



Ανάκληση Πληροφορίας

Information Retrieval

Διδάσκων –
Δημήτριος Κατσαρός



Εισαγωγή στο μάθημα

&

Εισαγωγή στην Ανάκτηση
Πληροφορίας

Διδακτικό βοήθημα 1

- Καλύπτει το αντικείμενο του μαθήματος με θέμα:
 - Διαβάθμιση με βάση την ανάλυση υπερσυνδέσμων (Link Analysis Ranking)
- Είναι στην ελληνική
- 1η έκδοση (2010)
- Αναλυτικό, αυστηρά θεμελιωμένο, αλλά κατανοητό

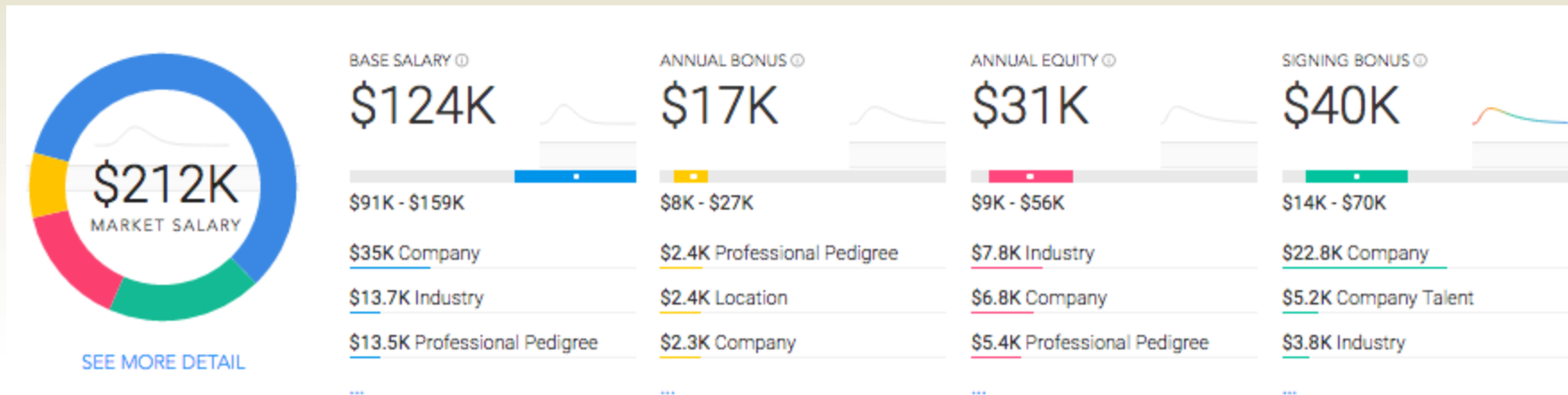
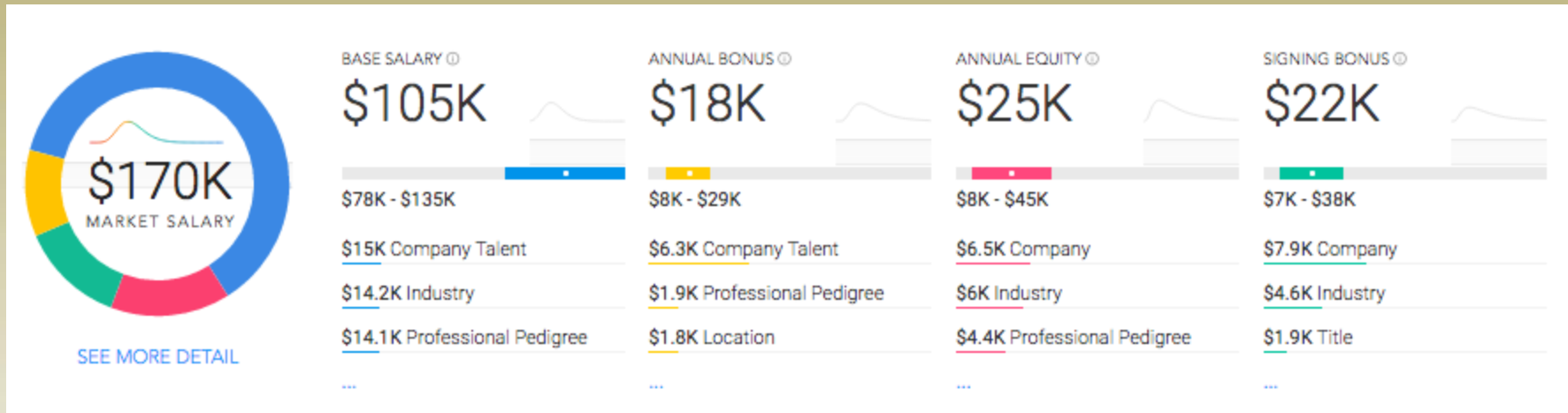


