

# Ένα γραφικό παραμύθι ν-γραμμάτων

Γιώργος Γιαννακόπουλος <sup>1,2</sup>

<sup>1</sup>Ε.Κ.Ε.Φ.Ε. Δημόκριτος, ΙΠΤ  
Εργαστήριο Μηχανικής Γνώσης και Λογισμικού  
ggianna@iit.demokritos.gr

<sup>2</sup>SciFY A.M.K.E.

Μάιος 2012

# Σκοπός

Να δούμε:

- (επιφανειακά) ενδιαφέροντες τομείς της Τεχνητής Νοημοσύνης

## Σκοπός

Να δούμε:

- (επιφανειακά) ενδιαφέροντες τομείς της Τεχνητής Νοημοσύνης
- πώς μπορεί η τεχνητή νοημοσύνη να βοηθήσει τη ζωή μας

## Σκοπός

Να δούμε:

- (επιφανειακά) ενδιαφέροντες τομείς της Τεχνητής Νοημοσύνης
- πώς μπορεί η τεχνητή νοημοσύνη να βοηθήσει τη ζωή μας
- πώς είναι ο δρόμος ενός ερευνητή (όπως τον έζησα)

## Στο λεωφορείο

Πώς μπορούμε να φτιάξουμε έξυπνη αναζήτηση;  
Μέχρι τότε

- Λέξεις κλειδιά: “πληροφορική”, “Δημόκριτος”
- Λογικές πράξεις: “έξυπνη ΚΑΙ όμορφη”, “έξυπνη Ή όμορφη”

Προβλήματα με ορθογραφία, συνώνυμες λέξεις

## Τα λάθη και ο εγκέφαλος

- Διαβάστε γρήγορα το ακόλουθο (εθελοντές;):

## Τα λάθη και ο εγκέφαλος

- Διαβάστε γρήγορα το ακόλουθο (εθελοντές;):
- Μπορείτε να με διαβσάετε;

## Τα λάθη και ο εγκέφαλος

- Διαβάστε γρήγορα το ακόλουθο (εθελοντές;):
- Μπορείτε να με διαβσάετε;
- Ναι, μοπρετίε.



## Τα λάθη και ο εγκέφαλος

- Διαβάστε γρήγορα το ακόλουθο (εθελοντές;):
- Μπορείτε να με διαβσάετε;
- Ναι, μοπρετίε.
- Πραά οτ γεγνοός οίτ ενίαι ακατανεμένα τα γάρματα;

Εντυπωσιάστηκα!

## Πρώτες σκέψεις

- Είμαστε ανεκτικοί στο θόρυβο.

## Πρώτες σκέψεις

- Είμαστε ανεκτικοί στο θόρυβο.
- Η ακριβής ακολουθία γραμμάτων είναι πολύ *χρήσιμη*, αλλά *όχι απόλυτα απαραίτητη*.

## Πρώτες σκέψεις

- Είμαστε ανεκτικοί στο θόρυβο.
- Η ακριβής ακολουθία γραμμάτων είναι πολύ *χρήσιμη*, αλλά *όχι απόλυτα απαραίτητη*.
- Η γειτονιά (γειτνίαση) παίζει ρόλο.



# Τα ν-γράμματα

N-γράμμα: Μερικά (N) γράμματα σε σειρά

## Παράδειγμα

Ίδρυμα

2-γράμματα: Ίδ, δρ, ρυ, υμ, μα

3-γράμματα: Ίδρ, δρυ, δυμ, υμα

## Γιατί όχι γειτονιές ν-γραμμάτων;

### Παράδειγμα

Κείμενο:

abcde

## Γιατί όχι γειτονιές ν-γραμμάτων;

### Παράδειγμα

Κείμενο:	abcde
3-γράμματα:	abc, bcd, cde

## Γιατί όχι γειτονιές ν-γραμμάτων;

### Παράδειγμα

Κείμενο:	abcde
3-γράμματα:	abc, bcd, cde
Ακμές (γειτονιά):	abc-bcd, bcd-cde



## Γιατί όχι γειτονιές ν-γραμμάτων;

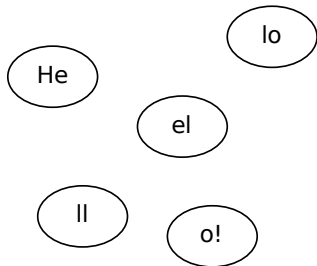
### Παράδειγμα

Κείμενο:	abcde
3-γράμματα:	abc, bcd, cde
Ακμές (γειτονιά):	abc-bcd, bcd-cde
Συνεμφανίσεις:	abc-bcd (1.0) , bcd-cde (1.0)

