

שיעור 2 - מנועי חיפוש



★ המטרות של מנוע חיפוש

★ איך מנוע חיפוש עובד

★ דינמיות של מנועי חיפוש

מהן המטרות של מנוע חיפוש?

★ למכור שטחי פרסום ולעשות ככל האפשר יותר כסף

Revenues	Full Year			2013			
	2011	2012	(unaudited) 2013	Q1	(unaudited) Q2	Q3	Q4
Google Websites	\$26,145	\$31,221	\$37,453	\$8,640	\$8,868	\$9,394	\$10,551
Y/Y Growth Rate	34%	19%	20%	18%	18%	22%	22%
Q/Q Growth Rate	NA	NA	NA	0%	3%	6%	12%
Google Network Members' Websites	\$10,386	\$12,465	\$13,125	\$3,262	\$3,193	\$3,148	\$3,522
Y/Y Growth Rate	18%	20%	5%	12%	7%	0%	3%
Q/Q Growth Rate	NA	NA	NA	-5%	-2%	-1%	12%
Total Advertising Revenues	\$36,531	\$43,686	\$50,578	\$11,902	\$12,061	\$12,542	\$14,073
Y/Y Growth Rate	29%	20%	16%	16%	15%	15%	17%
Q/Q Growth Rate	NA	NA	NA	-1%	1%	4%	12%

לקריאת הדו"ח המלא של גוגל:

<http://investor.google.com/financial/tables.html>

מהן המטרות של מנוע חיפוש?

★ לרכוש את אמון הגולש כדי שיחזור שוב ושוב ושוב...

★ לספק את התשובות הרלבנטיות ביותר לכל שאילתא

התפקידים של מנוע חיפוש

תפקידים בלתי תלויים

1. סריקה ובניית אינדקס

2. "מכונת תשובות"

סריקה ובניית אינדקס

מנוע חיפוש חייב לסרוק את כל האינטרנט ולבנות
אינדקס מקיף של כל המידע באמצעות:

Robots, Spiders, Web Crawlers



★ Seeds - רשימת urls המשמשים כנקודות מוצא לרובוטים

★ איתור כל הלינקים בדף הנסרק והוספתם לרשימת ה-urls לביקור. הרשימה נקראת "חזית הסריקה" crawl frontier

★ ביקור בכל לינק בתהליך המבוסס על אלגוריתם רקורסיבי

★ הורדת החומר לאינדקס מנוע החיפוש

סריקה ובניית אינדקס

מנועי חיפוש מגדירים לעצמם "מדיניות זחילה"
Crawling policy:



★ מדיניות ברירה (selection) - אילו דפים/חומרים לסרוק ו"לאנדקס".