Διδακτικό βοήθημα 2

- Καλύπτει το αντικείμενο του μαθήματος με θέμα:
 - Διαβάθμιση με βάση το περιεχόμενο (Αντεστραμμένο Ευρετήριο – Inverted index)
- Είναι στην ελληνική (ΔΩΡΕΑΝ το pdf της αγγλικής)
- 1η έκδοση (2012)
- Αναλυτικό & κατανοητό, διδακτικές ασκήσεις



Αγορά εργασίας (2016): Νέοι απόφοιτοι σε Google και Facebook (δεδομένα από Paysa)





Αγορά εργασίας (2016): Google και Yahoo! Software Engineers (δεδομένα από Payscale)

What is the average total annual pay for Google Software Engineer?

The average market salary for Google Software Engineer is \$233K per year, ranging from \$161K to \$310K. Average market salary includes \$132K Base Salary, \$22.2K Annual Bonus, \$20.1K Signing Bonus and \$59.4K Annual Equity. See below for additional details.

Average Market Salary



BASE SALARY

\$132K

\$105K - \$160K

ANNUAL BONUS

\$22K

\$12K - \$33K

ANNUAL EQUITY

\$59K

\$29K - \$94K

SIGNING BONUS

\$20K

\$11K - \$30K

What is the average total annual pay for Yahoo! Software Engineer?

The average market salary for Yahoo! Software Engineer is \$164K per year, ranging from \$116K to \$215K. Average market salary includes \$111K Base Salary, \$13.1K Annual Bonus, \$15.1K Signing Bonus and \$25.2K Annual Equity. See below for additional details.

Average Market Salary



BASE SALARY

\$111K

\$89K - \$134K

ANNUAL BONUS

\$13K

\$7K - \$20K

ANNUAL EQUITY

\$25K

\$11K - \$41K

SIGNING BONUS

\$15K

\$8K - \$23K

Google

- ❖ 7% studied at Carnegie Mellon University
- ❖ 70% know Java
- ❖ 83% have a Bachelors degree
- ❖ 37% are Asian

Yahoo

- ❖ 18% studied at University of Southern California
- ❖ 72% know Java
- ❖ 81% have a Bachelors degree
- ❖ 54% are Asian

Αγορά εργασίας (2016): Google και Yahoo! Search Engineers (δεδομένα από Paysa)



Google
Search Engineer

4

RANK



\$203K
MARKET SALARY

Base Salary	\$124K
Annual Bonus	\$18K
Annual Equity	\$47K
Signing Bonus	\$14K

Google

- ❖ 36% of Search Engineers studied at M.I.T.
- ❖ 78% Search Engineers know Distributed systems
- ❖ 85% Search Engineers have a Bachelors degree
- ❖ 54% of Search Engineers are white