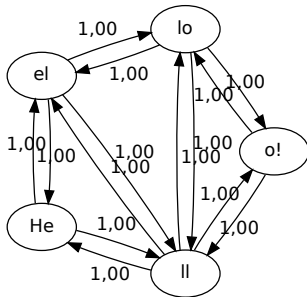
# Ο γράφος ν-γραμμάτων

*Hello!*

# Ο γράφος ν-γραμμάτων



# Ο γράφος ν-γραμμάτων



## Γράφος ν-γραμμάτων και μακαρονάδα

- Ένα έγγραφο γίνεται **γράφος ν-γραμμάτων**



## Γράφος ν-γραμμάτων και μακαρονάδα

- Ένα έγγραφο γίνεται **γράφος ν-γραμμάτων**



- Πολλά έγγραφα;



## Γράφος ν-γραμμάτων και μακαρονάδα

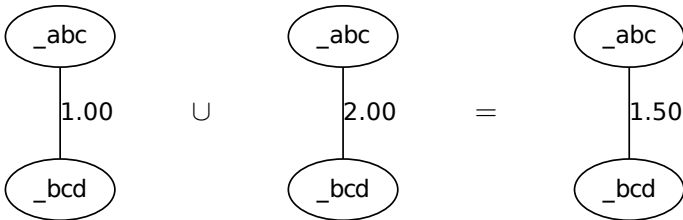
- Ένα έγγραφο γίνεται **γράφος ν-γραμμάτων**



- Πολλά έγγραφα;
- Ένας “μέσος”, πλούσιος γράφος (μοντέλο)



## Ένωση γράφων



$$\frac{1 + 2}{2} = 1.5$$

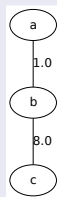


## Και πώς συγκρίνουμε δύο γράφους;

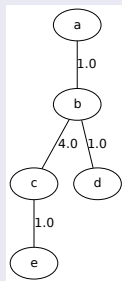
- Βλέπουμε αν έχουν ίδιο αριθμό ακμών (μέγεθος)
- Βλέπουμε αν έχουν ίδιες ακμές (συνύπαρξη)
- Βλέπουμε αν οι ακμές έχουν ίδιο βάρος (τιμή)

## Ομοιότητα μεγέθους (Size Similarity)

### Παράδειγμα



$$|G_1| = 2$$

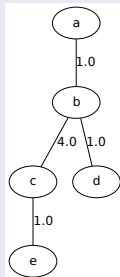
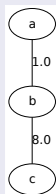


$$|G_2| = 4$$

Αποτέλεσμα:  $\frac{2}{4} = 0.5$

## Ομοιότητα συνύπαρξης (Co-occurrence Sim.)

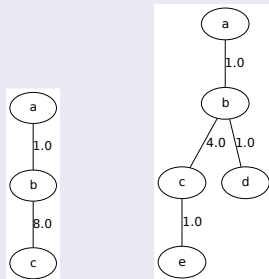
### Παράδειγμα



Αποτέλεσμα:  $\frac{1}{4} + \frac{1}{4} = 0.5$

## Ομοιότητα τιμής (Value Sim.)

### Παράδειγμα



Αποτέλεσμα:  $\frac{1.0}{\frac{1.0}{4}} + \frac{4.0}{\frac{8.0}{4}} = \frac{1}{4} + \frac{1}{8} = 0.375$

## Από τη μακαρονάδα στην ταξινόμηση κειμένων

- Αν ένα e-mail είναι κείμενο
- ...τότε μπορεί να γίνει γράφος ν-γραμμάτων
- ...και μπορούμε να αναπαραστήσουμε ομάδες e-mails (π.χ., ανά θέμα)
- ...και όταν έρχεται ένα νέο κείμενο να βλέπουμε με τι μοιάζει περισσότερο!

