעזרת תהליך ה"Crawling" המקדים, ניתנת למשתמשים האפשרות לחפש בצורה מהירה ויעילה מידע ממוקד מתוך כלל המידע הקיים באינטרנט (Brin, 1999 - ממקימי חברת "גוגל").

יעילות החיפוש של מנועי חיפוש מאפשרת גם חיפוש של מקורות ביבליוגרפים אקדמיים מתקדמים. ישנה הנחה כי בעתיד יחליפו מאגרי מידע אינטרנטיים את מאגרי המידע הביבליוגרפים האקדמיים המסורתיים, כמקורות מידע למחקרים מתקדמים. מנועי החיפוש מציגים תוצאות בצורה מהירה ונוחה יותר מאשר מאגרי מידע ביבליוגרפים אקדמיים, אך עקרונות ההיגיון הבוליאני וייצוא תקצירים עדיין מהווה מכשול. נדרש לשפר את איכות המאמרים ואופן החיפוש במאגרי מידע ביבליוגרפים, מתוך מחשבה שאם הדבר לא יקרה מנועי החיפוש אכן יצליחו להוות חלופה הוגנת (Bates, Best, Mcquilkiln and Taylor, 2017). בהתבסס על הנאמר לעיל, ניתן להסיק כי למידע הנמצא באינטרנט יש חשיבות רבה ולמנועי החיפוש יש השפעה רבה על המשתמשים הצורכים מידע זה. בעקבות כך, בדקנו את השוני בין תוצאות החיפוש, בין דפדפנים שונים, באזורים שונים בעולם ובכך לדעת האם מנועי החיפוש עושים שימוש בפרמטרים אלה לקביעת התוצאות אשר יוצגו למשתמש.

2.3 אלגוריתמים של מנועי חיפוש והשפעתם על תוצאות החיפוש

2.3.1 אלגוריתם של מנועי חיפוש

אלגוריתם של מנועי חיפוש הינו פורמולה מתמטית, המקבלת מילות מפתח כקלט ומחזירה תוצאות חיפוש רלוונטיות כפלט. זה נעשה ע"י התאמת מילות המפתח למאגר של תוצאות המאוחסן במסד נתונים. המאגר, בדרך כלל, כולל תוכן של אתרים, קשרים בין אתרים ופופולריות של אתרים ברחבי האינטרנט (זה נעשה ע"י תהליך ה" Crawling" שהוסבר קודם לכן). האופן בו מוגדרת הפורמולה המתמטית משתנה בין מנועי החיפוש. בנוסף, מנועי חיפוש יכולים להשתמש בפרמטרים נוספים שיוכנסו כקלט לאלגוריתם, ביניהם: אתרים פופולריים של המשתמש המחפש, זמן שהייה של משתמשים באתרים שונים, מיקום גאוגרפי ועוד (Cassidy, 2017).

באלגוריתמי חיפוש מתווסף מודל חשוב של דירוג אתרים (Ranking). קישורים של אתרים רבים המפנים לאתר מסוים, ללא קשר לתוכן של אותם האתרים, מעלה את דירוג האתר אליו הקישורים מפנים. כלומר, הסיכוי של אתר להופיע במיקום גבוה בתוצאות החיפוש תלוי, בין היתר, במספר האתרים המפנים אליו (Bar-Ilan, 2006).

בנוסף לכך, אתרים שתוכנם דומה אחד לשני, ידורגו גם כן גבוה, כלומר, יופיעו במיקום גבוה יותר בתוצאות החיפוש. ומכאן שדירוג אתרים נעשה גם על פי הדמיון בתוכן של אתרים. ככל שיש יותר אתרים הדומים בתוכנם אחד לשני, כך הדירוג שלהם בעת חיפוש מילות מפתח הקשורות לתוכן זה יהיה גבוה יותר (Brin, 2005). הדבר שונה במהות מאלגוריתם אשר יגדיל סיכויי אתר להופיע ראשון בתוצאות החיפוש על פי כמות האתרים המפנים אליו.

2.3.2 השפעת האלגוריתם על תוצאות החיפוש

מנועי חיפוש מאופיינים באלגוריתמים מורכבים שעוזרים לדרג ולהציג תוצאות חיפוש רלוונטיות למשתמש. מאחורי מנועי החיפוש עומדות חברות ענקיות כגון גוגל, מיקרוסופט, יאהו ועוד, אשר מנתחות ומעבדות מידע עצום על משתמשים ומתכננות אלגוריתמים מדויקים אשר משפיעים על אופן הצגתן של תוצאות החיפוש.

דוגמא לאלגוריתם כזה הוא אלגוריתם Hiltop של גוגל. מטרת האלגוריתם למצוא אתרים מהימנים אותם גוגל יכולה לדרג. תחילה האלגוריתם מוצא אתרים פופולריים אשר ממוקדים בנושא מסוים אך מובילים ע"י קישורים פנימיים לאתרים אחרים. האלגוריתם מדרג אתר על פי כמות האתרים האחרים המפנים אליו. בכך, כאשר המשתמש מחפש מילות מפתח כלשהן, התוצאות שיוצגו בפניו הן לא רק המתאימות ביותר למילות המפתח, אלא גם התוצאות שאתריהם מדורגים הכי גבוה במסד הנתונים של גוגל. מכאן שאלגוריתמים שונים, של מנועי חיפוש, משפיעים על תוצאות החיפוש וסדר הופעתן בדרכים מגוונות (Montti, 2018).

דבר מעניין נוסף הוא שניתן לחזות תוצאות חיפוש של מילת חיפוש כלשהי על פי פרמטרים ידועים של המשתמש. במחקר שניסה לחקות את אלגוריתם החיפוש של גוגל (Ao-Jan, Hu, Kuzmanovic)

(and Cheng-Kok, 2010), הצליחו החוקרים לזהות מראש יותר מ-9 תוצאות חיפוש מתוך 10 תוצאות החיפוש של גוגל עבור 77% ממושגי החיפוש שנוסו. כלומר, החוקרים הצליחו לזהות מהם הפרמטרים אשר ישפיעו על תוצאות החיפוש של גוגל. מאמר זה מאפשר לאתרים לאפיין עצמם על פי פרמטרים אלה מראש ובכך לקדם את עצמם בצורה משמעותית בתוצאות החיפוש של גוגל. אפיון הפרמטרים של אתר מסוים על פי אלגוריתמים של מנועי חיפוש נקרא Search Engine Optimization או בקיצור SEO והוא אחד ממנועי הקידום החשובים ביותר של אתרים בימינו.

בנוסף לאופי הטבעי של האלגוריתמים שתוארו לעיל, גוגל, למשל, מורידה אתרים מתוצאות החיפוש באופן מוחלט ומכוון, ללא קשר לדירוג שלהם, דבר המשפיע מאוד על תוצאות החיפוש ואופן קבלת ההחלטות. הדבר התגלה על ידי השוואה בין תוצאות חיפוש באמריקה אל מול תוצאות חיפוש ביפן של אותם המושגים (Yoshiura and Moroi, 2012). כלומר, מלבד תוצאות האלגוריתם הטבעיות, ישנה גם התערבות ידנית של החברות הגדולות בהצגת תוצאות החיפוש.

השפעת האלגוריתמים אינה רק על אופן ההצגה של תוצאות החיפוש, אלא גם על הביטחון והאמינות של המשתמשים, אשר בכל חיפוש מקבלים את מבוקשם באופן אופטימלי. לדוגמא, סטודנטים סמכו על תוצאות החיפוש של גוגל כמעט בעיוורון מוחלט והיו מוטים בחיפוש שלהם על פי דירוג תוצאות החיפוש. התגלה כי סטודנטים חקרו תקצירי מאמרים שהיו פחות רלוונטיים מאחרים רק בגלל שהופיעו בדירוג גבוה יותר בתוצאות החיפוש של גוגל (Pan, Hembrooke, 2007). מכאן אנו למדים כי אלגוריתמי חיפוש בעלי השפעה מכרעת על תרבות, חברה וגם על תעבורת משתמשים ברשת אל אתרים או מאמרים מסוימים.

2.4 הבדלים בין תוצאות חיפוש על גבי דפדפנים שונים

הגורמים המשפיעים על תוצאות החיפוש הם, בין היתר, היסטוריית החיפוש של המשתמש, סוג המכשיר עליו מתבצע החיפוש והתחברות של המשתמש לשירותים אחרים של מנוע החיפוש (כמו שירות Gmail של גוגל). הגורמים הנ"ל מאפשרים פרסונליזציה של תוצאות החיפוש, כלומר הצגת תוצאות חיפוש מותאמות אישית למשתמש. מכאן ששימוש שונה במנוע חיפוש כלשהו (היסטוריית גלישה שונה או חיבור לשירותים שונים), על גבי דפדפנים שונים, עלול להוביל לתוצאות שונות (McEvoy, 2015). על מנת לבצע מחקר ניטרלי, ביטלנו (ניטרלנו) משתנים אלה (היסטוריית חיפוש, חיבור לשירותים נוספים וכו').

בנוסף, ידוע כי אין הבדל משמעותי בין גברים ונשים באופן החיפוש ועל כן אין שוני בתוצאות החיפוש על פי אופן הקלדה, זמן ממוצע בכל חיפוש וכו' (Zhou, 2015), ולכן, לא נטרלנו משתנים אלה.

תוצאות חיפוש ישתנו בהתאם למכשיר עליו נעשה החיפוש, כגון מכשיר סלולרי או מחשב, או על פי סוג הדפדפן. לדוגמא, חיפוש מלונות במכשיר iPhone עם דפדפן Safari יציג תוצאות אחרות

ממכשיר Android עם דפדפן Chrome, מכאן ניתן ללמוד כי תוצאות חיפוש אכן יכולות להיות שונות בין מכשירים שונים עליהם מותקנים דפדפנים שונים (Ehrenberg, 2014), אך לא ניתן להסיק בבירור האם רק סוג המכשיר הוא המשפיע על התוצאות או גם סוג הדפדפן, את הדילמה הזו בדקנו במחקר.

מהמתואר מעלה ניתן להסיק כי על ידי ניקוי היסטוריית חיפוש, ניתוק המשתמש מחשבונותיו, ניקוי "עוגיות דפדפן", חיפוש מילות מפתח ע"י תוכנה שפועלת תמיד באותו אופן במחשב מסוג מסוים אחד בלבד, על דפדפן כלשהו, ואזור גיאוגרפי כלשהו ניתן לנטרל כל משתנה בלתי תלוי שאינו סוג הדפדפן ומיקום גיאוגרפי הנבדקים במחקר זה.

2.5 הבדלים בין תוצאות חיפוש על בסיס גאוגרפי

תוצאות החיפוש מושפעות בין היתר עקב מיקום גאוגרפי. הדבר נעשה על ידי מיפוי ה IP של המחשב באמצעותו המשתמש מבצע חיפוש למיקום גאוגרפי מאוד מדויק. תוצאות החיפוש ישתנו בהתאם למיקום הגאוגרפי בייחוד כשמדובר על מילות מפתח כגון: מלונות, מסעדות ועוד מושגים המשתנים בהתאם לאזור הגאוגרפי שלהם. גוגל יציגו בנוסף לתוצאות החיפוש גם מפה אשר תתאר למשתמש כיצד להגיע ממיקומו הנוכחי ליעד אותו הוא חיפש (Khorev, 2016).

במחקר הסוקר את השפעת המיקום הגיאוגרפי על תוצאות החיפוש (Miller, 2005), ניתן להתרשם כי אנשים ממקומות שונים בעולם יכולים להגיע לרופאים מקומיים ומספרי הטלפון שלהם בחיפוש אחר בעיה רפואית, פרחים מקומיים או ביקורות על בתי עסק מקומיים. תוצאות החיפוש יהיו שונות ממקום למקום, רלוונטיות למיקום הספציפי של המחפש ולעיתים יכללו מפה עם הכוונה ליעד. מכאן אנו למדים על רלוונטיות של תוצאות החיפוש למיקומו של המשתמש.

במאמר נוסף (Orkut, Junghoo, Hector, Luis and Narayanan, 1999) מוסבר כי חיפוש מידע על מסעדות, תיאטראות, דירות להשכרה ועוד, רלוונטי למיקום המשתמש המחפש. במאמר מתוארת דוגמא למימוש אלגוריתם חיפוש בו נעשה שימוש במיקום המשתמש כחלק מאלגוריתם החיפוש המציע תוצאות חיפוש. החוקרים מסבירים כי מימוש אלגוריתם חיפוש חייב לקחת בחשבון את מיקום המשתמש על מנת להציג תוצאות רלוונטיות עבורו.

מטופלים יכולים לחפש בעיה רפואית כלשהי ולקבל בתוצאות החיפוש גישה לשמות של רופאים באזורם, מספרי טלפון, כתובות וביקורות בלי לעזוב בשום שלב את עמוד תוצאות החיפוש (Glenn, 2011).

תוצאות חיפוש על בסיס גאוגרפי הם אינהרנטיים באלגוריתם החיפוש של גוגל, אך עדיין לא ברור האם תוצאות החיפוש יהיו שונות בין סוגי דפדפנים שונים באותו מיקום גאוגרפי.