★ מדיניות ביקורים חוזרים (re-visit) - באיזו תדירות יש לסרוק כל דף ודף

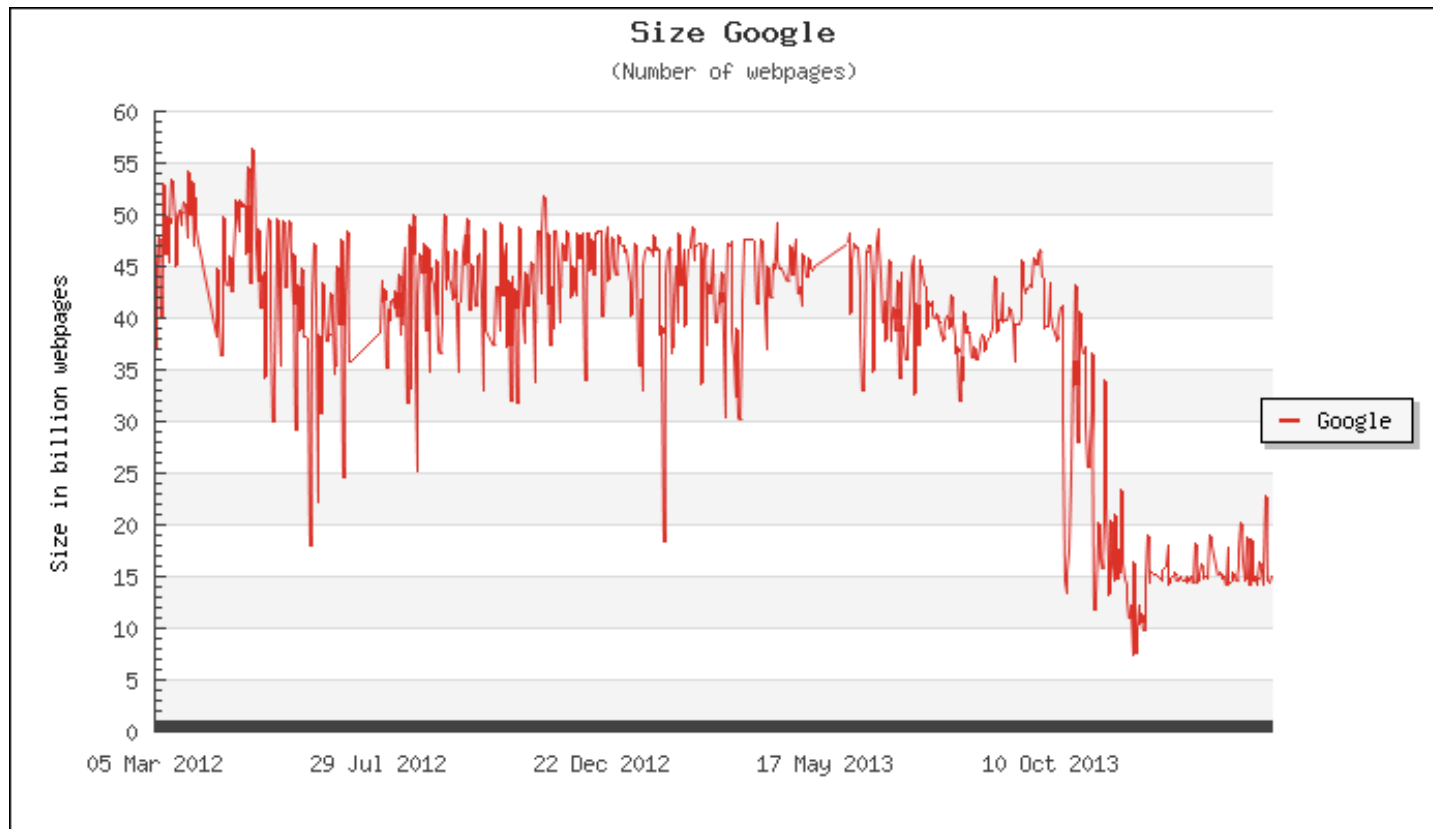
★ מדיניות נימוס (politeness) - האופן שבו הרובוט סורק את האתר מבלי לפגוע בפעולתו

★ מדיניות "מיקבול" (parallelisation) - האופן בו המשימה מתחלקת על פני מעבדים רבים מבלי שיוצרו הורדות כפולות של אותו החומר (פחות רלבנטי לנו)

האם הרובוטים עושים עבודת סריקה טובה?

הערכת מספר הדפים באינקס של גוגל בשנתיים האחרונות

(מקור: www.worldwidewebsize.com)



ומה עם הרשת האפלה? (dark web, invisible web, deep web...)

- מוערכת כגדולה בסדרי גודל מהרשת הנראית
- סיבות לדפים בלתי נראים: רשתות פרטיות, היעדר קישוריות, מבנה אתר פגום, דפים דינמיים, חסימות מכוונות ושגויות, תוכן שאינו ניתן להבנה ע"י הרובוט...

סריקה ובניית אינדקס

אנחנו וה-crawler

- ★ חייבים לדאוג למבנה אתר ודפים התומך בסריקת התוכן
- ★ יכולים להודיע לעכביש אילו דפים לסרוק ואילו לא, אילו לאנדקס ואילו לא, לינקים שאין לעקוב אחריהם.
- ★ יכולים להודיע לזוחל מהי החשיבות היחסית של דפים באתר שלנו (מספר בין 0 ל-1 כאשר 0.5 הוא ברירת המחדל)
- ★ יכולים לדווח לרובוט מהי תדירות העדכון האפיינית של כל דף (always, hourly, daily, weekly, monthly, yearly, never)
- ★ יכולים למנוע (עקרונית) מרובוטים מסויימים את סריקת האתר