Σημείωση: Αυτό συνδέεται με τη *Μηχανική Μάθηση*

## Πίσω στην ταξινόμηση...

- Ενοχλητική αλληλογραφία: **98% επιτυχία**

## Πίσω στην ταξινόμηση...

- Ενοχλητική αλληλογραφία: **98% επιτυχία**
- Θεματική ταξινόμηση: **71% - 95%** επιτυχία (Twitter και Reuters)
- Δουλεύει σε πολλές γλώσσες (Ελληνικά, Αγγλικά, Γαλλικά, Γερμανικά...)

## Πίσω στην ταξινόμηση...

- Ενοχλητική αλληλογραφία: **98% επιτυχία**
- Θεματική ταξινόμηση: **71% - 95%** επιτυχία (Twitter και Reuters)
- Δουλεύει σε πολλές γλώσσες (Ελληνικά, Αγγλικά, Γαλλικά, Γερμανικά...)

Ήταν σαφές: κάτι ενδιαφέρον γεννήθηκε...



## Εξαγωγή περιλήψεων από πολλαπλά κείμενα

### Ποια περίληψη είναι καλή;

- Αξιολόγηση περιλήψεων
- Έχουμε υπολογιστές ή/και ανθρώπους που βγάζουν περιλήψεις.
- Αλλά τις βγάζουν καλά; Πώς το κρίνουμε;
- Δεν μπορούμε να ρωτάμε συνέχεια ανθρώπους.

## Πίσω στη μακαρονάδα...

- Αν έχουμε 3 πολύ καλές μακαρονάδες



- ...μπορούμε να θεωρήσουμε καλή όποια μακαρονάδα τους μοιάζει
- ...και κακές αυτές που δεν τους μοιάζουν.

Γιατί όχι και στις περιλήψεις, αφού **μπορούν να γίνουν γράφοι ν-γραμμάτων**;

## ...και ξανά στις περιλήψεις...

- Ζητάμε 3 περιλήψεις από ανθρώπους για το ίδιο θέμα

## ...και ξανά στις περιλήψεις...

- Ζητάμε 3 περιλήψεις από ανθρώπους για το ίδιο θέμα
- Κάνουμε κάθε περίληψη-μοντέλο (από άνθρωπο) γράφο ν-γραμμάτων

## ...και ξανά στις περιλήψεις...

- Ζητάμε 3 περιλήψεις από ανθρώπους για το ίδιο θέμα
- Κάνουμε κάθε περίληψη-μοντέλο (από άνθρωπο) γράφο ν-γραμμάτων
- Όταν θέλουμε να κρίνουμε μία νέα περίληψη

## ...και ξανά στις περιλήψεις...

- Ζητάμε 3 περιλήψεις από ανθρώπους για το ίδιο θέμα
- Κάνουμε κάθε περίληψη-μοντέλο (από άνθρωπο) γράφο ν-γραμμάτων
- Όταν θέλουμε να κρίνουμε μία νέα περίληψη
  - Φτιάχνουμε το γράφο της

## ...και ξανά στις περιλήψεις...

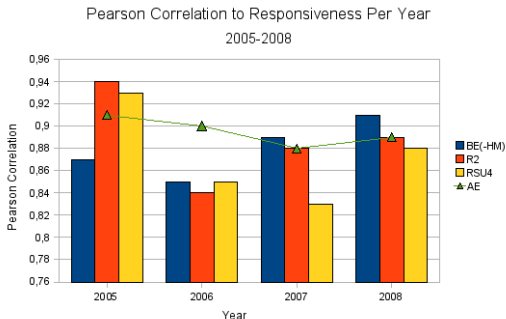
- Ζητάμε 3 περιλήψεις από ανθρώπους για το ίδιο θέμα
- Κάνουμε κάθε περίληψη-μοντέλο (από άνθρωπο) γράφο  $n$ -γραμμάτων
- Όταν θέλουμε να κρίνουμε μία νέα περίληψη
  - Φτιάχνουμε το γράφο της
  - Παίρνουμε την ομοιότητά του με κάθε ένα από τα μοντέλα  
(π.χ. 0.5, 0.6, 0.7)