2.6 סיכום

מהנזכר לעיל ניתן להסיק כי תוצאות החיפוש מושפעות מפרמטרים רבים כגון: מיקום גאוגרפי, סוג המכשיר, היסטוריית חיפוש של המשתמש ומידע מקדים על המשתמש משירותים שונים. כל אלה הינם משתנים בלתי תלויים העשויים להשפיע על המחקר. נדרש לנטרל את המשתנים הבלתי תלויים הלא רצויים כדי לקבל מחקר מהימן. בנוסף, ישנם פרמטרים שאינם משפיעים על תוצאות החיפוש, כגון: אופן החיפוש של המשתמש, מהירות הקלדה ועוד. לכן, לא נתייחס לפרמטרים אלה במחקר. לעומת זאת, לא ניתן לדעת בוודאות מהמחקרים שנסקרו לעיל האם סוג הדפדפן מהווה פרמטר חשוב באלגוריתמי חיפוש ועל כן משפיע על אופן הצגה ודירוג של תוצאות חיפוש. בהינתן נטרול המשתנים הלא רצויים והגדרה של המשתנים הבלתי תלויים הרצויים (סוג דפדפן, מיקום גאוגרפי, מילת חיפוש) נצפה לקבל מסקנה אודות השפעת סוג הדפדפן במיקום גאוגרפי כלשהו על תוצאות החיפוש, מה שלא הצלחנו להסיק ממחקרים אחרים שבוצעו עד כה.

מתודולוגיה

3.1 מערך המחקר, אוכלוסיית המחקר.

3.1.1 מערך המחקר

3.1.1.1 חלוקה אזורית

המחקר בוצע ב 6 אזורים גאוגרפים שונים על פני 8 מדינות/ערים שונות המפורטים מטה:

1. צפון אמריקה – אוהיו (ארה"ב)
2. דרום אמריקה - סאו פאולו (ברזיל)
3. אירופה – פריז
4. אסיה - מומבאי (הודו), טוקיו (יפן)
5. אוסטרליה - סידני (אוסטרליה)
6. המזרח התיכון – בחריין, ישראל

במטרה לבדוק האם יש שוני בין האזורים ובין על פני דפדפנים שונים עבור מנוע החיפוש של גוגל. האזורים נבחרו על פי אפשרויות התמיכה הגלובלית הקיימות בשירותי הענן של Amazon (שירותי AWS) וישראל. באמצעות שירותי הענן של אמזון ניתן להקים שרתים ברחבי העולם בלחיצת כפתור ולהתחבר אליהם מרחוק על ידי תוכנת RDP. כך, ניתן למעשה לייצר סימולציית חיפוש על פני דפדפנים שונים במדינות מרוחקות.

טבלה 1: אזורים נתמכים על ידי AWS

Region Name	Region
US East (Ohio)	us-east-2
US East (N. Virginia)	us-east-1
US West (N. California)	us-west-1
US West (Oregon)	us-west-2
Asia Pacific (Mumbai)	ap-south-1
Asia Pacific (Osaka-Local)	ap-northeast-3
Asia Pacific (Seoul)	ap-northeast-2

Asia Pacific (Singapore)	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Asia Pacific (Tokyo)	ap-northeast-1
Canada (Central)	ca-central-1
China (Beijing)	cn-north-1
China (Ningxia)	cn-northwest-1
EU (Frankfurt)	eu-central-1
EU (Ireland)	eu-west-1
EU (London)	eu-west-2
EU (Paris)	eu-west-3
EU (Stockholm)	eu-north-1
South America (São Paulo)	sa-east-1
Middle East (Bahrain)	me-south-1

מקור:

<https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Concepts.RegionsAndAvailabilityZones.html>

3.1.1.2 חלוקה לפי דפדפנים

המחקר בוצע על פני שלושת הדפדפנים הפופולריים ביותר בשוק הדפדפנים כיום:

1. Chrome של חברת Google

2. Internet Explorer של חברת Microsoft

3. Firefox של חברת Mozilla

כאשר על כל דפדפן הורצו שאילתות חיפוש דומות עם פרמטרים דומים.

טבלה 2: נתח שוק של דפדפנים

Browser	Share
<input type="checkbox"/> Chrome	64.15%
<input type="checkbox"/> Internet Explorer	10.83%
<input type="checkbox"/> Firefox	9.89%
<input type="checkbox"/> Edge	4.30%
<input type="checkbox"/> Safari	3.80%
<input type="checkbox"/> Opera	1.58%
<input type="checkbox"/> Sogou Explorer	1.49%
<input type="checkbox"/> QQ	1.47%
<input type="checkbox"/> Yandex	0.62%
<input type="checkbox"/> UC Browser	0.61%

מקור: <https://netmarketshare.com/browser-market-share.aspx>

3.2 הפרדיגמה המחקרית – מחקר כמותי

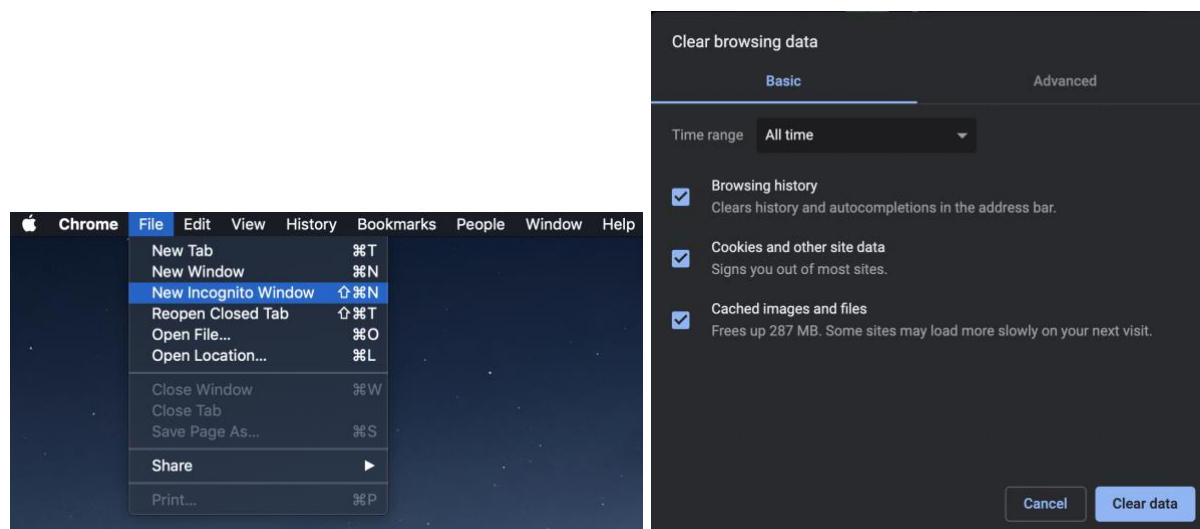
הסיבות לכך כי המחקר משלב אלמנטים של מחקר כמותי:

1. במחקר הוגדר המדד הכמותי הבא: "מהו המרחק בין תוצאות החיפוש של דפדפנים שונים באותה המדינה". הסבר על ה-"מרחק" יורחב בהמשך המחקר.
2. כלי איסוף הנתונים הוא הדפדפנים השונים עליהם התבצע החיפוש. את תוצאות השונות נשמור בקובץ CSV בו יהיה מיפוי בין שאילתת החיפוש, הדפדפן, האזור, ותוצאות החיפוש שהתקבלו.
3. המחקר מבקש למצוא קשר בין שאילתת החיפוש, אזור וסוג הדפדפן לתוצאות החיפוש.
4. המדגם כלל את סוגי הדפדפנים המייצגים מעל 85% מנתח שוק הדפדפנים כיום ואת המדינות בהן ניתן להקים שרת ניטרלי (ללא שייכות או היסטוריה של משתמש כלשהו) על ידי שירותי AWS של אמזון.
5. החלק הכמותי של המחקר אובייקטיבי לחלוטין, זאת אומרת נבדקו את אותם הפרמטרים ואותה שאילתת החיפוש על פני כל הדפדפנים בכל האזורים המוזכרים.

3.3 תכנון מחקר גישוש

מחקר גישוש הינו כלי חשוב שסייע לנו לתכנן את המחקר. מחקר גישוש, באופן כללי, מכוון את החוקרים להתנהל מחקרית בדרכים שאינן ברורות להם במידה מספקת. ערכו של מחקר הגישוש כאמצעי להבהרת מוקד המחקר ושאלותיו משמעותי ביותר. כל פרויקט מחקרי מתבסס על ידע מוקדם. ידע זה מעוגן בניסיון ובידע אישי ו / או בספרות מקצועית רלוונטית. במחקר הגישוש שלנו ביצענו חיפוש ידני מיפן (Tokyo, ע"י תוכנת VPN המדמה מיקום) של מילות המפתח: "Facebook" ו-"New York" על גבי דפדפני Firefox ו-Chrome. במקביל ביצענו את אותו החיפוש מארה"ב (Ohio, ע"י אותה תוכנת VPN). ווידאנו כי הדפדפנים רצים ללא היסטוריית חיפוש או כל מידע מקדים על המשתמש (כמו עוגיות, נתונים של אתרים אחרים ועוד). הדבר נעשה על ידי מחיקת עוגיות בדפדפנים הנ"ל והרצת האלגוריתם במצב "פרטי" אשר אינו מכיל היסטוריית משתמש או עוגיות מעקב.

איור 2: ניקוי נתוני משתמש בדפדפן Chrome



התוצאות שהתקבלו בארה"ב היו שונות מהתוצאות ביפן, אך יתרה מזאת, התוצאות שהתקבלו ביפן עצמה היו שונות בין שני הדפדפנים שנבדקו. כלומר, ביפן בלבד, עבור שאילתת החיפוש "New York" הופיעה תוצאה (קישור לאתר) בדפדפן Chrome שלא הופיעה בדפדפן Firefox ולהפך. ניתן לראות זאת בטבלה 3 מטה, אשר מציגה את המדינה בה בוצע החיפוש, מילת החיפוש, סוג הדפדפן והתוצאות (קישורים לאתרים) שפלט מנוע החיפוש גוגל.

טבלה 3: תוצאות מחקר גישוש

Date	State/City	Search Query	Browser Type	Result Link
2019-01-05	Tokyo, Japan	New York	Chrome	https://ja.wikipedia.org/wiki/ニューヨーク

				https://www.nycgo.com/
				https://ejje.weblio.jp/content/New+York
				https://ejje.weblio.jp/content/New+York+City
				https://www.newyork.jp/
				https://www.nyc.gov/
				https://www.iloveny.com/
				https://www.timeout.com/newyork
				https://www.nytimes.com/section/nyregion
				https://basikny.com/
2019-01-05	Tokyo, Japan	New York	Firefox	https://ja.wikipedia.org/wiki/ニューヨーク
				https://www.nycgo.com/
				https://ejje.weblio.jp/content/New+York
				https://www.newyork.jp/
				https://www.nyc.gov/
				https://www.iloveny.com/
				https://www.timeout.com/newyork
				https://www.nytimes.com/section/nyregion
				https://basikny.com/
				https://ny-pg.com/

תוצאות מחקר הגימוש אפשרו לנו להמשיך בניסוי והראו כי קיים שוני בין אזורים גיאוגרפיים ואפילו בין דפדפנים שונים, כאשר ביפן ישנה תוצאה חסרה בדפדפן Chrome הקיימת ב Firefox (ולהפך) עבור מילת החיפוש "New York".

3.4 המדגם

מסגרת הדגימה מורכבת משלוש קטגוריות:

1. מספר דפדפנים: 3
2. מספר מדינות: 8
3. מספר מילות החיפוש: 20

$N = \text{Number of browsers} = 3$

$M = \text{Number of regions} = 8$

$K = \text{Number of keywords} = 20$

$$\Omega = N * M * K = 3 * 8 * 20 = 480$$

4. כדי לקבל תוצאות מהימנות מילות החיפוש חולקו ל 6 קטגוריות (משתנים קטגוריאליים) ומכאן ש:

$$\Omega = N * M * C = 3 * 8 * 6 = 144$$

3.5 שיטת המחקר

3.5.1 הקדמה

המחקר הוא ניסוי, אשר כלל כתיבת קוד (תוכנה) שביצעה חיפוש באמצעות מילות חיפוש ידועות מראש בכל אחד מהאזורים המוזכרים על פני כל אחד מהדפדפנים שצוינו למעלה. בנוסף, נכתב אלגוריתם למציאת מרחק בין תוצאות החיפוש (יתואר בפירוט בהמשך).

