



# Ανάκτηση πληροφορίας

## Ενότητα 2: Μέτρηση Αποτελεσματικότητας Συστημάτων

Απόστολος Παπαδόπουλος  
Τμήμα Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

# Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.





ΑΡΙΣΤΟΤΕΛΕΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΑΝΟΙΚΤΑ  
ΑΚΑΔΗΜΑΪΚΑ  
ΜΑΘΗΜΑΤΑ



# Αποτίμηση Αποτελεσματικότητας



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ  
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ  
*επένδυση στην κοινωνία της γνώσης*

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ  
2007-2013  
πρόγραμμα για την ανάπτυξη  
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

# Περιεχόμενα ενότητας

---

## 1. Αποτίμηση Αποτελεσματικότητας





ΑΡΙΣΤΟΤΕΛΕΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΕΣΣΑΛΟΝΙΚΗΣ

---

# Αποτίμηση Αποτελεσματικότητας

# Μέτρα Απόδοσης

- **Precision** =  $\frac{\# \text{ σχετικών κειμένων που επιστρέφονται}}{\# \text{ κειμένων που επιστρέφονται}}$
- **Recall** =  $\frac{\# \text{ σχετικών κειμένων που επιστρέφονται}}{\# \text{ συνολικών σχετικών κειμένων}}$
- Ο στόχος είναι να μεγιστοποιηθούν και τα δύο.
- Συνήθως, η σχετικότητα (Relevance) ενός κειμένου ως προς κάποιο ερώτημα είναι κάτι το υποκειμενικό.



# Εισαγωγή-1

- **Μέτρηση Απόδοσης**

- Πόσο καλό είναι ένα σύστημα.
- Συγκρίσεις μεταξύ συστημάτων.

- **Τι πρέπει να μετρήσουμε**

- Πόσο είναι ικανοποιημένος ο χρήστης από το αποτέλεσμα.
- Κάλυψη πληροφορίας.
- Τρόπος παρουσίασης.
- Δυσκολία που υπάρχει.

αποδοτικότητα



- Αποδοτικό ως προς χρόνο και χώρο.
- Recall
  - τμήμα των σχετικών κειμένων που ανακτώνται.
- Precision
  - τμήμα των ανακτώμενων κειμένων που είναι σχετικά.



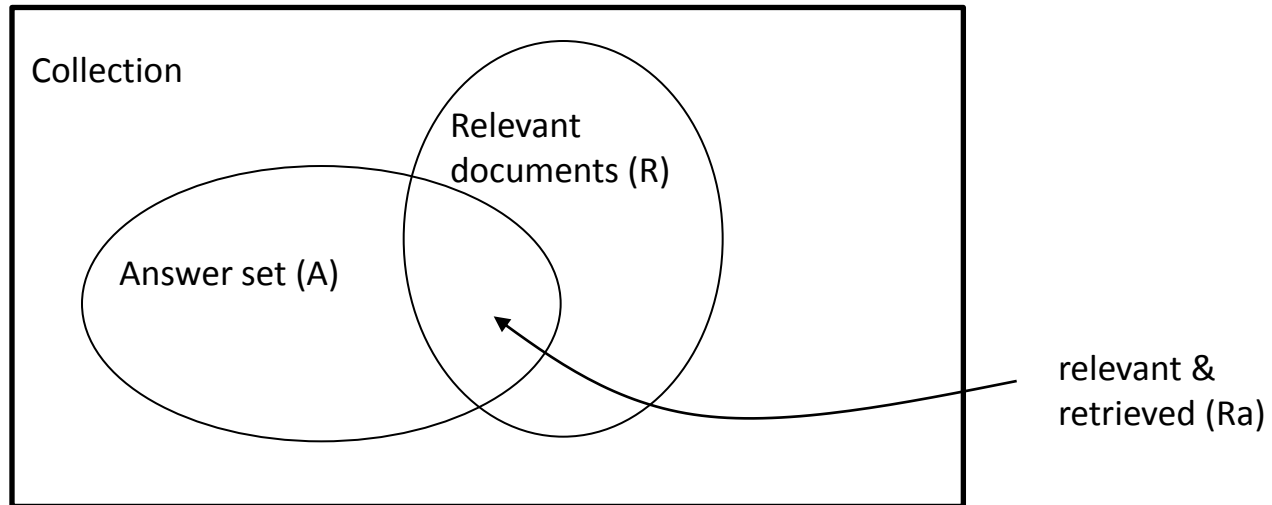


# Εισαγωγή-2

- **Πως μετράμε την απόδοση?**
  - Test reference collections:
    - TIPSTER/TREC
    - CACM
    - CISI
    - Cystic Fibrosis