## ...και ξανά στις περιλήψεις...

- Ζητάμε 3 περιλήψεις από ανθρώπους για το ίδιο θέμα
- Κάνουμε κάθε περίληψη-μοντέλο (από άνθρωπο) γράφο ν-γραμμάτων
- Όταν θέλουμε να κρίνουμε μία νέα περίληψη
  - Φτιάχνουμε το γράφο της
  - Παίρνουμε την ομοιότητά του με κάθε ένα από τα μοντέλα (π.χ. 0.5, 0.6, 0.7)
  - Παίρνουμε το μέσο όρο (π.χ. τελικός βαθμός: 0.6)

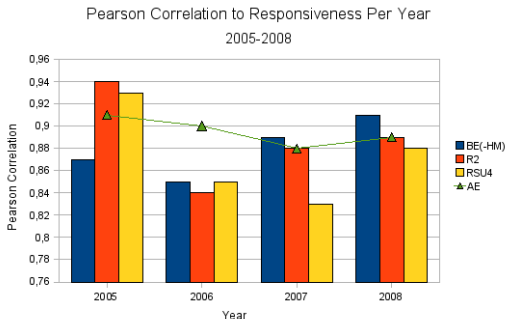


## Και δούλεψε!



**Σχήμα:** Συσχέτιση μεταξύ αυτόματης και ανθρώπινης αξιολόγησης σε συστήματα εξαγωγής περιλήψεων

## Και δούλεψε!



**Σχήμα:** Συσχέτιση μεταξύ αυτόματης και ανθρώπινης αξιολόγησης σε συστήματα εξαγωγής περιλήψεων

...αλλά κανείς δεν το πίστευε...

## Και τι άλλο έγινε;

- Βγάλαμε περιλήψεις
- Δουλέψαμε με βίντεο: μία συμπεριφορά είναι μία γειτονιά από κινήσεις  
...και άλλα πολλά.

# Όλα καλά;



# Ανάλυση συναισθήματος

- Το κείμενο εκφράζει συναίσθημα
- ...που είναι δύσκολο να το ταξινομήσεις
- ...αν δεν είσαι άνθρωπος
- Πάμπολλες πηγές (ιστολόγια, κοινωνικά δίκτυα)

Πώς γίνεται αυτόματα;

## Από την πρόταση στην ακολουθία

Για κείμενα που μας είναι γνωστό αν είναι θετικά ή αρνητικά:

- Από τις λέξεις τους  
(π.χ. *“Θα ήταν όμορφη μέρα, αν δεν πεινούσαμε...”*)

## Από την πρόταση στην ακολουθία

Για κείμενα που μας είναι γνωστό αν είναι θετικά ή αρνητικά:

- Από τις λέξεις τους  
(π.χ. *“Θα ήταν όμορφη μέρα, αν δεν πεινούσαμε...”*)
- φτάνουμε στις έννοιες  
(*“είναι”, “όμορφος”, “μέρα”, “πεινώ”*)

## Από την πρόταση στην ακολουθία

Για κείμενα που μας είναι γνωστό αν είναι θετικά ή αρνητικά:

- Από τις λέξεις τους  
(π.χ. *“Θα ήταν όμορφη μέρα, αν δεν πεινούσαμε...”*)
- φτάνουμε στις έννοιες  
(*“είναι”, “όμορφος”, “μέρα”, “πεινώ”*)
- κι από εκεί σε ακολουθία θετικού/αρνητικού  
(*ουδέτερο, θετικό, ουδέτερο, αρνητικό*)



## Από την πρόταση στην ακολουθία

Για κείμενα που μας είναι γνωστό αν είναι θετικά ή αρνητικά:

- Από τις λέξεις τους  
(π.χ. “Θα ήταν όμορφη μέρα, αν δεν πεινούσαμε...”)
- φτάνουμε στις έννοιες  
(“είναι”, “όμορφος”, “μέρα”, “πεινώ”)
- κι από εκεί σε ακολουθία θετικού/αρνητικού  
(ουδέτερο, θετικό, ουδέτερο, αρνητικό)
- και βάζουμε το σύστημα να μάθει τη σύνδεση  
(ουδέτερο, θετικό, ουδέτερο, αρνητικό =>  
αρνητικό κείμενο)

## Και πώς πηγαι;

- Σε τίτλους εφημερίδων

## Και πώς πηγαι;

- Σε τίτλους εφημερίδων
- Αν επιλέγαμε τυχαία θα είχαμε: 50%

## Και πώς πηγαι;

- Σε τίτλους εφημερίδων
- Αν επιλέγαμε τυχαία θα είχαμε: 50%
- Μεταφορές **~72% επιτυχία**  
“Αποκεφαλισμός Αλ Κάιντα στο Ιράκ”

## Και πώς πηγαι;

- Σε τίτλους εφημερίδων
- Αν επιλέγαμε τυχαία θα είχαμε: 50%
- Μεταφορές **~72% επιτυχία**  
“Αποκεφαλισμός Αλ Κάιντα στο Ιράκ”
- Επεκταμένες έννοιες **~68% επιτυχία**  
“Γυναίκα μάχεται να κρατήσει μεθυσμένο οδηγό στη φυλακή”

## Πώς αξιοποιείται η ανάλυση συναισθήματος;

### Opengov.gr

- Άνθρωποι συζητούν και εκφράζονται για νόμους
- Κάποιοι χρειάζεται να αναλύσουν τις πολλές απαντήσεις.
- Μπορούμε να το κάνουμε αυτόματα!

# Αξιοποιώντας απόψεις

Άρθρο 1 Ορισμό

Βόλου Τόλη

Επικίνδυνο ένα σύνταγμα είναι το ένα σύνταγμα που εκδηλώνει απόλυτη ένταση επιθετική συμπεριφορά προς τον άνθρωπο... Η περίπτωση αυτή αφορά κατά βάση το σχολείο, προβάλλεται με ποικίλο τρόπο να αποδοκιμασθεί η επιθετική συμπεριφορά. Θα πρέπει να περιληφθεί ως ασφαλιστική διαίρεση ότι η επιθετική αυτή συμπεριφορά θα αποδοκιμασθεί με βιβλική νουανσέρ. Η πρακτική μάχη σήμερα μας έχει δείξει ότι πολλές καταγγελίες για όλην διάρκεια στα σχολεία δεν ανταποκρίνονται στην πραγματικότητα. Διαφορετικά ο κίνδυνος που έχει προκύψει με κάποιο στυλ θα επικαλούτο επιθετική συμπεριφορά.

Περσσότερες πληροφορίες

Σχετικές ομάδες

Αρνητική άποψη

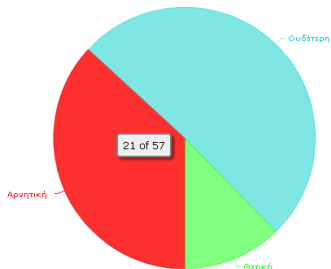
ΑΔΕΛΦΟΤΑ ΖΩΩΝ	ΖΩΑ ΖΩΟ	ΑΓΑΠΗΝΗ ΔΕΣΜΕΥΕΤΑΙ ΕΠΕΝΔΥΣΗ ΕΦΑΡΜΟΣΩΝ	ΟΙ ΑΠΟΔΕΙΚΝΥΤΑΙ ΔΙΑΦΟΡΕΤΙΚΑ ΕΠΙΘΕΤΙΚΑ	CETY ΕΠΙΚΙΝΔΥΝΟ ΕΣΚΥΦΕ	ΒΕΒΑΙΩΣΗ
------------------	------------	---	---	---------------------------	----------

Εύκολο Γινώ

Όταν αφορά τον χαρακτηρισμό "επιθετικό ζώο" να προστεθεί στο απόλυτο και η φράση "και απολαυσιζόμενης επιθετική συμπεριφορά, η οποία θα αποδοκιμασθεί με περσσότερες της μιας μαθητικές και με γινώριση σκόλο που θα εξετάσει το ζώο".