3.5.2 מילות החיפוש

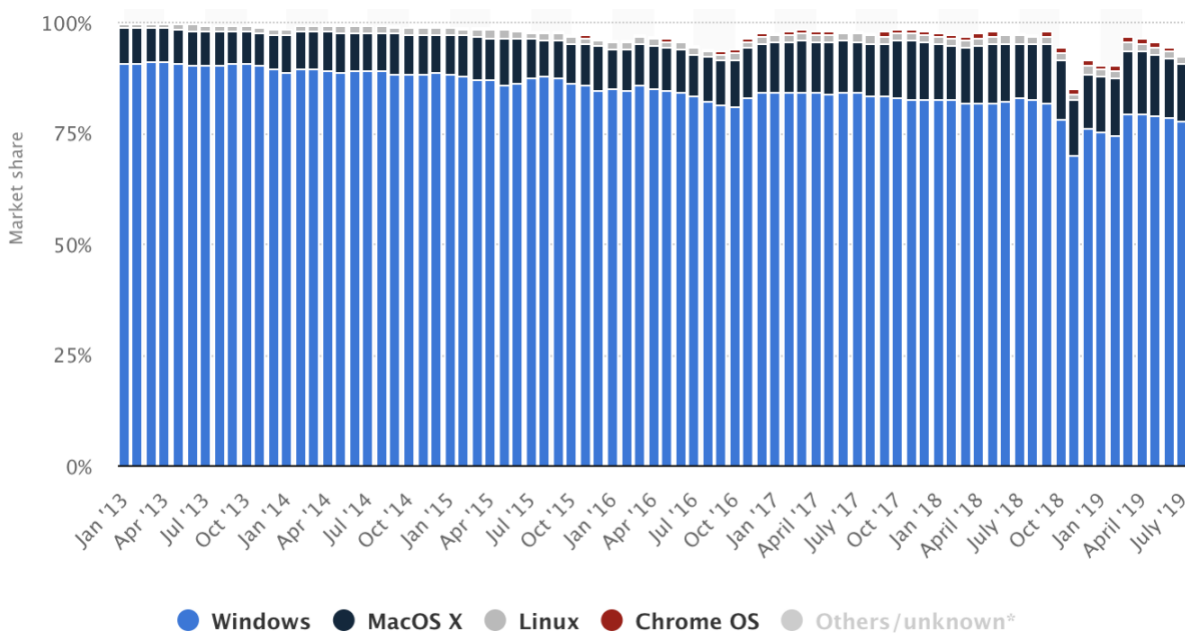
1. מילות החיפוש שנבחרו היו סביב קטגוריה ספציפית אחת: "תינוקות"
2. נבחר מדגם רחב יחסית של מילות חיפוש הכולל: colic, diapers, infant formula, newborn, baby chair, porridge, baby bottle, playpen, baby dummy, teether, labor, bassinet, reflux, baby romper, baby carrier, nanny, diaper rash, bib, stroller, baby car seat

3.5.3 מהלך הניסוי

- יצירת חשבון בשירות AWS של Amazon.

- הקמת 8 שרתי Windows Server 10 (מערכת ההפעלה הפופולארית ביותר בשוק כיום – ראו גרף 1) בכל אחד מהאזורים: אוהיו (ארה"ב), פריז (צרפת), מומביי (הודו), סידני (אוסטרליה), טוקיו (יפן), סאו פאולו (ברזיל), בחריין.
- המכונות היו מסוג t2.micro משיקולי תקציב (חינם למשך שנה ביצירת חשבון AWS חדש).
- שימוש במחשב אישי כדי להריץ את התוכנה בישראל. על המחשב האישי הותקנה מכונה וירטואלית חדשה, גם היא מעל מערכת ההפעלה Windows 10 (הפופולארית בשוק המחשבים כיום), זאת כדי שהחיפוש יבוצע ממחשב "חדש" ללא פרטים אישיים או מידע אישי כלשהו.
- כתיבת קוד בטכנולוגיית NodeJS אשר משתמש בכלי Open Source הנקרא Selenium המאפשר לדמות התנהגות של בנאדם על כל סוגי הדפדפנים (לחיצה, הקלדה ועוד).
- הרצת התוכנה על כל השרתים המוזכרים לעיל פעם ביום במשך שלושה ימים.
- לא היה שוני בין ההרצות עבור אוסף של הגדרות מסוים (דפדפן, מילת חיפוש, אזור גאוגרפי).
- בכל הרצה המחשב עלה מחדש, על כל הגדרותיו, ללא כל מידע קודם או היסטוריית חיפוש.
- בכל חיפוש הדפדפן נוקה מחדש כך שלא תהיה היסטוריית חיפוש או מידע מקדים על המשתמש.
- כלל התוצאות הגולמיות נשמרו בקבצי CSV עליהם בהמשך הרצתי אלגוריתמים נוספים.

גרף 1: נתח שוק מערכות הפעלה



מקור: <https://www.statista.com/statistics/218089/global-market-share-of-windows-7/>

טבלה 4: דוגמא לקובץ הנתונים הגולמי

Date And Time	Country	City	Query	Browser	Result Link
---------------	---------	------	-------	---------	-------------

2019-08-27T20:10:21.206Z	United States	Columbus	newborn	chrome	https://www.parents.com/baby/care/
2019-08-27T20:10:21.206Z	United States	Columbus	newborn	chrome	https://www.verywellfamily.com/difference-between-baby-newborn-infant-toddler-293848
2019-08-27T20:10:21.206Z	United States	Columbus	newborn	chrome	https://en.wikipedia.org/wiki/Infant

3.5.4 אלגוריתם חישוב המרחק

1. לכל מיקום גאוגרפי נאספו כ 10 תוצאות חיפוש עבור מילת חיפוש כלשהי וסוגי דפדפנים שונים.
2. עבור דפדפן מסוג כלשהו נבדק מה המרחק של תוצאות החיפוש, באותה המדינה, כנגד תוצאות החיפוש של דפדפן אחר.
3. הנ"ל נעשה על ידי הרצת אלגוריתם מרחק הבדק את דירוג התוצאות בין שני דפדפנים שונים.
4. אלגוריתם המרחק: ניתן ניקוד (מרחק) לדירוג תוצאות שונות על ידי הפרש אבסולוטי בין המיקומים של תוצאות החיפוש (לדוגמא: בעת חיפוש המילה "reflux", אם תוצאת חיפוש מסוימת, המיוצגת על ידי קישור לאתר מסוים, דורגה במיקום 3 בדפדפן Firefox אך במיקום 5 בדפדפן Chrome על ידי מנוע החיפוש של גוגל, הדבר היווה $2 = 3 - 5$ נקודות, כלומר המרחק הוא 2)
5. תוצאת חיפוש שאינה נמצאת בדפדפן אחד אך נמצאת לאחר זיכתה ב 10 נקודות, כלומר המרחק הוא 10.
6. הפלט של האלגוריתם הנ"ל גם הוא בוצע לקובץ CSV.

טבלה 5: דוגמא לפלט אלגוריתם המרחק

Query	Browsers	Distance
infant formula	chrome-firefox	262
diapers	chrome-firefox	72
infant formula	firefox-internet explorer	142
diapers	firefox-internet explorer	32

3.6 כלי איסוף הנתונים

כפי שתואר לעיל, נעשה שימוש בשירותי הענן של Amazon הנקראים AWS, כמו כן, בוצע שימוש בתוכנה שנכתבה למטרת המחקר הזה בשפת Javascript מעל טכנולוגיית NodeJS ובכלי Open

Source הנקרא Selenium. חיבור כל הגורמים האלה אפשר איסוף נתונים לתוך קבצי CSV שהומרו מאוחר יותר לקבצים אשר תוכנת R או Weka יודעות לפענח.

3.7 ניתוח הנתונים

ניתוח הנתונים בוצע באמצעות תוכנת R ותוכנת Weka. תחילה בוצעה רגרסיה לינארית פשוטה שאינה הניבה תוצאות טובות באמצעות תוכנת R ולאחר מכן נעשה שימוש באלגוריתם לומד J48 אשר מצוי בתוכנת Weka ולמעשה מייצר עץ החלטות. עבור תוכנת R חולקו המדינות אוסטרליה, ברייטון, ברזיל, הודו, יפן, ישראל, צרפת, ארה"ב למספרים 1-8 בהתאמה. מילות החיפוש והמרחקים חולקו לקטגוריות (יתואר בהמשך).

3.7.1 הסבר על תוכנת Weka

Weka היא תוכנה המכילה אוסף של כלי ויזואליזציה ואלגוריתמים לניתוח נתונים וחיזוי מודלים, יחד עם ממשקי משתמש גרפיים עבור גישה קלה לפונקציות אלה.

כל הטכניקות של Weka מושתתות על ההנחה כי הנתונים זמינים כקובץ או ביחס שטוח אחד, כאשר כל נקודת נתונים מתוארת על ידי מספר קבוע של תכונות (בדרך כלל, תכונות מספריות או נומינליות).

3.7.2 בניית קובץ הנתונים בפורמט ARFF עבור תוכנת Weka

1. קובץ ARFF (Attribute-Relation File Format) מורכב משני חלקים:

1. Header – שם, מאפיינים וסוגי המאפיינים
2. Data – הנתונים עצמם, שורות אשר בכל אחת מהן הנתונים מופרדים בסימן פיסוק ללא רווחים.
2. הנתונים חולקו לקטגוריות באופן הבא:
 1. 8 מדינות
 2. 3 סוגי דפדפנים
 3. 6 קטגוריות של מילות חיפוש (שם עצם, בסיס, אוכל, אביזרים, מחלה, סחיבה)

3.7.3 עץ החלטות

עץ החלטה הוא עץ בינארי מלא המורכב מצמתי החלטה שבכל אחד מהם נבדק תנאי מסוים על מאפיין מסוים של התצפיות ועלים המכילים את הערך החזוי עבור התצפית המתאימה למסלול שמוביל אליהם בעץ.

למידה מבוססת עץ החלטה היא אחת הטכניקות המוצלחות ביותר ללמידה מפוקחת באמצעות סיווג. עץ יכול "ללמוד" על ידי פיצול קבוצת המקור לתתי קבוצות, המתבססות על מתן ערך לתכונה. תהליך זה חוזר על עצמו בכל תת-קבוצה באופן רקורסיבי, ונקרא מחיצות רקורסיביות. הרקורסיה מושלמת כאשר כל קבוצות המשנה בצומת בעלות אותו ערך מטרה, או כאשר הפיצול כבר אינו מוסיף ערך לתחזיות. הדבר שונה מרגרסיה בה יש ניסיון לחזות מודל לינארי (נוסחה) בין כל הנקודות המייצגות את התוצאות שהתקבלו.

3.7.4 הסבר על אלגוריתם J48

האלגוריתם J48 מבוסס על אלגוריתם עץ החלטות בשם C4.5 שפותח על ידי Ross Quinlan, חוקר בתחום מדעי המחשב, כריית נתונים ותאוריית ההחלטות. למעשה, האלגוריתם מנסה לפצל כל מאפיין לענפים ("החלטה") שיובילו למאפיין הבא ובודק בסוף יצירת העץ על ההחלטות שקיבל לפי הנתונים האמיתיים. על ידי כמה מעברים ולמידה מטעויות אפשר לקבל עץ מדויק שמצביע על החלטות המתאימות למציאות.

נבחנו אלגוריתמים נוספים, כגון Rotation Forest אשר מנסה לייצר עץ החלטות מצומצם על ידי ירידה ועליה בעץ ההחלטות וניסיון צמצום העץ עד לקבלת העץ הקטן ביותר (נספח IX). אך, התוצאות הטובות ביותר התקבלו באמצעות אלגוריתם J48.