Yahoo!
Search Engineer



\$189K
MARKET SALARY

Base Salary	\$124K
Annual Bonus	\$16K
Annual Equity	\$33K
Signing Bonus	\$15K

Yahoo

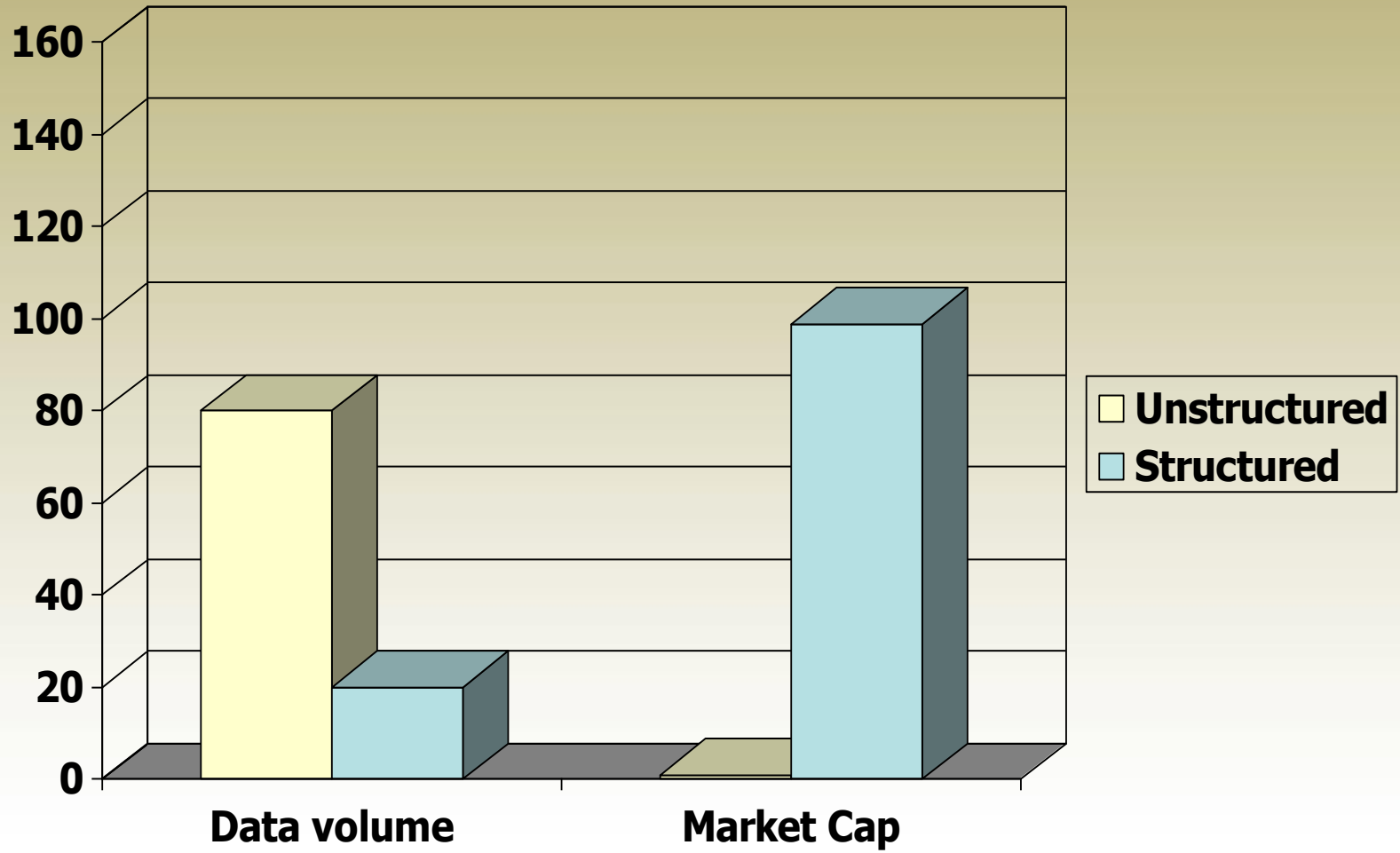
- ❖ 14% of Search Engineers studied at University of California Los Angeles
- ❖ 64% of Search Engineers know Scalability
- ❖ 69% of Search Engineers have a Bachelors degree
- ❖ 51% of Search Engineers are Asian