Περσσότερες πληροφορίες

Αριθμός θετικών, αρνητικών και ουδέτερων απόψεων στο άρθρο.



<http://gov.insight.iit.demokritos.gr>  
**Γιατί; Ο λόγος μας αποκτά αξία**

## Προσαρμοστικά συστήματα

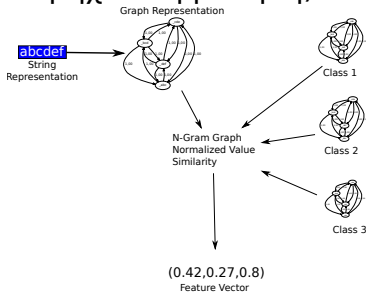
- Ένα σύστημα που μαθαίνει τι σου αρέσει (π.χ. ποιες ειδήσεις σε ενδιαφέρουν)
- Σου προτείνει κάθε μέρα ειδήσεις
- Και όσο το διορθώνεις (feedback) βελτιώνεται



## Πώς γίνεται;

Με γράφους  $n$ -γραμμάτων (και μηχανική μάθηση)!

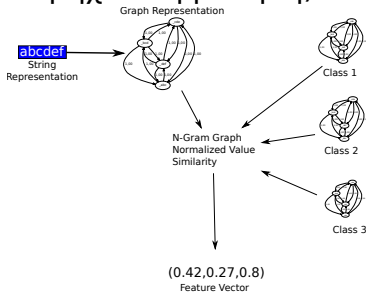
- Γράφοι για πολύ σημαντικές, ενδιαφέρουσες, ή αδιάφορες ειδήσεις.



## Πώς γίνεται;

Με γράφους  $n$ -γραμμάτων (και μηχανική μάθηση)!

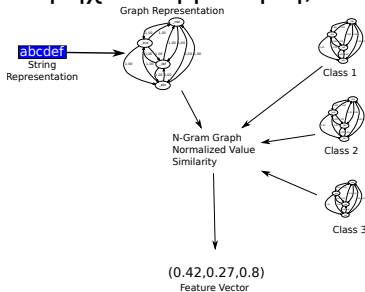
- Γράφοι για πολύ σημαντικές, ενδιαφέρουσες, ή αδιάφορες ειδήσεις.
- Μετράμε ομοιότητες με τους γράφους των κατηγοριών.



## Πώς γίνεται;

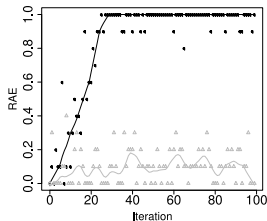
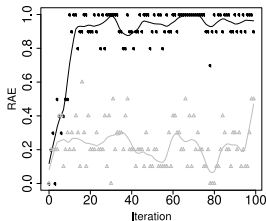
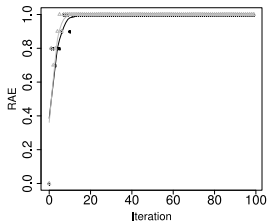
Με γράφους  $n$ -γραμμάτων (και μηχανική μάθηση)!

- Γράφοι για πολύ σημαντικές, ενδιαφέρουσες, ή αδιάφορες ειδήσεις.
- Μετράμε ομοιότητες με τους γράφους των κατηγοριών.
- Παίρνουμε την απόφαση από όλες τις ομοιότητες μαζί.



## Τελικά μαθαίνει το σύστημα;

Τρεις χρήστες. Από τον πιο εύκολο, στον πιο δύσκολο.  
Γκρι: χωρίς γράφους. Μαύρο: εμείς!