איור 3: דוגמא לעותק מקובץ הנתונים עבור תוכנת R

```
country,query,browsers,distance
1,1,1,2
1,2,1,3
1,3,1,3
1,4,1,3
1,3,1,3
1,1,1,3
1,5,1,1
1,5,1,3
1,5,1,2
```

איור 4: דוגמא לעותק מקובץ הנתונים עבור תוכנת Weka

```
@RELATION browsers

@ATTRIBUTE country {australia,bahrain,brazil,france,india,israel,japan,usa}
@ATTRIBUTE query {noun,basic,food,accessory,carry,disease}
@ATTRIBUTE browsers {chrome-internet-explorer,chrome-firefox,firefox-internet-explorer}
@ATTRIBUTE distance {small,medium,big}

@DATA
australia,noun,chrome-firefox,medium
australia,food,chrome-firefox,big
australia,basic,chrome-firefox,big
australia,disease,chrome-firefox,big
australia,basic,chrome-firefox,big
australia,noun,chrome-firefox,big
australia,accessory,chrome-firefox,small
australia,accessory,chrome-firefox,big
australia,accessory,chrome-firefox,medium
australia,food,chrome-firefox,big
australia,basic,chrome-firefox,big
australia,carry,chrome-firefox,small
australia,accessory,chrome-firefox,big
australia,basic,chrome-firefox,medium
australia,food,chrome-firefox,big
australia,disease,chrome-firefox,big
australia,noun,chrome-firefox,big
australia,carry,chrome-firefox,small
australia,basic,chrome-firefox,small
australia,disease,chrome-firefox,big
australia,noun,chrome-internet-explorer,medium
australia,food,chrome-internet-explorer,small
```

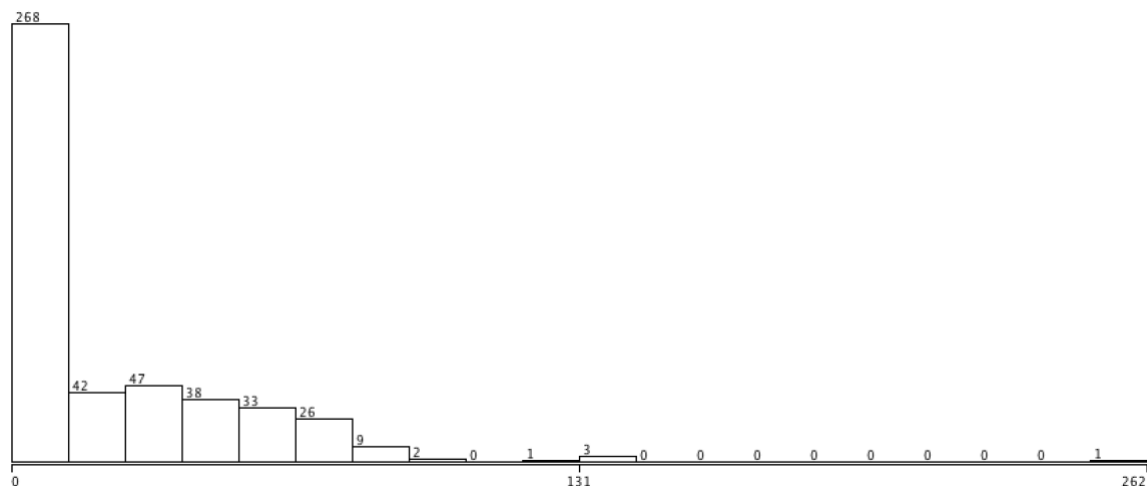
4. תוצאות

4.1 ניתוח תוצאות על ידי רגרסיה לינארית

4.1.1 פיזור התוצאות ואלגוריתם הקטגוריזציה

1. נעשה שימוש בתוכנת Weka על מנת למצוא את פיזור המרחקים על גבי כלל התוצאות.
2. התקבל הפיזור הבא:
 - i. 268 תוצאות קיבלו מרחקים: 0 – 13.1
 - ii. 127 תוצאות קיבלו מרחקים: 13.1 – 52.4
 - iii. 75 תוצאות קיבלו מרחקים הגדולים מ 52.4
3. על כן, הוחלט לבצע קטגוריזציה של התוצאות על פי הפיזור למעלה, כך שכל קטגוריה מקבלת את A,B,C בהתאמה.
4. בנוסף, מכיוון שכמות מילות החיפוש גדולה, גם אותה חילקנו ל 6 קטגוריות: מזון (B), מוצרי בסיס (C), אביזרים נלווים (E), מחלות ילדים (D), מוצרי הנגשה (F) ושמות עצם כלליים (A).

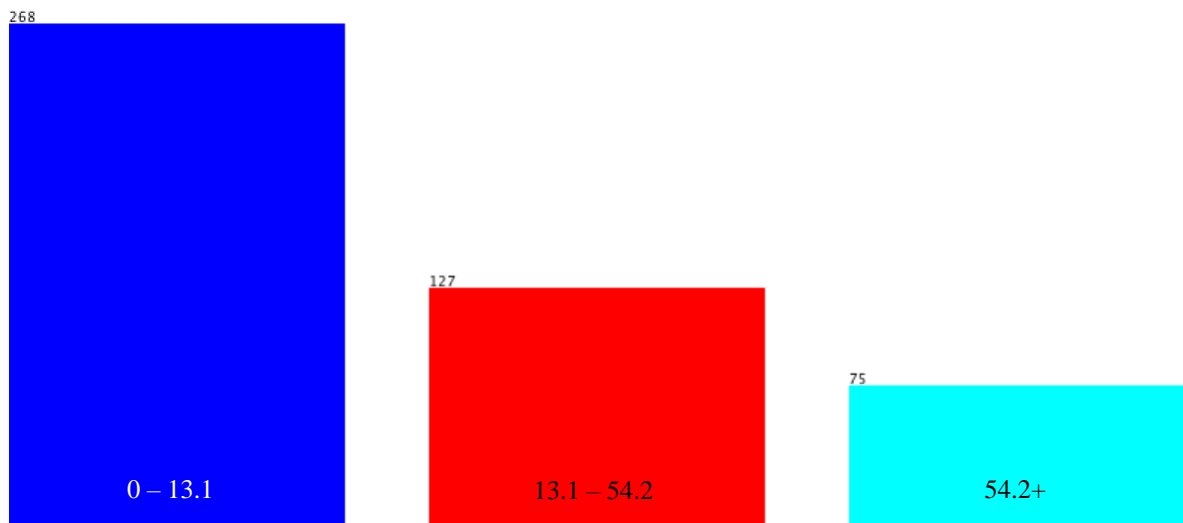
גרף 2: פיזור התוצאות למרחקים



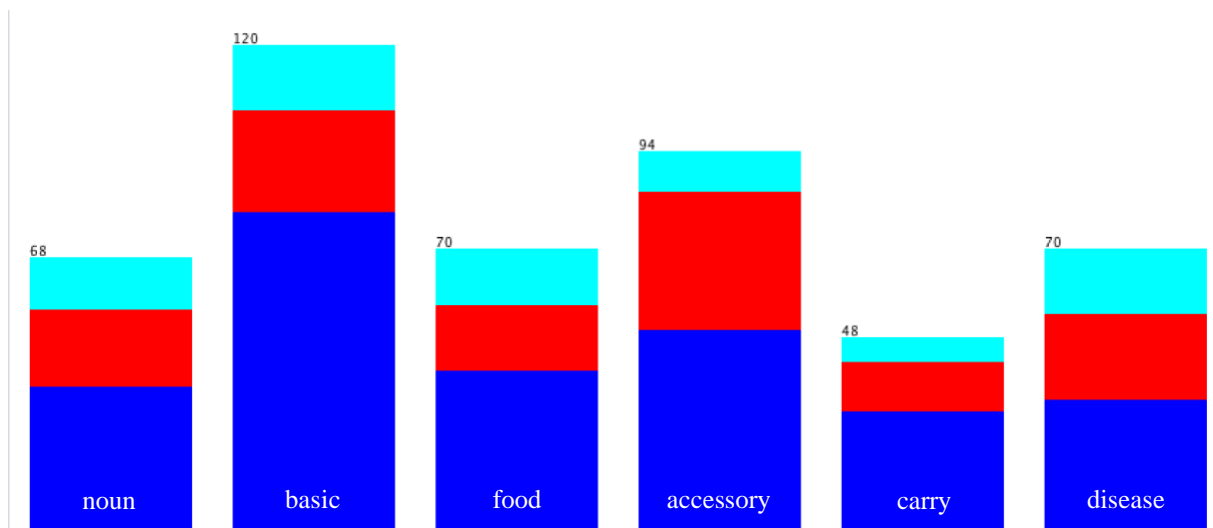
טבלה 4: דוגמא לפלט אלגוריתם הקטגוריזציה

Query	Browsers	Distance
food	chrome-firefox	C
basic	chrome-firefox	C
food	firefox-internet explorer	C
basic	firefox-internet explorer	B

גרף 3: פיזור מרחק לאחר קטגוריזציה



גרף 4: פיזור מילות חיפוש לאחר קטגוריזציה



4.1.2 מתאם פירסון

שימוש בתוכנת R, ראשית בדקנו האם יש קשר בין סוג הדפדפן, המדינה ומילת החיפוש למרחק על ידי מתאם פירסון.

טבלה 5: תוצאות מתאם פירסון

	Country	Query	Browsers
Distance	P-Val: 0.002847 Cor: 0.1373433	P-Val: 0.5199 Cor: -0.0297518	P-Val: 0.008946 Cor: -0.1204613

תוצאות מבחן מתאם פירסון מעידות כי נמצא קשר חיובי מובהק בין מדינה ומרחק, קשר שלילי מובהק בין דפדפן ומרחק וקשר שלילי שאינו מובהק בין מילת חיפוש למרחק.

4.1.2 מבחן T

שימוש בתוכנת R, נתקדם לביצוע מבחן T להשוואת קבוצות בלתי תלויות על מנת לבדוק האם ההפרשים בין הממוצעים של המשתנים הם מובהקים סטטיסטית.

טבלה 6: פלט תוכנת R - תוצאות מבחן T שוניות שוות

t.test(distance, country + browsers + query, var.equal = TRUE)					
T	DF	P-Val	95% confidence interval	Mean of X	Mean of Y
2.9774	938	0.002981	1.409195 – 6.858890	21.19362	17.05957

טבלה 7: פלט תוכנת R - תוצאות מבחן T שוניות שונות

t.test(distance, country + browsers + query, var.equal = FALSE)					
T	DF	P-Val	95% confidence interval	Mean of X	Mean of Y
2.9774	510.93	0.003045	1.406249 – 6.861836	21.19362	17.05957

טבלה 8: תוצאות מבחן T מאוחדות

	T	DF	P-Value	95 percent confidence interval
Equal variances assumed	2.9774	938	0.002981	1.409195 - 6.858890
Equal variances not assumed	2.9774	510.93	0.003045	1.406249 - 6.861836

מתוצאות המבחן, ניתן להסיק ברמת מובהקות טובה מ 5% כי קיים הבדל בתוצאות החיפוש בין דפדפנים שונים בתלות בסוג הדפדפן, מילת החיפוש והמדינה. נעבור לביצוע רגרסיה לינארית.

4.1.3 ביצוע רגרסיה לינארית

שימוש בתוכנת R, נתקדם לביצוע הרגרסיה הלינארית. לצורך כך העברנו את כל המשתנים למשתני דמה קטגוריאליים כך שכל משתנה מפורק לאופציות הקטגוריאליות שלו. משתנה Country (מדינה) מפורק ל8 משתני דמה קטגוריאליים C1,C2,C3,C4,C5,C6,C7,C8 המייצגים את כל המדינות: אוסטרליה, בחריין, ברזיל, צרפת, הודו, ישראל, יפן, ארה"ב. משתנה Query (מילת חיפוש) מפורק ל6 משתני דמה קטגוריאליים Q1,Q2,Q3,Q4,Q5,Q6 המייצגים את כל מילות החיפוש בקטגוריות: שם עצם, אוכל, מחלה, בסיס, אביזרים, סחיבה משתנה Browsers (דפדפן) מפורק ל3 משתני דמה קטגוריאליים B1,B2,B3 המייצגים את כל הדפדפנים שנבדקו: IE-Firefox, Chrome-IE, Chrome-Firefox. הרכבה של כל המשתנים הנ"ל תהווה 144 אפשרויות צירוף, כך ש C1B1Q1 מציין כי החיפוש נעשה במדינת אוסטרליה על דפדפנים Chrome-Firefox עבור מילת חיפוש מקטגוריית שם עצם. הפורמולה לייצוג הרגרסיה הינה ללא החותך על מנת לבטל אפשרות של תלות בין משתנים:

$$\begin{aligned} \text{formula} = \text{distance} \sim & C1B1Q1 + C1B1Q2 + C1B1Q3 + C1B1Q4 + C1B1Q5 + \\ & C1B1Q6 + C1B2Q1 + C1B2Q2 + C1B2Q3 + C1B2Q4 + C1B2Q5 + C1B2Q6 + \\ & C1B3Q1 + C1B3Q2 + C1B3Q3 + C1B3Q4 + C1B3Q5 + C1B3Q6 + C2B1Q1 + \\ & C2B1Q2 + C2B1Q3 + C2B1Q4 + C2B1Q5 + C2B1Q6 + C2B2Q1 + C2B2Q2 + \\ & C2B2Q3 + C2B2Q4 + C2B2Q5 + C2B2Q6 + C2B3Q1 + C2B3Q2 + C2B3Q3 + \\ & C2B3Q4 + C2B3Q5 + C2B3Q6 + C3B1Q1 + C3B1Q2 + C3B1Q3 + C3B1Q4 + \\ & C3B1Q5 + C3B1Q6 + C3B2Q1 + C3B2Q2 + C3B2Q3 + C3B2Q4 + C3B2Q5 + \\ & C3B2Q6 + C3B3Q1 + C3B3Q2 + C3B3Q3 + C3B3Q4 + C3B3Q5 + C3B3Q6 + \\ & C4B1Q1 + C4B1Q2 + C4B1Q3 + C4B1Q4 + C4B1Q5 + C4B1Q6 + C4B2Q1 + \\ & C4B2Q2 + C4B2Q3 + C4B2Q4 + C4B2Q5 + C4B2Q6 + C4B3Q1 + C4B3Q2 + \\ & C4B3Q3 + C4B3Q4 + C4B3Q5 + C4B3Q6 + C5B1Q1 + C5B1Q2 + C5B1Q3 + \\ & C5B1Q4 + C5B1Q5 + C5B1Q6 + C5B2Q1 + C5B2Q2 + C5B2Q3 + C5B2Q4 + \\ & C5B2Q5 + C5B2Q6 + C5B3Q1 + C5B3Q2 + C5B3Q3 + C5B3Q4 + C5B3Q5 + \\ & C5B3Q6 + C6B1Q1 + C6B1Q2 + C6B1Q3 + C6B1Q4 + C6B1Q5 + C6B1Q6 + \\ & C6B2Q1 + C6B2Q2 + C6B2Q3 + C6B2Q4 + C6B2Q5 + C6B2Q6 + C6B3Q1 + \\ & C6B3Q2 + C6B3Q3 + C6B3Q4 + C6B3Q5 + C6B3Q6 + C7B1Q1 + C7B1Q2 + \\ & C7B1Q3 + C7B1Q4 + C7B1Q5 + C7B1Q6 + C7B2Q1 + C7B2Q2 + C7B2Q3 + \\ & C7B2Q4 + C7B2Q5 + C7B2Q6 + C7B3Q1 + C7B3Q2 + C7B3Q3 + C7B3Q4 + \\ & C7B3Q5 + C7B3Q6 + C8B1Q1 + C8B1Q2 + C8B1Q3 + C8B1Q4 + C8B1Q5 + \\ & C8B1Q6 + C8B2Q1 + C8B2Q2 + C8B2Q3 + C8B2Q4 + C8B2Q5 + C8B2Q6 + \\ & C8B3Q1 + C8B3Q2 + C8B3Q3 + C8B3Q4 + C8B3Q5 + C8B3Q6 - 1 \end{aligned}$$