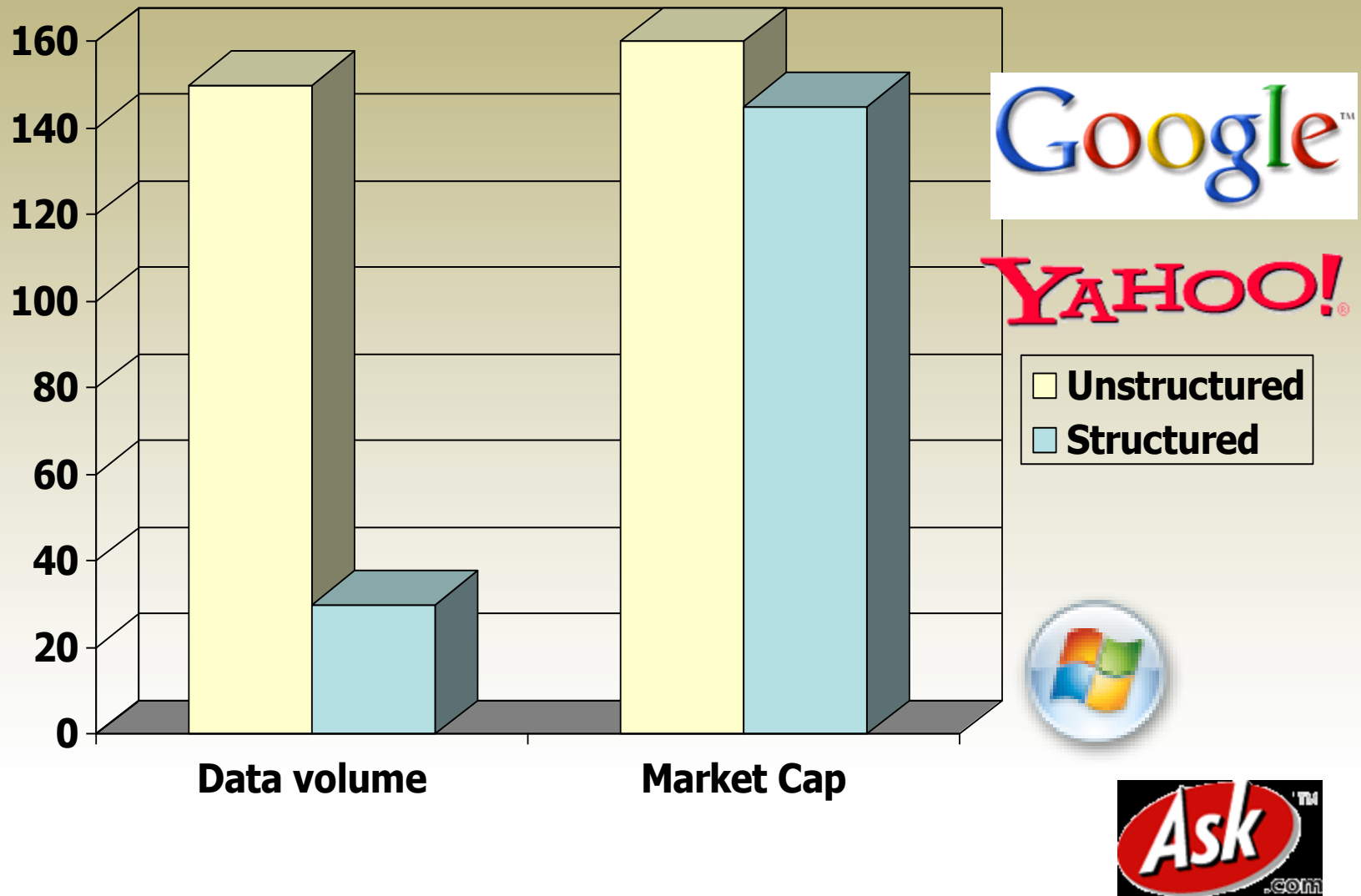
Εισαγωγικά



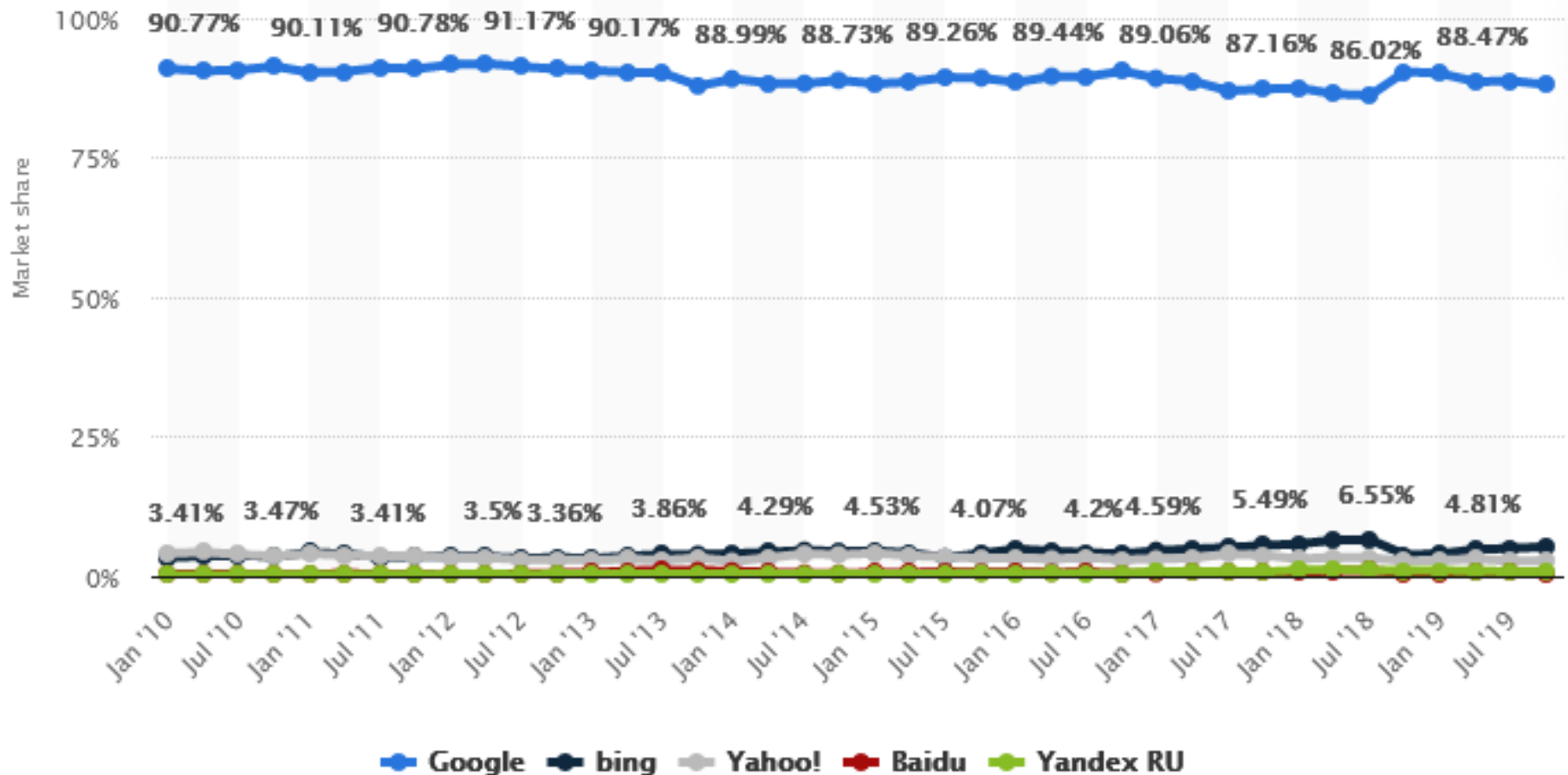
Αδόμητα (text) vs. Δομημένα (database) δεδομένα το 1996