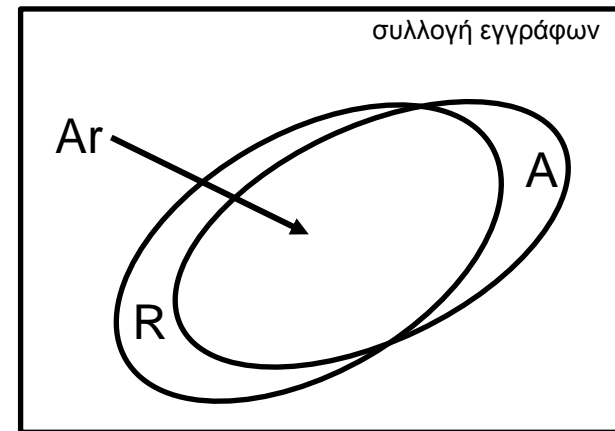
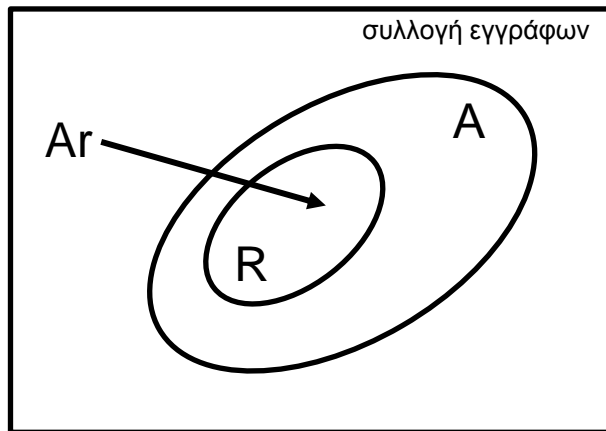
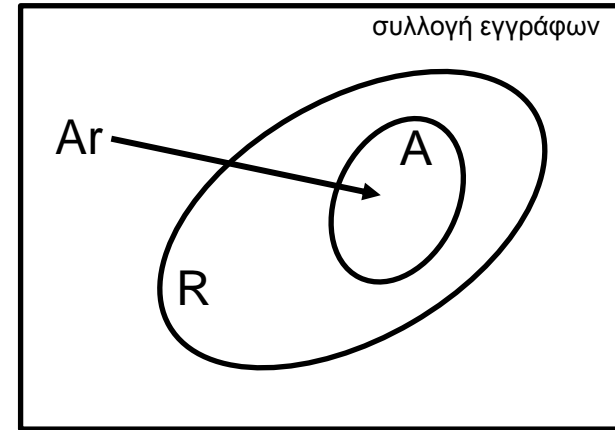
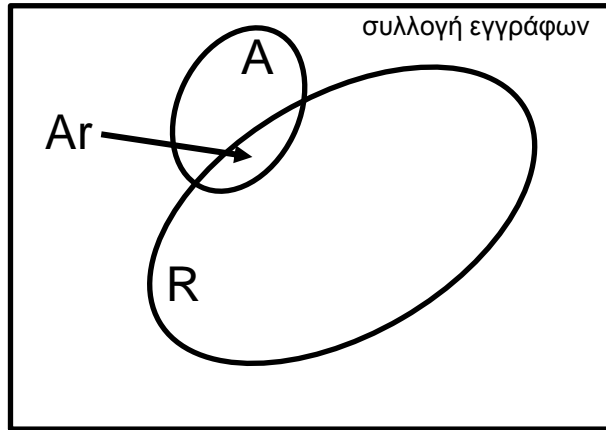