טבלה 9: תוצאות רגרסיה לינארית

	Estimate	Std. Error	T Value	P-Value
C1B1Q1	49.333	12.004	4.11	5.01E-05
C1B1Q2	73.333	12.004	6.109	2.85E-09
C1B1Q3	80	12.004	6.665	1.13E-10
C1B1Q4	49.6	9.298	5.334	1.80E-07
C1B1Q5	45.5	10.396	4.377	1.62E-05
C1B1Q6	0	14.701	0	1
C1B2Q1	7.333	12.004	0.611	0.541676
C1B2Q2	16.667	12.004	1.388	0.165944
C1B2Q3	46.333	12.004	3.86	0.000137
C1B2Q4	2.8	9.298	0.301	0.7635
C1B2Q5	2	10.396	0.192	0.847556
C1B2Q6	34	14.701	2.313	0.021362
C1B3Q1	29.333	12.004	2.444	0.015068
C1B3Q2	29.333	12.004	2.444	0.015068
C1B3Q3	37.667	12.004	3.138	0.001857
C1B3Q4	19.6	9.298	2.108	0.035796
C1B3Q5	18	10.396	1.732	0.084305
C1B3Q6	34	14.701	2.313	0.021362
C2B1Q1	20.333	12.004	1.694	0.091235
C2B1Q2	51.333	12.004	4.276	2.50E-05
C2B1Q3	50.667	12.004	4.221	3.16E-05
C2B1Q4	26.6	9.298	2.861	0.004498
C2B1Q5	17.5	10.396	1.683	0.093251
C2B1Q6	36	14.701	2.449	0.014862
C2B2Q1	8.333	12.004	0.694	0.488032
C2B2Q2	3.333	12.004	0.278	0.781424
C2B2Q3	14.333	12.004	1.194	0.233316
C2B2Q4	12.6	9.298	1.355	0.176315
C2B2Q5	10.5	10.396	1.01	0.313219
C2B2Q6	0	14.701	0	1
C2B3Q1	10.333	12.004	0.861	0.389955
C2B3Q2	24.667	12.004	2.055	0.040683

C2B3Q3	29.333	12.004	2.444	0.015068
C2B3Q4	10.2	9.298	1.097	0.273447
C2B3Q5	14	10.396	1.347	0.179
C2B3Q6	16	14.701	1.088	0.277254
C3B1Q1	2.667	12.004	0.222	0.824333
C3B1Q2	0	12.004	0	1
C3B1Q3	0	12.004	0	1
C3B1Q4	0	9.298	0	1
C3B1Q5	0.5	10.396	0.048	0.961668
C3B1Q6	7	14.701	0.476	0.634291
C3B2Q1	2.667	12.004	0.222	0.824333
C3B2Q2	0	12.004	0	1
C3B2Q3	0	12.004	0	1
C3B2Q4	0	9.298	0	1
C3B2Q5	0.5	10.396	0.048	0.961668
C3B2Q6	0	14.701	0	1
C3B3Q1	2.667	12.004	0.222	0.824333
C3B3Q2	0	12.004	0	1
C3B3Q3	0	12.004	0	1
C3B3Q4	0	9.298	0	1
C3B3Q5	0.5	10.396	0.048	0.961668
C3B3Q6	7	14.701	0.476	0.634291
C4B1Q1	0	12.004	0	1
C4B1Q2	0	12.004	0	1
C4B1Q3	0	12.004	0	1
C4B1Q4	2.4	9.298	0.258	0.796477
C4B1Q5	0	12.004	0	1
C4B1Q6	0	14.701	0	1
C4B2Q1	16.667	12.004	1.388	0.165944
C4B2Q2	0	12.004	0	1
C4B2Q3	4.667	12.004	0.389	0.6977
C4B2Q4	2.6	9.298	0.28	0.779939
C4B2Q5	6.25	10.396	0.601	0.54811
C4B2Q6	0	14.701	0	1

C4B3Q1	16.667	12.004	1.388	0.165944
C4B3Q2	0	12.004	0	1
C4B3Q3	4.667	12.004	0.389	0.6977
C4B3Q4	2.6	9.298	0.28	0.779939
C4B3Q5	4.667	12.004	0.389	0.6977
C4B3Q6	0	14.701	0	1
C5B1Q1	74.667	12.004	6.22	1.52E-09
C5B1Q2	61	14.701	4.149	4.26E-05
C5B1Q3	44	12.004	3.666	0.000288
C5B1Q4	64	9.298	6.883	3.01E-11
C5B1Q5	64	10.396	6.157	2.18E-09
C5B1Q6	35	14.701	2.381	0.017853
C5B2Q1	16.667	12.004	1.388	0.165944
C5B2Q2	9.333	12.004	0.778	0.437405
C5B2Q3	26	12.004	2.166	0.031036
C5B2Q4	4.8	9.298	0.516	0.606037
C5B2Q5	23	10.396	2.212	0.027625
C5B2Q6	3	14.701	0.204	0.838433
C5B3Q1	43.333	12.004	3.61	0.000354
C5B3Q2	21	14.701	1.428	0.154125
C5B3Q3	34	12.004	2.832	0.004906
C5B3Q4	25.2	9.298	2.71	0.007079
C5B3Q5	36.75	10.396	3.535	0.000466
C5B3Q6	12.5	14.701	0.85	0.395807
C6B1Q1	7.333	12.004	0.611	0.541676
C6B1Q2	21.333	12.004	1.777	0.076462
C6B1Q3	0	14.701	0	1
C6B1Q4	0.8	9.298	0.086	0.931488
C6B1Q5	9.75	10.396	0.938	0.348987
C6B1Q6	0	14.701	0	1
C6B2Q1	50.667	12.004	4.221	3.16E-05
C6B2Q2	43.667	12.004	3.638	0.00032
C6B2Q3	48	12.004	3.999	7.88E-05
C6B2Q4	40.4	9.298	4.345	1.86E-05

C6B2Q5	34.75	10.396	3.343	0.000926
C6B2Q6	48.5	14.701	3.299	0.001078
C6B3Q1	49	12.004	4.082	5.62E-05
C6B3Q2	47	12.004	3.915	0.00011
C6B3Q3	36.5	14.701	2.483	0.01354
C6B3Q4	40.4	9.298	4.345	1.86E-05
C6B3Q5	30.75	10.396	2.958	0.003323
C6B3Q6	48.5	14.701	3.299	0.001078
C7B1Q1	5	14.701	0.34	0.733998
C7B1Q2	0	12.004	0	1
C7B1Q3	6	12.004	0.5	0.61752
C7B1Q4	4.8	9.298	0.516	0.606037
C7B1Q5	6.75	10.396	0.649	0.516589
C7B1Q6	0	14.701	0	1
C7B2Q1	5	14.701	0.34	0.733998
C7B2Q2	0	12.004	0	1
C7B2Q3	6.667	12.004	0.555	0.579012
C7B2Q4	2	9.298	0.215	0.829824
C7B2Q5	0.5	10.396	0.048	0.961668
C7B2Q6	16	14.701	1.088	0.277254
C7B3Q1	0	12.004	0	1
C7B3Q2	0	12.004	0	1
C7B3Q3	1.333	12.004	0.111	0.911624
C7B3Q4	2.8	9.298	0.301	0.7635
C7B3Q5	7.25	10.396	0.697	0.486039
C7B3Q6	16	14.701	1.088	0.277254
C8B1Q1	79.333	12.004	6.609	1.58E-10
C8B1Q2	129.667	12.004	10.802	<2.00E-16
C8B1Q3	96	12.004	7.998	2.23E-14
C8B1Q4	52.6	9.298	5.657	3.37E-08
C8B1Q5	52	10.396	5.002	9.28E-07
C8B1Q6	51.5	14.701	3.503	0.000524
C8B2Q1	10	14.701	0.68	0.496857
C8B2Q2	21.333	12.004	1.777	0.076462

C8B2Q3	64.667	12.004	5.387	1.37E-07
C8B2Q4	5	9.298	0.538	0.591117
C8B2Q5	31	10.396	2.982	0.003079
C8B2Q6	2	14.701	0.136	0.891873
C8B3Q1	52.5	14.701	3.571	0.000409
C8B3Q2	65.333	12.004	5.443	1.03E-07
C8B3Q3	72.667	12.004	6.054	3.89E-09
C8B3Q4	19.2	9.298	2.065	0.039717
C8B3Q5	28.5	10.396	2.742	0.006452
C8B3Q6	21.5	14.701	1.462	0.144584
Residual standard error	0.79 on 326 degrees of freedom			
Multiple R-squared	0.7719			
Adjusted R-squared	0.6712			
F-statistic	7.662 on 144 and 326 DF			
p-value	< 2.2e-16			

4.1.4 מסקנות הרגרסיה

ניתן לראות כי רמת המובהקות עבור המשתנים המסבירים מדינה וסוגי דפדפנים טובה מ-5% אך R^2 נמוך מעט, כלומר לא ניתן באופן מובהק להסביר את השינויים של המשתנה התלוי מרחק (distance) על ידי שינויים של המשתנים הבלתי תלויים (מדינה, מילת חיפוש וסוג דפדפן).

4.2 ניתוח תוצאות על ידי אלגוריתם עץ J48

4.2.1

איור 5: תוצאות הרצת אלגוריתם J48

=== Summary ===

Correctly Classified Instances	353	75.1064 %
Incorrectly Classified Instances	117	24.8936 %
Kappa statistic	0.5481	
Mean absolute error	0.2396	
Root mean squared error	0.3601	
Relative absolute error	62.2783 %	
Root relative squared error	82.1186 %	
Total Number of Instances	470	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.892	0.312	0.791	0.892	0.839	0.599	0.843	0.854	small
	0.551	0.114	0.642	0.551	0.593	0.460	0.733	0.522	medium
	0.587	0.038	0.746	0.587	0.657	0.606	0.853	0.572	big
Weighted Avg.	0.751	0.215	0.744	0.751	0.743	0.563	0.815	0.719	

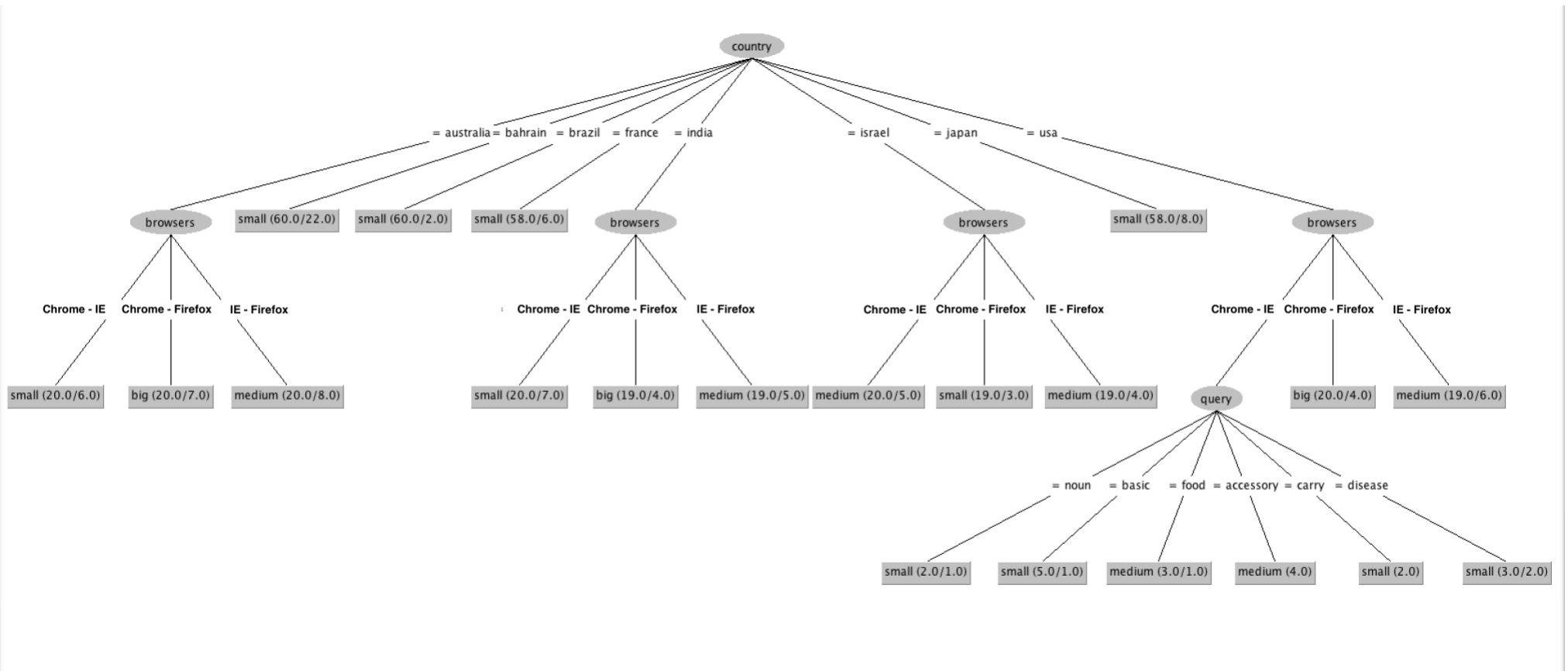
=== Confusion Matrix ===

a	b	c	<-- classified as
239	22	7	a = small
49	70	8	b = medium
14	17	44	c = big

4.2.3 מסקנות הרצת אלגוריתם J48:

1. ניתן לראות כי המודל חזה ב 75% נכונה את ההחלטות שבקובץ נתונים.
2. ניתן לראות במטריצת הטעויות (Confusion Matrix) כי האלגוריתם חזה נכון 239 תוצאות אשר מרחקן קטן (0-13), אך, טעה וחזה כי 22 תוצאות בעלות מרחק בינוני ו 7 תוצאות בעלות מרחק גדול למרות שהיו אמורות להיות בעלות מרחק קטן. הנ"ל מהווה 89% סיכוי לאלגוריתם לחזות נכונה מרחק קטן.
3. בנוסף, ניתן לראות במטריצת הטעויות (Confusion Matrix) כי האלגוריתם חזה נכון 70 תוצאות אשר מרחקן בינוני (13-52), אך, טעה וחזה כי 49 תוצאות בעלות מרחק קטן ו 8 תוצאות בעלות מרחק גדול למרות שהיו אמורות להיות בעלות מרחק בינוני גם כן. הנ"ל מהווה 55% סיכוי לאלגוריתם לחזות נכונה מרחק בינוני.
4. ולבסוף, ניתן לראות במטריצת הטעויות (Confusion Matrix) כי האלגוריתם חזה נכון 44 תוצאות אשר מרחקן גדול (52 ומעלה), אך, טעה וחזה כי 14 תוצאות בעלות מרחק קטן ו 17 תוצאות בעלות מרחק בינוני למרות שהיו אמורות להיות בעלות מרחק גדול. הנ"ל מהווה 58% סיכוי לאלגוריתם לחזות נכונה מרחק גדול.

איור 6: עץ ההחלטות של אלגוריתם J48



4.2.4 הסבר עץ ההחלטות:

1. קיימות מדינות בהן יש הבדל בתוצאות החיפוש בין דפדפנים ללא קשר למילת החיפוש שנבחרה, והן: אוסטרליה, הודו וישראל.
2. קיימות מדינות בהן אין הבדל בתוצאות החיפוש בין סוגי דפדפנים, והן: בחריין, ברזיל, צרפת ויפן.
3. בארה"ב יש הבדל בין תוצאות החיפוש בין דפדפנים מסוגים שונים, כאשר עבור ההבדלים בין דפדפן Chrome ודפדפן Internet Explorer מושפעים גם ממילת החיפוש שנדגמה.
4. באוסטרליה, ארה"ב ובהודו ההבדל הגדול ביותר בין תוצאות החיפוש יהיה בין דפדפן Chrome לדפדפן Firefox, ואילו בישראל ההפך הוא הנכון.

5. סיכום ודיון

במחקר זה נבדק האם קיים שוני בדרוג תוצאות החיפוש בין דפדפנים מסוגים שונים במקום גאוגרפי כלשהו. נדגמו 8 מדינות ברחבי העולם (אוסטרליה, ברזיל, בחריין, ישראל, יפן, צרפת, ארה"ב, הודו), על גבי מנוע החיפוש של גוגל. בכל מדינה, הורצה תוכנת חיפוש אוטומטית ללא היסטוריה או מידע מקדים על המשתמש לפי 20 מילות חיפוש בנושא תינוקות. נבדקו 3 הדפדפנים הפופולאריים ביותר בשוק כיום.

מתוצאות המחקר ניתן להסיק כי קיים שוני בדרוג תוצאות החיפוש בתלות בסוג הדפדפן במדינות אוסטרליה, הודו, ישראל וארה"ב. בארה"ב, בניגוד לשאר המדינות בהם יש שוני בדירוג תוצאות החיפוש, ראינו כי השוני בתוצאות החיפוש בין דפדפן Chrome ל Internet Explorer מוסבר גם באמצעות מילת החיפוש.

ראינו בתוצאות המחקר כי אין שוני בדרוג תוצאות החיפוש במדינות יפן, בחריין, ברזיל וצרפת בין סוגי הדפדפנים השונים. הדבר מעיד על כך שבמדינות מסוימות בעולם דירוג תוצאות החיפוש בין דפדפנים יהיה זהה לעומת מדינות אחרות בהן הדירוג יהיה שונה.

בנוסף, רואים בתוצאות המחקר כי ההבדל הגדול ביותר בין תוצאות החיפוש הוא בין דפדפן Chrome לדפדפן Firefox באוסטרליה, ארה"ב והודו. בניגוד לכך, בישראל, ההפך הוא הנכון (ההבדל בין תוצאות החיפוש בין דפדפן Chrome לדפדפן Firefox הוא הקטן ביותר).

המסקנה למטרת קידום אתרים היא כי יש שוני בדרוג תוצאות החיפוש בין דפדפנים במדינות מסוימות בעולם. כלומר, השערת המחקר הוכחה. השוני התקבל בסבירות הגבוהה ביותר (89%) עבור מילות חיפוש שהמרחקים ביניהם היו קטנים (0-13), ולכן, כאשר עושים אופטימיזציה למנועי חיפוש, יש לבדוק כי האופטימיזציה מצליחה על כל סוגי הדפדפנים הפופולאריים בשוק עבור כל מדינה שרלוונטית לקידום האתר.

המלצות להמשך:

1. הרצת תוכנת החיפוש האוטומטית על מרחב מדגם גדול יותר:
 - i. כל המדינות בעולם (יש 193 מדינות בעולם).
 - ii. מילות חיפוש מתחומים נוספים הרלוונטיים למדגם רחב של אתרי אונליין כגון: ספורט, בגדים, מוצרי אלקטרוניקה ועוד.
 - iii. הוספת דפדפן Safari לתוכנת החיפוש באמצעות שימוש במחשבי Mac של חברת Apple (לא נתמך ע"י AWS בשלב זה).
2. מומלץ לבדוק האם מערכת ההפעלה, עליה רצה התוכנה, משפיעה על תוצאות המחקר. במחקר זה בדקנו את מערכת ההפעלה Windows 10 בלבד (הפופולארית ביותר בשוק מערכות ההפעלה כיום).

5. רשימת מקורות

מקורות אקדמיים

- Ao-Jan Su, Hu Y C, Kuzmanovic A., Cheng-Kok Koh (2010) How to Improve Your Google Ranking: Myths and Reality. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 1, pp. 50-57.
- Bar-Ilan, Judit (2006) Web links and search engine ranking: The case of Google and the query "jew". *Journal of the American Society for Information Science and Technology*, Vol. 57, No. 12, pp. 1581-1589.
- Bates J., Best P., Mcquilkin J., Taylor B. (2017) Will Web Search Engines Replace Bibliographic Databases in the Systematic Identification of Research? *The Journal of Academic Librarianship*, Vol.43, No. 1, pp. 8-17.
- Ehrenberg Rachel (2014) Math & technology: Online retailers personalize search results to try to maximize profits: Websites alter results depending on whether consumers use smartphones or particular web browsers. *Science News*, Vol. 186, No. 11, pp.7-7
- Henzinger M., Chang B., Milch B., Brin S. (2005) Query-Free News Search. *World Wide Web*, Vol.8, No. 2, pp. 101-126.
- Heting Chu and Marilyn Rosenthal (1996) Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology. *Long Island University Brookville, New York*.
- Miller, Ron (2005) Local Search Brings Results Home. *EContent*, Vol. 28, No. 5, p. 30.
- Orkut B., Junghoo C., Hector G., Luis G., Narayanan S. (1999) Exploiting Geographical Location Information of Web Pages, Department of Computer Science. *Stanford University*.
- Pan B., Hembrooke H., Joachims T., Lorigo L, Gay G., Granka L. (2007) In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication*, Vol. 12, No. 3, pp. 801-823.
- Sergey Brin (1999) The anatomy of a large-scale hypertextual Web search engine, Computer Science Department. *Stanford University*, Vol. 56, No. 18, pp. 3825-3833.
- Yoshiura Noriaki, Moroi Takanori (2012) Discovery and Analysis of Web Pages Censored by Google Search Engine in Japan. *Journal of Socio-Informatics*, Vol.5, No. 1, pp. 51-64.
- Zhou, Mingming (2014) Gender difference in web search perceptions and behavior: Does it vary by task performance? *Computers & Education*, Vol.78, pp.174-184.

- 7 Reasons Google Search Results Vary Dramatically,
<https://www.webpresencesolutions.net/7-reasons-google-search-results-vary-dramatically/>
-
- C4.5 algorithm, https://en.wikipedia.org/wiki/C4.5_algorithm
- Search Engines and Algorithms: Optimizing for MSN's RankNet Technology,
<https://web.archive.org/web/20170426011650/http://www.seochat.com:80/c/a/msn-optimization-help/search-engines-and-algorithms-optimizing-for-msns-ranknet-technology/>
- Google's Hilltop Algorithm – A Foundation for Modern SEO,
<https://www.searchenginejournal.com/hilltop-algorithm/253893/>
- Web search engine, https://en.wikipedia.org/wiki/Web_search_engine
- Weka, [https://en.wikipedia.org/wiki/Weka_\(machine_learning\)](https://en.wikipedia.org/wiki/Weka_(machine_learning))
- Why Do Different Browsers and Devices Show Different Search Results on Google?
<https://mikekhorev.com/why-do-different-browsers-show-different-search-results-on-google>

6. נספחים

- I. קוד תוכנת החיפוש: <https://github.com/assafey/hit-browser-search-cross-regions>
- II. שירותי AWS: <https://aws.amazon.com/>
- III. תוכנת Weka: <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>
- IV. תוכנת RStudio: <https://rstudio.com/products/rstudio/download/>
- V. טבלת גישוש מלאה:

Date	State/City	Search Query	Browser Type	Result Link
2019-01-05	Ohio, USA	Facebook	Chrome	https://www.facebook.com/
				https://twitter.com/facebook
				https://newsroom.fb.com/
				https://play.google.com/store/apps/details?id=com.facebook.katana&hl=en_US
				https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html
				https://www.newyorker.com/magazine/2018/09/17/can-mark-zuckerberg-fix-facebook-before-it-breaks-democracy
				https://www.wired.com/story/inside-facebook-mark-zuckerberg-2-years-of-hell/
				https://itunes.apple.com/us/app/facebook/id284882215?mt=8
2019-01-05	Ohio, USA	Facebook	Firefox	https://www.facebook.com/
				https://twitter.com/facebook
				https://newsroom.fb.com/

				https://play.google.com/store/apps/details?id=com.facebook.katana&hl=en_US
				https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html
				https://www.newyorker.com/magazine/2018/09/17/can-mark-zuckerberg-fix-facebook-before-it-breaks-democracy
				https://www.wired.com/story/inside-facebook-mark-zuckerberg-2-years-of-hell/
				https://itunes.apple.com/us/app/facebook/id284882215?mt=8
2019-01-05	Tokyo, Japan	Facebook	Chrome	https://ja-jp.facebook.com/
				https://ja.wikipedia.org/wiki/Facebook
				https://play.google.com/store/apps/details?id=com.facebook.katana&hl=ja
				https://newsroom.fb.com/
				https://ja.newsroom.fb.com/
				news.livedoor.com/Facebook/topics/keyword/14084/
				https://itunes.apple.com/jp/app/facebook/id284882215