סריקה ובניית אינדקס

איך מתקשרים עם הרובוט

★ robot.txt - מאוחסן בראש האתר ובוא ניתן לקבוע לאיזה רובוט מותר ולאיזה אסור לסרוק את האתר (user-agent), דפים להמנע מהם, מעקב אחרי לינקים וכו'

★ sitemap.xml - קובץ המכיל רשימה של כל קבצי האתר עם דירוג חשיבות יחסית (priority) ותדירות עדכון אפיינית

★ שימוש בתגית html בחלק ה-head של האתר למניעת אינדקס או מעקב לינקים בדף מסויים
<"META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">

סריקה ובניית אינדקס

לקריאה נוספת

איך הרובוטים עובדים: http://en.wikipedia.org/wiki/Web_crawler

מפות אתר Site Maps: http://en.wikipedia.org/wiki/Site_map

קבצי robot.txt - <http://en.wikipedia.org/wiki/Robot.txt>

הרשת האפלה - http://en.wikipedia.org/wiki/Deep_Web

תגית "מטא" לרובוטים - <http://www.robotstxt.org/meta.html>

מנוע חיפוש: מתן מענה לשאלות



מנוע החיפוש סורק את האינדקס ושולף ממנו את התוצאות המתאימות ביותר לשאלתא המסויימת. לפי מה?

1. רלבנטיות Relevancy

2. חשיבות Importance

3. עמידנות Endurance

רלבנטיות



המטרה של מנוע החיפוש היא "להבין" על
מה כל דף "מדבר" ומה ההקשר שלו באתר.

★ ניתוח טקסט באמצעות אלגוריתמים של בינה
מלאכותית

★ ניתוח הקשרים בתוך האתר

★ ניתוח האתר בהקשר לרשת כולה

★ הבאה בחשבון של הסטוריית גלישה וחיפוש, מיקום
גיאוגרפי, נתונים אישיים, פלטפורמה וכו'

★ אלגוריתמים מורכבים המביאים בחשבון מאות
פרמטרים

חשיבות



מהי החשיבות היחסית של התוצאות
הרלבנטיות שהתקבלו?