# Ανάκληση και Ακρίβεια



- $Recall = |Ra| / |R|$
- $Precision = |Ra| / |A|$

# Περιπτώσεις Ανάκλησης-Ακρίβειας



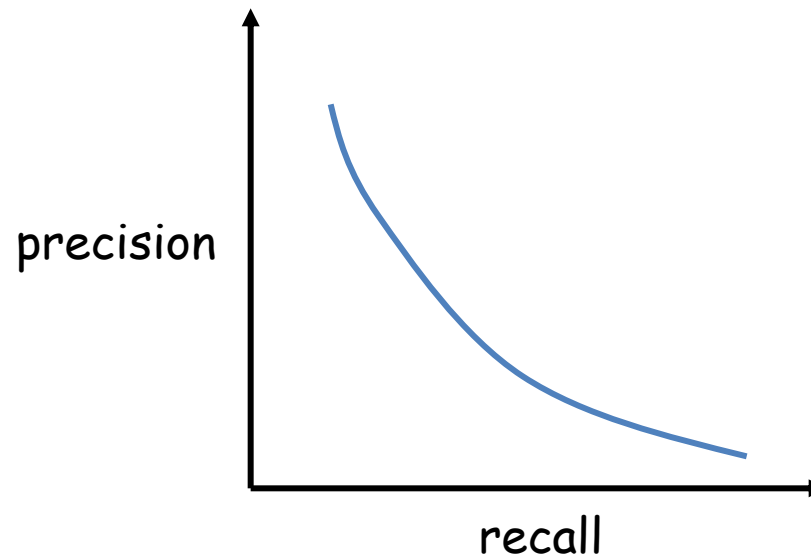
# Σχετικότητα Εγγράφων

- Το σύνολο  $R$  δεν είναι γνωστό.
- Relevance:
  - Υποκειμενική.
  - Μπορεί να μετρηθεί μέχρι κάποιο βαθμό.
- Πως ένα κείμενο χαρακτηρίζεται ως σχετικό προς το ερώτημα;
  - Σαφής απάντηση σε σαφές ερώτημα.
  - Partial Matching.
  - Προτείνεται πηγή για περισσότερες πληροφορίες.
  - Πληροφορίες για το background.



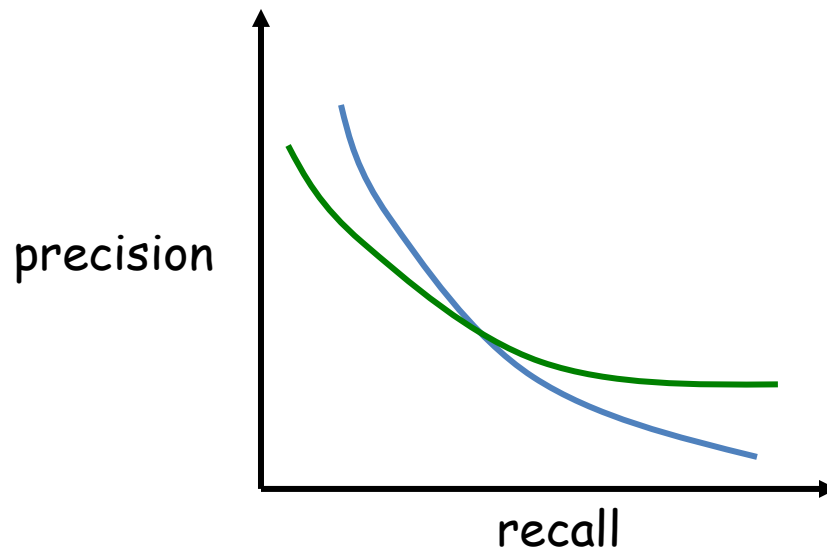
# Precision/ Recall Καμπύλη-1

- Precision - Recall trade-off.
- Μετρούμε το Precision σε διαφορετικά επίπεδα Recall.



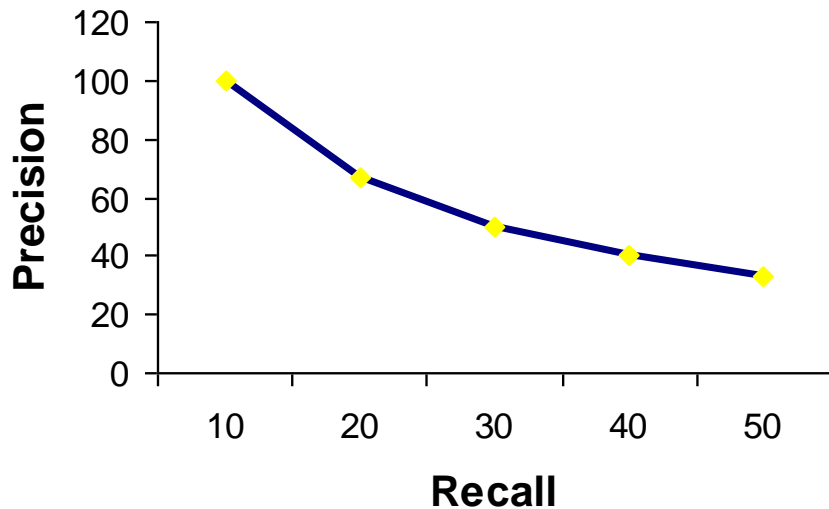
# Precision/ Recall Καμπύλη-2

- Είναι δύσκολο να διακρίνουμε ποια από τις παρακάτω καμπύλες είναι η καλύτερη:



# Παράδειγμα

$$R_q = \begin{cases} d_3, d_9, d_{25}, d_{39}, \\ d_{44}, d_{56}, d_{71}, d_{89}, d_{123} \end{cases}$$



Rank	Doc	Rel	R <sub>recall</sub>	P <sub>recision</sub>
0			0 %	0 %
1	d <sub>123</sub>	✓	10%	100%
2	d <sub>84</sub>		10%	50%
3	d <sub>56</sub>	✓	20%	67%
4	d <sub>6</sub>		20%	50%
5	d <sub>84</sub>		20%	40%
6	d <sub>9</sub>	✓	30%	50%
7	d <sub>511</sub>		30%	43%
8	d <sub>129</sub>		30%	38%
9	d <sub>187</sub>		30%	33%
10	d <sub>25</sub>	✓	40%	40%
11	d <sub>38</sub>		40%	36%
12	d <sub>48</sub>		40%	33%
13	d <sub>250</sub>		40%	31%
14	d <sub>113</sub>		40%	29%
15	d <sub>3</sub>	✓	50%	33%



# Μέση Ακρίβεια

- Μέση ακρίβεια για κάθε επίπεδο ανάκλισης:

$$\bar{P}(r) = \sum_{i=1}^{N_q} \frac{P_i(r)}{N_q}$$

- $\bar{P}(r)$  - μέση ακρίβεια για το επίπεδο ανάκλισης  $r$ .
- $N_q$  - πλήθος ερωτημάτων
- $P_i(r)$  - ακρίβεια στο επίπεδο ανάκλισης  $r$  για το  $i$ -οστό ερώτημα





# Επίπεδα Ανάκλησης

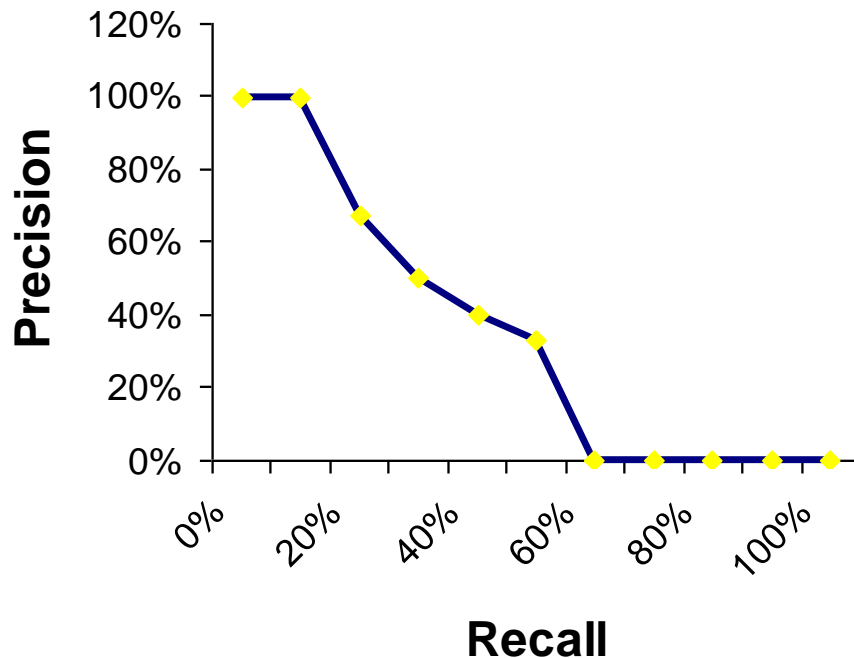
- Η καμπύλη δημιουργείται με 11 επίπεδα recall: 0%, 10%, 20%, ..., 100%
- Έστω  $r_j, j \in \{0, 1, 2, \dots, 10\}$  το  $j$ -ιοστό επίπεδο recall. Τότε,

$$P(r_j) = \max_{r_j \leq r \leq r_{j+1}} P(r)$$



# Παρεμβολή

- Εφαρμογή παρεμβολής για τα 11 επίπεδα ανάκλησης.