2019-01-05	Tokyo, Japan	Facebook	Firefox	https://ja-jp.facebook.com
				https://ja.wikipedia.org/wiki/Facebook
				https://play.google.com/store/apps/details?id=com.facebook.katana&hl=ja
				https://newsroom.fb.com/
				https://ja.newsroom.fb.com/
				news.livedoor.com/Facebook/topics/keyword/14084/
				https://itunes.apple.com/jp/app/facebook/id284882215
2019-01-05	Ohio, USA	New York	Chrome	https://en.wikipedia.org/wiki/New_York_City
				https://www.nycgo.com/
				https://www.nyc.gov/
				https://www.iloveny.com/
				https://www.ny.gov/
				https://www.timeout.com/newyork
				https://www.nytimes.com/
				https://www.theatlantic.com/ideas/archive/2018/10/new-york-retail-vacancy/572911/
				https://www.timeout.com/newyork/attractions/best-beaches

				https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html
				https://www.nyu.edu/
				https://www.lonelyplanet.com/usa/new-york-city
2019-01-05	Ohio, USA	New York	Firefox	https://en.wikipedia.org/wiki/New_York_City
				https://www.nycgo.com/
				https://www.nyc.gov/
				https://www.iloveny.com/
				https://www.ny.gov/
				https://www.timeout.com/newyork
				https://www.nytimes.com/
				https://www.theatlantic.com/ideas/archive/2018/10/new-york-retail-vacancy/572911/
				https://www.timeout.com/newyork/attractions/best-beaches
				https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html
				https://www.nyu.edu/
				https://www.lonelyplanet.com/usa/new-york-city
2019-01-05	Tokyo, Japan	New York	Chrome	https://ja.wikipedia.org/wiki/ニューヨーク

				https://www.nycgo.com/
				https://ejje.weblio.jp/content/New+York
				https://ejje.weblio.jp/content/New+York+City
				https://www.newyork.jp/
				https://www.nyc.gov/
				https://www.iloveny.com/
				https://www.timeout.com/newyork
				https://www.nytimes.com/section/nyregion
				https://basikny.com/
2019-01-05	Tokyo, Japan	New York	Firefox	https://ja.wikipedia.org/wiki/ニューヨーク
				https://www.nycgo.com/
				https://ejje.weblio.jp/content/New+York
				https://www.newyork.jp/
				https://www.nyc.gov/
				https://www.iloveny.com/
				https://www.timeout.com/newyork
				https://www.nytimes.com/section/nyregion
				https://basikny.com/
				https://ny-pg.com/

.VI פלט תוכנת R - תוצאות מבחן T שוניות שוות

```
> t.test(distance, country+query+browsers, var.equal = TRUE)
```

Two Sample t-test

```
data: distance and country + query + browsers
t = 2.9774, df = 938, p-value = 0.002981
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.409195 6.858890
sample estimates:
mean of x mean of y
21.19362 17.05957
```

.VII פלט תוכנת R - תוצאות מבחן T שוניות שונות

```
> t.test(distance, country+query+browsers, var.equal = FALSE)
```

Welch Two Sample t-test

```
data: distance and country + query + browsers
t = 2.9774, df = 510.93, p-value = 0.003045
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.406249 6.861836
sample estimates:
mean of x mean of y
21.19362 17.05957
```

.VIII תוצאות רגרסיה לינארית

.IX תוצאות RotationForest:

Time taken to build model: 0.26 seconds

```
=== Stratified cross-validation ===
=== Summary ===
```

Correctly Classified Instances	343	72.9787 %
Incorrectly Classified Instances	127	27.0213 %
Kappa statistic	0.5099	
Mean absolute error	0.2533	
Root mean squared error	0.3629	
Relative absolute error	65.849 %	
Root relative squared error	82.7945 %	
Total Number of Instances	470	

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.873	0.337	0.775	0.873	0.821	0.554	0.843	0.859	small
	0.535	0.114	0.636	0.535	0.581	0.447	0.774	0.550	medium
	0.547	0.051	0.672	0.547	0.603	0.541	0.885	0.624	big
Weighted Avg.	0.730	0.231	0.721	0.730	0.721	0.523	0.831	0.738	

```
=== Confusion Matrix ===
```

```

a   b   c  <-- classified as
234 22 12 | a = small
 51 68  8 | b = medium
 17 17 41 | c = big
```

X. נתונים גולמיים של משתני דמה עבור רגרסיה:

https://docs.google.com/spreadsheets/d/e/2PACX-1vTsKJh5MiU13TIQ9NYIoV6T1VE6TNlaT6rID-UdSPu6d8Bg1Yq6Ec_zFyrbJf6e6FuHSZObaP4LH3-7/pub?output=csv

XI. נתונים גולמיים של תוצאות חיפוש:

Country	Query	Browsers	Distance
australia	newborn	chrome-firefox	14
australia	infant formula	chrome-firefox	88
australia	diapers	chrome-firefox	64
australia	colic	chrome-firefox	92
australia	bassinet	chrome-firefox	72
australia	labor	chrome-firefox	54
australia	teether	chrome-firefox	0
australia	baby dummy	chrome-firefox	66
australia	playpen	chrome-firefox	40
australia	baby bottle	chrome-firefox	64
australia	baby wipes	chrome-firefox	72
australia	carry cot	chrome-firefox	0
australia	tetra	chrome-firefox	76
australia	stroller	chrome-firefox	40
australia	bib	chrome-firefox	68
australia	diaper rash	chrome-firefox	67
australia	nanny	chrome-firefox	80
australia	baby carrier	chrome-firefox	0
australia	baby romper	chrome-firefox	0
australia	baby reflux	chrome-firefox	81
australia	newborn	chrome-internet explorer	14
australia	infant formula	chrome-internet explorer	12
australia	diapers	chrome-internet explorer	14
australia	colic	chrome-internet explorer	42
australia	bassinet	chrome-internet explorer	0
australia	labor	chrome-internet explorer	8
australia	teether	chrome-internet explorer	0
australia	baby dummy	chrome-internet explorer	8
australia	playpen	chrome-internet explorer	0
australia	baby bottle	chrome-internet explorer	38
australia	baby wipes	chrome-internet explorer	0

australia	carry cot	chrome-internet explorer	0
australia	tetra	chrome-internet explorer	0
australia	stroller	chrome-internet explorer	0
australia	bib	chrome-internet explorer	0
australia	diaper rash	chrome-internet explorer	10
australia	nanny	chrome-internet explorer	0
australia	baby carrier	chrome-internet explorer	68
australia	baby romper	chrome-internet explorer	0
australia	baby reflux	chrome-internet explorer	87
australia	newborn	firefox-internet explorer	14
australia	infant formula	firefox-internet explorer	48
australia	diapers	firefox-internet explorer	34
australia	colic	firefox-internet explorer	62
australia	bassinet	firefox-internet explorer	32
australia	labor	firefox-internet explorer	34
australia	teether	firefox-internet explorer	0
australia	baby dummy	firefox-internet explorer	36
australia	playpen	firefox-internet explorer	0
australia	baby bottle	firefox-internet explorer	12
australia	baby wipes	firefox-internet explorer	32
australia	carry cot	firefox-internet explorer	0
australia	tetra	firefox-internet explorer	36
australia	stroller	firefox-internet explorer	0
australia	bib	firefox-internet explorer	28
australia	diaper rash	firefox-internet explorer	37
australia	nanny	firefox-internet explorer	40
australia	baby carrier	firefox-internet explorer	68

australia	baby romper	firefox-internet explorer	0
australia	baby reflux	firefox-internet explorer	14
bahrain	newborn	chrome-firefox	61
bahrain	infant formula	chrome-firefox	86
bahrain	diapers	chrome-firefox	0
bahrain	colic	chrome-firefox	0
bahrain	bassinet	chrome-firefox	0
bahrain	labor	chrome-firefox	0
bahrain	teether	chrome-firefox	60
bahrain	baby dummy	chrome-firefox	0
bahrain	playpen	chrome-firefox	10
bahrain	baby bottle	chrome-firefox	68
bahrain	baby wipes	chrome-firefox	76
bahrain	carry cot	chrome-firefox	0
bahrain	tetra	chrome-firefox	0
bahrain	stroller	chrome-firefox	0
bahrain	bib	chrome-firefox	0
bahrain	diaper rash	chrome-firefox	71
bahrain	nanny	chrome-firefox	0
bahrain	baby carrier	chrome-firefox	72
bahrain	baby romper	chrome-firefox	57
bahrain	baby reflux	chrome-firefox	81
bahrain	newborn	chrome-internet explorer	25
bahrain	infant formula	chrome-internet explorer	10
bahrain	diapers	chrome-internet explorer	2
bahrain	colic	chrome-internet explorer	0
bahrain	bassinet	chrome-internet explorer	0
bahrain	labor	chrome-internet explorer	0
bahrain	teether	chrome-internet explorer	40
bahrain	baby dummy	chrome-internet explorer	0
bahrain	playpen	chrome-internet explorer	2
bahrain	baby bottle	chrome-internet explorer	0
bahrain	baby wipes	chrome-internet explorer	0
bahrain	carry cot	chrome-internet explorer	0
bahrain	tetra	chrome-internet explorer	0
bahrain	stroller	chrome-internet explorer	0
bahrain	bib	chrome-internet explorer	0

bahrain	diaper rash	chrome-internet explorer	18
bahrain	nanny	chrome-internet explorer	0
bahrain	baby carrier	chrome-internet explorer	0
bahrain	baby romper	chrome-internet explorer	61
bahrain	baby reflux	chrome-internet explorer	25
bahrain	newborn	firefox-internet explorer	31
bahrain	infant formula	firefox-internet explorer	46
bahrain	diapers	firefox-internet explorer	2
bahrain	colic	firefox-internet explorer	0
bahrain	bassinet	firefox-internet explorer	0
bahrain	labor	firefox-internet explorer	0
bahrain	teether	firefox-internet explorer	45
bahrain	baby dummy	firefox-internet explorer	0
bahrain	playpen	firefox-internet explorer	11
bahrain	baby bottle	firefox-internet explorer	28
bahrain	baby wipes	firefox-internet explorer	36
bahrain	carry cot	firefox-internet explorer	0
bahrain	tetra	firefox-internet explorer	0
bahrain	stroller	firefox-internet explorer	0
bahrain	bib	firefox-internet explorer	0
bahrain	diaper rash	firefox-internet explorer	41
bahrain	nanny	firefox-internet explorer	0
bahrain	baby carrier	firefox-internet explorer	32
bahrain	baby romper	firefox-internet explorer	13
bahrain	baby reflux	firefox-internet explorer	47
brazil	newborn	chrome-firefox	0
brazil	infant formula	chrome-firefox	0
brazil	diapers	chrome-firefox	0
brazil	colic	chrome-firefox	0
brazil	bassinet	chrome-firefox	0

brazil	labor	chrome-firefox	8
brazil	teether	chrome-firefox	0
brazil	baby dummy	chrome-firefox	0
brazil	playpen	chrome-firefox	0
brazil	baby bottle	chrome-firefox	0
brazil	baby wipes	chrome-firefox	0
brazil	carry cot	chrome-firefox	14
brazil	tetra	chrome-firefox	2
brazil	stroller	chrome-firefox	0
brazil	bib	chrome-firefox	0
brazil	diaper rash	chrome-firefox	0
brazil	nanny	chrome-firefox	0
brazil	baby carrier	chrome-firefox	0
brazil	baby romper	chrome-firefox	0
brazil	baby reflux	chrome-firefox	0
brazil	newborn	chrome-internet explorer	0
brazil	infant formula	chrome-internet explorer	0
brazil	diapers	chrome-internet explorer	0
brazil	colic	chrome-internet explorer	0
brazil	bassinet	chrome-internet explorer	0
brazil	labor	chrome-internet explorer	8
brazil	teether	chrome-internet explorer	0
brazil	baby dummy	chrome-internet explorer	0
brazil	playpen	chrome-internet explorer	0
brazil	baby bottle	chrome-internet explorer	0
brazil	baby wipes	chrome-internet explorer	0
brazil	carry cot	chrome-internet explorer	0
brazil	tetra	chrome-internet explorer	2
brazil	stroller	chrome-internet explorer	0
brazil	bib	chrome-internet explorer	0
brazil	diaper rash	chrome-internet explorer	0
brazil	nanny	chrome-internet explorer	0
brazil	baby carrier	chrome-internet explorer	0
brazil	baby romper	chrome-internet explorer	0

brazil	baby reflux	chrome-internet explorer	0
brazil	newborn	firefox-internet explorer	0
brazil	infant formula	firefox-internet explorer	0
brazil	diapers	firefox-internet explorer	0
brazil	colic	firefox-internet explorer	0
brazil	bassinet	firefox-internet explorer	0
brazil	labor	firefox-internet explorer	8
brazil	teether	firefox-internet explorer	0
brazil	baby dummy	firefox-internet explorer	0
brazil	playpen	firefox-internet explorer	0
brazil	baby bottle	firefox-internet explorer	0
brazil	baby wipes	firefox-internet explorer	0
brazil	carry cot	firefox-internet explorer	14
brazil	tetra	firefox-internet explorer	2
brazil	stroller	firefox-internet explorer	0
brazil	bib	firefox-internet explorer	0
brazil	diaper rash	firefox-internet explorer	0
brazil	nanny	firefox-internet explorer	0
brazil	baby carrier	firefox-internet explorer	0
brazil	baby romper	firefox-internet explorer	0
brazil	baby reflux	firefox-internet explorer	0
france	newborn	chrome-firefox	0
france	infant formula	chrome-firefox	0
france	diapers	chrome-firefox	0
france	colic	chrome-firefox	0
france	bassinet	chrome-firefox	12
france	labor	chrome-firefox	0
france	baby dummy	chrome-firefox	0
france	playpen	chrome-firefox	0
france	baby bottle	chrome-firefox	0
france	baby wipes	chrome-firefox	0
france	carry cot	chrome-firefox	0
france	tetra	chrome-firefox	0
france	stroller	chrome-firefox	0