גוגל היו הראשונים שקשרו בין חשיבות
לפופולריות.
פופולריות = כמה מדברים עליך



מנוע חיפוש - תוצאות

SERP - Search Engine Result Positions

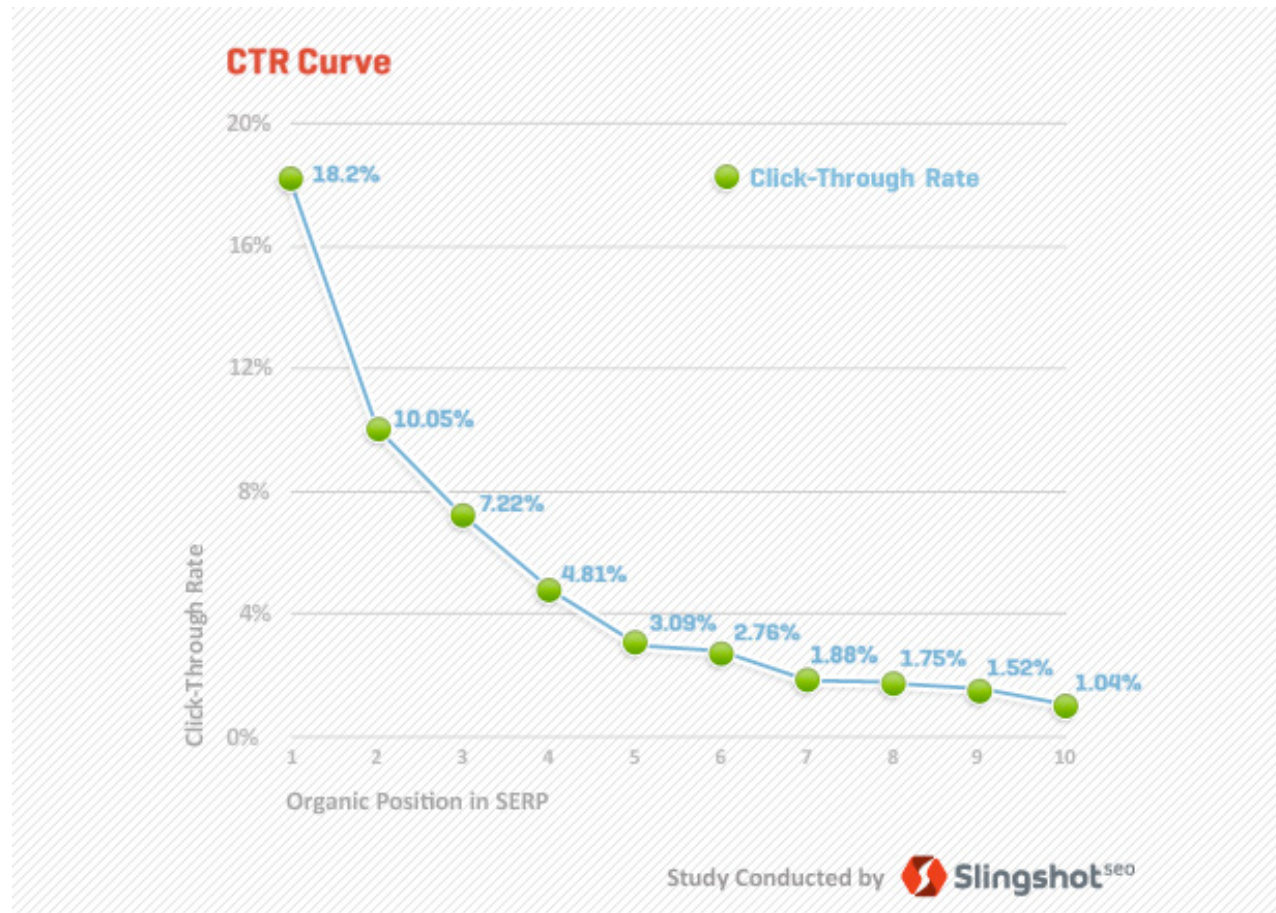
ככל שאנחנו מאמינים יותר למנוע החיפוש
כך ה-serp הופך להיות חשוב יותר

SEO = better SERP

קידום אתר = מיקום גבוה יותר בתוצאות

גוגל ואנחנו

ניתוח ההקלקות על ה-SERP מראה שהאמון שלנו בגוגל הוא כמעט מוחלט



מקור: <http://moz.com/blog/mission-imposserpble-establishing-clickthrough-rates>

מסקנות

- ★ הרוב המוחלט של הגולשים סומך באופן עיוור על גוגל
- ★ המקום הראשון ב-serp מקבל בערך פי 20 יותר תנועה מהמקום ה-10! אין בכלל מה לדבר על העמודים הפנימיים ב-serp.
- ★ התפקיד של קידום האתר הוא להשיג תוצאות טובות יותר ב-serp.
- ★ הנקודות בהן פועל מקדם האתר הן: עזרה לרובוטים בסריקת חומרי המטרה, שיפור הרלבנטיות של התכנים והאתר, העלאת חשיבות האתר.

לקידום אתר יש פוטנציאל להכניס הרבה כסף לעסק

הדינמיות של מנועי חיפוש

- ★ משכללים כל הזמן את איכות ויעילות הסריקה
- ★ משפרים את האלגוריתמים הקובעים רלבנטיות וחשיבות
- ★ מוסיפים עוד ועוד פרמטרים אישיים בהתאם ליכולת הגוברת להוציא מידע אישי
- ★ יחסי "אהבה-שנאה" עם מקדמי האתרים. איתור מתמיד אחר טכניקות black hat ו-grey hat ופיתוח מנגנוני "ענישה"
- ★ צורך בשינויים בגלל תחרות מול מנועים אחרים
- ★ שינויים מסיביים של אלגוריתמים בתדירות בלתי ידועה

קידום אתר הוא תהליך דינמי ואינו מסתיים לעולם

לקריאה אודות הסטורית העדכונים של גוגל: <http://moz.com/google-algorithm-change>

במה תלויה הצלחת קידום אתר?

- ★ איכות האתר
- ★ איכות התוכן
- ★ איכות המקדם
- ★ תקציב הקידום
- ★ מיהם המתחרים
- ★ מה עושים המתחרים
- ★ מה עושה גוגל...

לא ניתן להתחייב להשגת מיקום מסויים!!!