Level	Interpolated precision
0%	100%
10%	100%
20%	67%
30%	50%
40%	40%
50%	33%
60%	0%
70%	0%
80%	0%
90%	0%
100%	0%



# Κατώφλια

- Άλλος τρόπος υπολογισμού:
  - Σταθεροποιούμε τον αριθμό των κειμένων ανά επίπεδο:
    - top 5, top 10, top 20, top 50, top 100, top 500
  - Μέτρηση του precision για αυτά τα επίπεδα.
  - Μέση τιμή.
- Τρόπος να εστιάσουμε σε **high precision**.



# Μονότιμες Συνόψεις

- **Average Precision vs Recall** είναι χρήσιμες τιμές για την περιγραφή της απόδοσης ενός συστήματος σε ένα σύνολο ερωτημάτων.
- Πως θα μετρήσουμε την απόδοση **για κάθε ερώτημα χωριστά**;
- Single value measures:
  - Average Precision at Seen Relevant Documents;
  - R-Precision;
  - Precision Histograms;
  - Summary Table Statistics.



# Αρμονικός Μέσος

- The harmonic mean combines Recall & Precision into a single number ranging from 0 to 1:

$$F(j) = \frac{2}{\frac{1}{r(j)} + \frac{1}{P(j)}}$$

$P(j)$  - precision of  $j$ -th document in ranking;

$r(j)$  – recall of  $j$ -th document in ranking;

- If  $F(j) = 0$  – no relevant docs have been retrieved;
- If  $F(j) = 1$  – all ranked docs are relevant;
- The harmonic mean assumes **high** value only when both recall & precision are **high**.



# E-Measure

- Combine Precision and Recall into one number and allow user to specify whether s/he is more interested in recall or precision:

$$E(j) = 1 - \frac{1+b^2}{\frac{b^2}{r(j)} + \frac{1}{P(j)}}$$

$P(j)$  – precision of  $j$ -th document in ranking;

$r(j)$  – recall of  $j$ -th document in ranking;

$b$  - measure of relative importance of precision or recall:

$b > 1$  emphasizes precision,  $b < 1$  emphasizes recall

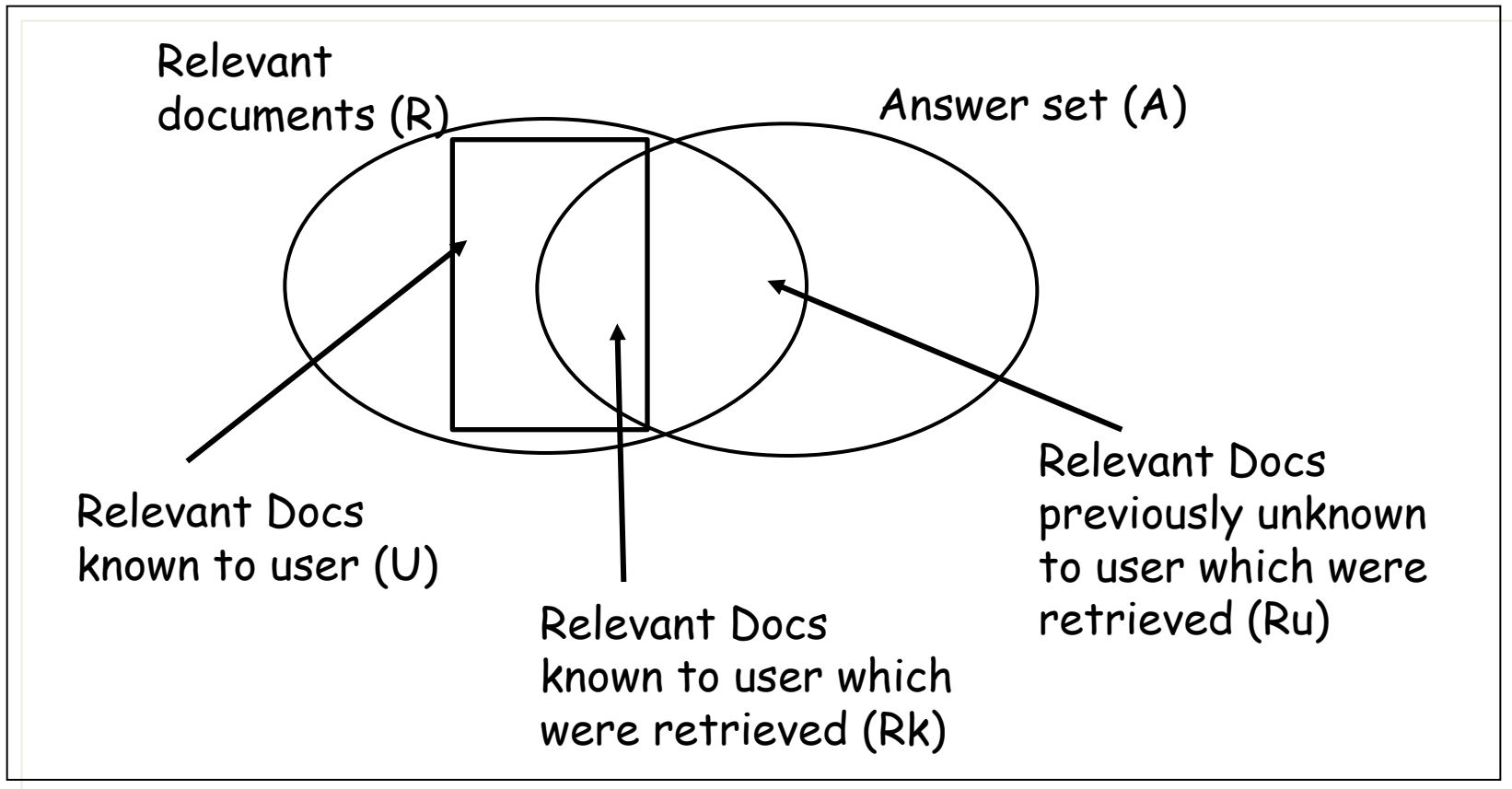


# Μέτρα Προσανατολισμένα στο Χρήστη

- Different users might have a different interpretation of which document is relevant or not
- User-oriented measures:
  - Coverage ratio:  $\text{coverage} = \frac{|Rk|}{|U|}$
  - Novelty ratio:  $\text{novelty} = \frac{|Ru|}{|Ru|+|Rk|}$
  - Relative ratio – the ratio between the # of relevant docs found and the # of relevant docs the user expected to find
  - Recall effort – the ratio between the # of relevant docs the user expected to find and the # of docs examined in an attempt to find the expected relevant docs



# Μέτρα Προσανατολισμένα στο Χρήστη





# Συλλογές Αναφοράς

- **TREC** collection (Text REtrieval Conference) – 5.8 GB, >1.5 Million Docs.
- **CACM** - computer science – 3024 articles.
- **ISI (CISI)** – library science – 1460 docs.
- **Cystic Fibrosis (CF)** - medicine – 1239 docs.
- **CRAN** – aeronautics – 1400 docs.
- **Time** – general articles – 423 docs.
- **NPL** – electrical engineering – 11429 docs.
- κλπ



# Σύνοψη

- Βασικά μέτρα αποτίμησης της αποτελεσματικότητας ενός IR συστήματος:
  - Recall & Precision.
  - Μονότιμες συνόψεις.
  - Εναλλακτικά μέτρα.
  - Μέτρα προσανατολισμένα στο χρήστη.
- Συλλογές αναφοράς.



# Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Απόστολος Παπαδόπουλος. «Ανάκτηση πληροφορίας. Αποτίμηση Αποτελεσματικότητας». Έκδοση: 1.0. Θεσσαλονίκη 2014. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://eclass.auth.gr/courses/OCRS388/>



# Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά - Μη Εμπορική Χρήση - Όχι Παράγωγα Έργα 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>





# Τέλος ενότητας

Επεξεργασία: <Μαυρίδης Απόστολος>  
Θεσσαλονίκη, <Εαρινό εξάμηνο 2013-2014>



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ  
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ  
*επένδυση στην κοινωνία της γνώσης*  
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ  
2007-2013  
πρόγραμμα για την ανάπτυξη  
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



ΑΡΙΣΤΟΤΕΛΕΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΕΣΣΑΛΟΝΙΚΗΣ

---

# Σημειώματα

# Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