Αδόμητα (text) vs. Δομημένα (database) δεδομένα το 2006

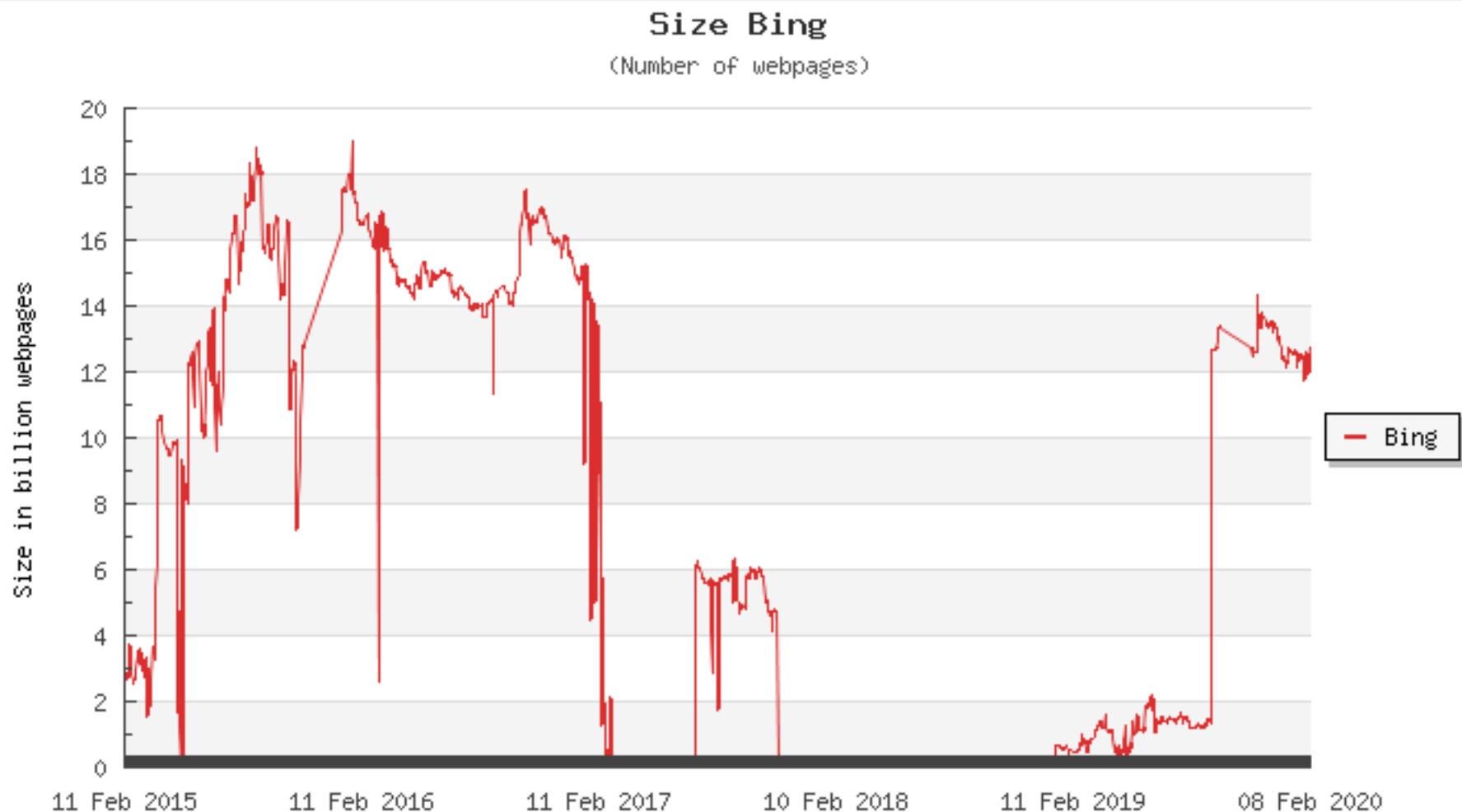