france	bib	chrome-firefox	0
france	diaper rash	chrome-firefox	0
france	nanny	chrome-firefox	0
france	baby carrier	chrome-firefox	0
france	baby romper	chrome-firefox	0
france	baby reflux	chrome-firefox	0
france	newborn	chrome-internet explorer	0
france	infant formula	chrome-internet explorer	0
france	diapers	chrome-internet explorer	0
france	colic	chrome-internet explorer	14
france	bassinet	chrome-internet explorer	2
france	labor	chrome-internet explorer	39
france	teether	chrome-internet explorer	11
france	baby dummy	chrome-internet explorer	0
france	playpen	chrome-internet explorer	0
france	baby bottle	chrome-internet explorer	0
france	baby wipes	chrome-internet explorer	0
france	carry cot	chrome-internet explorer	0
france	tetra	chrome-internet explorer	14
france	stroller	chrome-internet explorer	0
france	bib	chrome-internet explorer	0
france	diaper rash	chrome-internet explorer	0
france	nanny	chrome-internet explorer	11
france	baby carrier	chrome-internet explorer	0
france	baby romper	chrome-internet explorer	11
france	baby reflux	chrome-internet explorer	0
france	newborn	firefox-internet explorer	0
france	infant formula	firefox-internet explorer	0
france	diapers	firefox-internet explorer	0
france	colic	firefox-internet explorer	14

france	bassinet	firefox-internet explorer	2
france	labor	firefox-internet explorer	39
france	baby dummy	firefox-internet explorer	0
france	playpen	firefox-internet explorer	0
france	baby bottle	firefox-internet explorer	0
france	baby wipes	firefox-internet explorer	0
france	carry cot	firefox-internet explorer	0
france	tetra	firefox-internet explorer	14
france	stroller	firefox-internet explorer	0
france	bib	firefox-internet explorer	0
france	diaper rash	firefox-internet explorer	0
france	nanny	firefox-internet explorer	11
france	baby carrier	firefox-internet explorer	0
france	baby romper	firefox-internet explorer	11
france	baby reflux	firefox-internet explorer	0
india	newborn	chrome-firefox	72
india	diapers	chrome-firefox	72
india	colic	chrome-firefox	86
india	bassinet	chrome-firefox	68
india	labor	chrome-firefox	88
india	teether	chrome-firefox	67
india	baby dummy	chrome-firefox	62
india	playpen	chrome-firefox	63
india	baby bottle	chrome-firefox	60
india	baby wipes	chrome-firefox	68
india	carry cot	chrome-firefox	0
india	tetra	chrome-firefox	64
india	stroller	chrome-firefox	40
india	bib	chrome-firefox	62
india	diaper rash	chrome-firefox	0
india	nanny	chrome-firefox	64
india	baby carrier	chrome-firefox	70
india	baby romper	chrome-firefox	72
india	baby reflux	chrome-firefox	46
india	newborn	chrome-internet explorer	12
india	infant formula	chrome-internet explorer	24

india	diapers	chrome-internet explorer	12
india	colic	chrome-internet explorer	32
india	bassinet	chrome-internet explorer	0
india	labor	chrome-internet explorer	26
india	teether	chrome-internet explorer	18
india	baby dummy	chrome-internet explorer	44
india	playpen	chrome-internet explorer	18
india	baby bottle	chrome-internet explorer	0
india	baby wipes	chrome-internet explorer	0
india	carry cot	chrome-internet explorer	0
india	tetra	chrome-internet explorer	12
india	stroller	chrome-internet explorer	0
india	bib	chrome-internet explorer	4
india	diaper rash	chrome-internet explorer	0
india	nanny	chrome-internet explorer	12
india	baby carrier	chrome-internet explorer	6
india	baby romper	chrome-internet explorer	12
india	baby reflux	chrome-internet explorer	46
india	newborn	firefox-internet explorer	38
india	diapers	firefox-internet explorer	32
india	colic	firefox-internet explorer	56
india	bassinet	firefox-internet explorer	28
india	labor	firefox-internet explorer	58
india	teether	firefox-internet explorer	37
india	baby dummy	firefox-internet explorer	43
india	playpen	firefox-internet explorer	33
india	baby bottle	firefox-internet explorer	20

india	baby wipes	firefox-internet explorer	28
india	carry cot	firefox-internet explorer	0
india	tetra	firefox-internet explorer	34
india	stroller	firefox-internet explorer	0
india	bib	firefox-internet explorer	22
india	diaper rash	firefox-internet explorer	0
india	nanny	firefox-internet explorer	34
india	baby carrier	firefox-internet explorer	25
india	baby romper	firefox-internet explorer	38
india	baby reflux	firefox-internet explorer	46
israel	newborn	chrome-firefox	0
israel	infant formula	chrome-firefox	0
israel	diapers	chrome-firefox	4
israel	colic	chrome-firefox	0
israel	bassinet	chrome-firefox	0
israel	labor	chrome-firefox	22
israel	teether	chrome-firefox	0
israel	baby dummy	chrome-firefox	39
israel	playpen	chrome-firefox	0
israel	baby bottle	chrome-firefox	64
israel	baby wipes	chrome-firefox	0
israel	carry cot	chrome-firefox	0
israel	tetra	chrome-firefox	0
israel	stroller	chrome-firefox	0
israel	bib	chrome-firefox	0
israel	diaper rash	chrome-firefox	0
israel	nanny	chrome-firefox	0
israel	baby carrier	chrome-firefox	0
israel	baby romper	chrome-firefox	0
israel	newborn	chrome-internet explorer	42
israel	infant formula	chrome-internet explorer	53
israel	diapers	chrome-internet explorer	30
israel	colic	chrome-internet explorer	33
israel	bassinet	chrome-internet explorer	35
israel	labor	chrome-internet explorer	49
israel	teether	chrome-internet explorer	42

israel	baby dummy	chrome-internet explorer	41
israel	playpen	chrome-internet explorer	23
israel	baby bottle	chrome-internet explorer	33
israel	baby wipes	chrome-internet explorer	65
israel	carry cot	chrome-internet explorer	57
israel	tetra	chrome-internet explorer	33
israel	stroller	chrome-internet explorer	38
israel	bib	chrome-internet explorer	45
israel	diaper rash	chrome-internet explorer	40
israel	nanny	chrome-internet explorer	61
israel	baby carrier	chrome-internet explorer	40
israel	baby romper	chrome-internet explorer	34
israel	baby reflux	chrome-internet explorer	71
israel	newborn	firefox-internet explorer	42
israel	infant formula	firefox-internet explorer	53
israel	diapers	firefox-internet explorer	30
israel	colic	firefox-internet explorer	33
israel	bassinet	firefox-internet explorer	35
israel	labor	firefox-internet explorer	44
israel	teether	firefox-internet explorer	42
israel	baby dummy	firefox-internet explorer	25
israel	playpen	firefox-internet explorer	23
israel	baby bottle	firefox-internet explorer	43
israel	baby wipes	firefox-internet explorer	65
israel	carry cot	firefox-internet explorer	57
israel	tetra	firefox-internet explorer	33
israel	stroller	firefox-internet explorer	38

israel	bib	firefox-internet explorer	45
israel	diaper rash	firefox-internet explorer	40
israel	nanny	firefox-internet explorer	61
israel	baby carrier	firefox-internet explorer	40
israel	baby romper	firefox-internet explorer	34
japan	infant formula	chrome-firefox	0
japan	diapers	chrome-firefox	0
japan	colic	chrome-firefox	4
japan	bassinet	chrome-firefox	10
japan	labor	chrome-firefox	10
japan	teether	chrome-firefox	25
japan	baby dummy	chrome-firefox	0
japan	playpen	chrome-firefox	2
japan	baby bottle	chrome-firefox	0
japan	baby wipes	chrome-firefox	0
japan	carry cot	chrome-firefox	0
japan	tetra	chrome-firefox	0
japan	stroller	chrome-firefox	0
japan	bib	chrome-firefox	0
japan	diaper rash	chrome-firefox	0
japan	nanny	chrome-firefox	0
japan	baby carrier	chrome-firefox	0
japan	baby romper	chrome-firefox	14
japan	baby reflux	chrome-firefox	14
japan	infant formula	chrome-internet explorer	0
japan	diapers	chrome-internet explorer	0
japan	colic	chrome-internet explorer	6
japan	bassinet	chrome-internet explorer	10
japan	labor	chrome-internet explorer	10
japan	teether	chrome-internet explorer	0
japan	baby dummy	chrome-internet explorer	0
japan	playpen	chrome-internet explorer	2
japan	baby bottle	chrome-internet explorer	0
japan	baby wipes	chrome-internet explorer	0
japan	carry cot	chrome-internet explorer	0
japan	tetra	chrome-internet explorer	0

japan	stroller	chrome-internet explorer	0
japan	bib	chrome-internet explorer	0
japan	diaper rash	chrome-internet explorer	0
japan	nanny	chrome-internet explorer	0
japan	baby carrier	chrome-internet explorer	32
japan	baby romper	chrome-internet explorer	0
japan	baby reflux	chrome-internet explorer	14
japan	infant formula	firefox-internet explorer	0
japan	diapers	firefox-internet explorer	0
japan	colic	firefox-internet explorer	2
japan	bassinet	firefox-internet explorer	0
japan	labor	firefox-internet explorer	0
japan	teether	firefox-internet explorer	25
japan	baby dummy	firefox-internet explorer	0
japan	playpen	firefox-internet explorer	4
japan	baby bottle	firefox-internet explorer	0
japan	baby wipes	firefox-internet explorer	0
japan	carry cot	firefox-internet explorer	0
japan	tetra	firefox-internet explorer	0
japan	stroller	firefox-internet explorer	0
japan	bib	firefox-internet explorer	0
japan	diaper rash	firefox-internet explorer	0
japan	nanny	firefox-internet explorer	0
japan	baby carrier	firefox-internet explorer	32
japan	baby romper	firefox-internet explorer	14
japan	baby reflux	firefox-internet explorer	2
japan	newborn	firefox-internet explorer	0
usa	newborn	chrome-firefox	63

usa	infant formula	chrome-firefox	262
usa	diapers	chrome-firefox	72
usa	colic	chrome-firefox	80
usa	bassinet	chrome-firefox	68
usa	labor	chrome-firefox	100
usa	teether	chrome-firefox	60
usa	baby dummy	chrome-firefox	60
usa	playpen	chrome-firefox	26
usa	baby bottle	chrome-firefox	61
usa	baby wipes	chrome-firefox	68
usa	carry cot	chrome-firefox	51
usa	tetra	chrome-firefox	62
usa	stroller	chrome-firefox	55
usa	bib	chrome-firefox	66
usa	diaper rash	chrome-firefox	70
usa	nanny	chrome-firefox	75
usa	baby carrier	chrome-firefox	52
usa	baby romper	chrome-firefox	0
usa	baby reflux	chrome-firefox	138
usa	infant formula	chrome-internet explorer	46
usa	diapers	chrome-internet explorer	0
usa	colic	chrome-internet explorer	46
usa	bassinet	chrome-internet explorer	2
usa	labor	chrome-internet explorer	2
usa	teether	chrome-internet explorer	41
usa	baby dummy	chrome-internet explorer	39
usa	playpen	chrome-internet explorer	26
usa	baby bottle	chrome-internet explorer	18
usa	baby wipes	chrome-internet explorer	0
usa	carry cot	chrome-internet explorer	0
usa	tetra	chrome-internet explorer	18
usa	stroller	chrome-internet explorer	23
usa	bib	chrome-internet explorer	0
usa	diaper rash	chrome-internet explorer	8
usa	nanny	chrome-internet explorer	18
usa	baby carrier	chrome-internet explorer	4

usa	baby romper	chrome-internet explorer	0
usa	baby reflux	chrome-internet explorer	140
usa	infant formula	firefox-internet explorer	142
usa	diapers	firefox-internet explorer	32
usa	colic	firefox-internet explorer	60
usa	bassinet	firefox-internet explorer	28
usa	labor	firefox-internet explorer	60
usa	teether	firefox-internet explorer	38
usa	baby dummy	firefox-internet explorer	46
usa	playpen	firefox-internet explorer	16
usa	baby bottle	firefox-internet explorer	28
usa	baby wipes	firefox-internet explorer	28
usa	carry cot	firefox-internet explorer	21
usa	tetra	firefox-internet explorer	14
usa	stroller	firefox-internet explorer	8
usa	bib	firefox-internet explorer	26
usa	diaper rash	firefox-internet explorer	40
usa	nanny	firefox-internet explorer	45
usa	baby carrier	firefox-internet explorer	22
usa	baby romper	firefox-internet explorer	0
usa	baby reflux	firefox-internet explorer	118