מטריצה של Google

שתי בעיות של מנועי חיפוש:

* כל התשובות שגוגל יודע לתת, גם מנועים אחרים יודעים לתת – הם פשוט לא יודעים לשים את זה במקום הנכון. הייחוד של גוגל זה שהוא יודע לדרג את התשובות בצורה טובה.
* בהינתן טקסט, איך להבין אם הוא רלוונטי או לא.

גוגל מתחבר לכל דף ודף, ומסמן אם הוא רלוונטי או לא. ברגע ששולחים חיפוש, גוגל כבר יודע את התשובה. יש לו רשימה של כל האתרים הרלוונטים.

גוגל מפעיל דירוג על כל אתר. החידוש של גוגל זה שבהתחלה, גוגל בכלל לא מתייחס לתוכן האתר, אלא לכמה הוא חשוב – כלומר כמה לינקים יש אליו מאתרים אחרים.

לכן, אם למשל מייקרוסופט יחליטו שהאתר שלהם עוסק בחתולים, וישימו שם סיפורים על חתולים – מהר מאוד האתר של מייקרוסופט יגיע למקום הראשון בחיפוש על חתולים, פשוט כי יש המון לינקים שמובילים לשם, בלי קשר לכמות או לאיכות המידע על חתולים שיש שם.

רוצים להתייחס לא רק לכמות הלינקים שמובילים לאתר, אלא לחשיבות של האתרים שמכילים את אותם לינקים. אם למשל האתר של מייקרוסופט או של גוגל מקשר לדף מסויים, זה נותן לו יותר חשיבות מאשר אם סתם אתר אנונימי מקשר לשם.

# אלגוריתם PageRank(על שם לארי פייג)

דפים – קודקודים של גרף של אינטרנט:   
לקודקוד נסמן – כל הלינקים ל. – מספר לינקים יוצאים מ  
רוצים לתת חשיבות לכל דף כך ש::  
נרצה לפתור מערכת משוואות: לכל ,

## דוגמה

1

3

2

*זה יוצר לנו מטריצה – רוצים למצוא לא פתרון לא טריוויאלי.*

*מטריצה כך ש:*

מתקיים: , לכל j. זה נקרא מאירצה סתוכלסטית(לפי עמודות)

# משפט

לA(או לכל מטריצה סתוכסטית) הוא ע"ע

## תוצאה

קיים פתרון לא טריוויאלי.

## הוכחה

נתבונן ב ו.  
לA ול יש אותם ע"ע ⇦ לA יש ע"ע 1

# נחזור לדוגמה

⇦ זה פתרון שמקיים

**בעיות**: מה קורה אם מקבלים פתרון עם ערכים שליליים? או יותר מפתרון אחד?(כפילות בע"ע)

פתרון: זה לא יכול לקרות במטריצות סתוכלסטיות.

# משפט

יהי M מטריצה סתוכלסטית רגולרית(ז"א קיים כך של כך האיברים הם חיוביים(גרף הוא קשיר)). אזי:

1. ע"ע של M עם ריבוע 1.  
   קיים ו"ע של M עם הוא אי שלילי.  
   (⇦ אפשר לבחור אותו מקיים )
2. כל הו"ע האחרים מקיימים ואם v הוא ווקטור עצמי (עם ) אזי

## הסבר

רוצים לדעת את הפתרון למשוואה   
זה נקרא "random walk"

# משפט

כאשר ו ו.

הסבר: זה נותן הסתברות לאיפה גולש אקראי הגיע בסוף יום של גלישה.

זוהי התכנסות אקפוננציאלית – לכן זה מתכנס מאוד מהר, אפילו שהאינטרנט הוא עצום.

# "למה"

לA יש מרחב עצמי

## הוכחה

⇦

# "למה" 2

לצימצום של יש 2 ע"ע.

## הוכחה

"קיים" בסיס כך ש:

ע"ע 1 נשאר 1, אבל שכן , ולכן

יהי ו ו"ע של A עם ו, . אזי:  
|

אפשר לפתור . אפשר לבחור ולחשב

גוגל יודע לחשב את התשובה כי מטריצה A היא לא מטריצה סתכסטית אקראית. גוגל עושה שינויים במטריצה בשביל לייעל את החישובים.

בנוסף, המטריצה די דלילה – באתר אינטרנט ממוצע יש 50 לינקים. זה אומר שכמעט בכל עמודה יש מעט מאוד איברים שונים מאפס.

שיעור חזרה ביום חמישי 23/6 בשעה 16:00(לא סופי, תתפרסם הודעה)