Μερίδιο αγοράς στην Αναζήτηση στο Web σε desktops



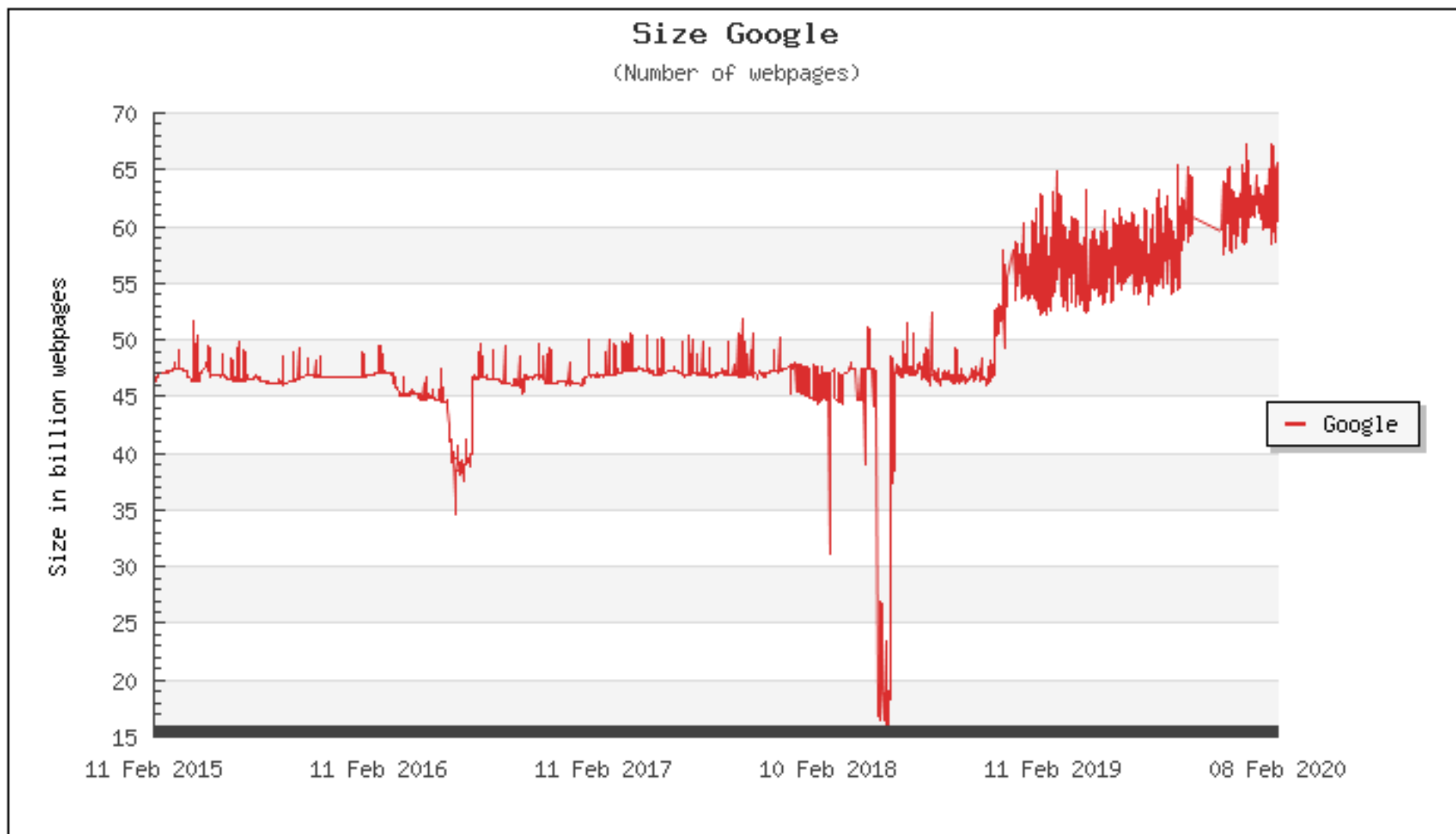
Μέγεθος του index της Bing (2015-'20)

(<http://worldwidewebsize.com/>)



Μέγεθος του index της Google (2015-'20)

(<http://worldwidewebsize.com/>)





Αδόμητα δεδομένα το 1650

- Ποια έργα του Shakespeare περιέχουν τις λέξεις *Brutus AND Caesar* αλλά *NOT Calpurnia*?
- Θα μπορούσαμε να κάνουμε grep όλα τα έργα του Shakespeare για τις λέξεις *Brutus* και *Caesar*, και κατόπιν να διώξουμε τις γραμμές που περιέχουν την λέξη *Calpurnia*?
 - Αργή προσέγγιση (για μεγάλες συλλογές)
 - Το *NOT Calpurnia* δεν είναι εύκολο στον χειρισμό
 - Άλλες λειτουργίες (π.χ., εύρεση της λέξης *Romans* κοντά στην *countrymen*) δεν είναι εφικτές
 - Διατεταγμένη ανάκτηση (ranked retrieval) (να επιστραφούν τα καλύτερα έγγραφα)
 - Σε επόμενες διάλέξεις



Πίνακας Terms-Documents

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	1	1	0	0	0	1
Brutus	1	1	0	1	0	0
Caesar	1	1	0	1	1	1
Calpurnia	0	1	0	0	0	0
Cleopatra	1	0	0	0	0	0
mercy	1	0	1	1	1	1
worser	1	0	1	1	1	0

Brutus AND Caesar but NOT Calpurnia

1 εάν το έργο
περιέχει την λέξη,
0 διαφορετικά



Διανύσματα όρων

- Έτσι, έχουμε ένα διάνυσμα με 0/1 για κάθε όρο
- Για να απαντήσουμε σε ένα ερώτημα: παίρνουμε τα διανύσματα για τους όρους ***Brutus, Caesar*** και ***Calpurnia*** (συμπληρωματικά) \rightarrow bitwise *AND*
- $110100 \text{ AND } 110111 \text{ AND } 101111 = 100100$



Απαντήσεις στο ερώτημα

- Antony and Cleopatra, Act III, Scene ii
 - *Agrippa* [Aside to DOMITIUS ENOBARBUS]: Why, Enobarbus,
 - When Antony found Julius **Caesar** dead,
 - He cried almost to roaring; and he wept
 - When at Philippi he found **Brutus** slain.
- Hamlet, Act III, Scene ii
 - *Lord Polonius*: I did enact Julius **Caesar** I was killed i' the
 - Capitol; **Brutus** killed me.



Πώς δουλεύει μια μηχανή αναζήτησης

- Για παράδειγμα η Google

http://inf-server.inf.uth.gr/courses/CE634/noexternalweb/How_Google_Search_Works.mp4