Σχήμα: Οριζόντια: Χρόνος μάθησης, Κατακόρυφα: Επίδοση

# Από υπολογιστές σε ανθρώπους

Από το κείμενο στα γονίδια

- Το γονιδίωμα αποτελείται από σύμβολα-βάσεις (A, T, G, C)

## Από υπολογιστές σε ανθρώπους

Από το κείμενο στα γονίδια

- Το γονιδίωμα αποτελείται από σύμβολα-βάσεις (A, T, G, C)
- ...άρα είναι σαν κείμενο (εντάξει, λίγο χαζό: AAACCATTC A)

# Από υπολογιστές σε ανθρώπους

Από το κείμενο στα γονίδια

- Το γονιδίωμα αποτελείται από σύμβολα-βάσεις (A, T, G, C)
- ...άρα είναι σαν κείμενο (εντάξει, λίγο χαζό: AAACCATTC A)
- ...άρα, **μπορεί να γίνει...**



## Από υπολογιστές σε ανθρώπους

Από το κείμενο στα γονίδια

- Το γονιδίωμα αποτελείται από σύμβολα-βάσεις (A, T, G, C)
- ...άρα είναι σαν κείμενο (εντάξει, λίγο χαζό: AAACCATTC A)
- ...άρα, **μπορεί να γίνει...**  
**Σωστά! Γράφος ν-γραμμάτων!**



# Υπερσυντηρημένες μη-κωδικές ακολουθίες (CNEs)

## CNEs

Ακολουθίες που υπάρχουν για χροοοοοοοονια, αλλά δεν ξέρουμε γιατί.

Παραδοσιακά δεν είναι εύκολο να διαχωριστούν  
(αναλογία βάσεων).

## Διαχωρίζονται με χρήση των γράφων;

- 490 ακολουθίες CNE από το ανθρώπινο γονιδίωμα
- 490 φυσικές αλληλουχίες ίσου μήκους και ανάλογης σύστασης
- 490 τυχαίες αλληλουχίες ίσου μήκους και ανάλογης σύστασης

## Διαχωρίζονται με χρήση των γράφων;

- 490 ακολουθίες CNE από το ανθρώπινο γονιδίωμα
- 490 φυσικές αλληλουχίες ίσου μήκους και ανάλογης σύστασης
- 490 τυχαίες αλληλουχίες ίσου μήκους και ανάλογης σύστασης

**Και ~80% κατηγοριοποιήθηκαν σωστά!**

Περίπου 18% καλύτερα από ότι με απλή αναλογία βάσεων.

## Επιστημογνωσία (Εν εξελίξει...)

### Από μαθητής, βοηθός

- Συμβουλευτική υποστήριξη
- Έμπρακτη υποστήριξη

### Σε ποιούς;

Φοιτητές και νέους ερευνητές

### Γιατί;

- Απερίσπαστη ακαδημαϊκή δημιουργία
- Ενίσχυση της διεπιστημονικής συνεργασίας

## Ποιοι είμαστε

- Υποψήφιοι διδάκτορες κοντά στην παρουσίαση
- Μεταδιδακτορικοί ερευνητές — Νέοι καθηγητές
- Νέοι ελεύθεροι επαγγελματίες και επιχειρηματίες
- Μέχρι 10 χρόνια από την εισαγωγή μας στο χώρο
- Τομείς:
  - Βιολογία
  - Πληροφορική
  - Μικροηλεκτρονική
  - Ψυχολογία
  - Εμπόριο — Βιομηχανία

## SciFY — [www.scify.org](http://www.scify.org)

### Από πατέντες σε κοινή ωφέλεια

Η επιστήμη είναι ο αγώνας λίγων να κατακτήσουν ένα κομμάτι της γνώσης που ανήκει σε όλους.

Ονειρευόμαστε έναν κόσμο όπου:

- η τεχνολογία είναι διαθέσιμη σε όλους, δωρεάν.

## SciFY — [www.scify.org](http://www.scify.org)

### Από πατέντες σε κοινή ωφέλεια

Η επιστήμη είναι ο αγώνας λίγων να κατακτήσουν ένα κομμάτι της γνώσης που ανήκει σε όλους.

Ονειρευόμαστε έναν κόσμο όπου:

- η τεχνολογία είναι διαθέσιμη σε όλους, δωρεάν.
- η ανακάλυψη δε σχετίζεται με πατέντες.

## SciFY — [www.scify.org](http://www.scify.org)

### Από πατέντες σε κοινή ωφέλεια

Η επιστήμη είναι ο αγώνας λίγων να κατακτήσουν ένα κομμάτι της γνώσης που ανήκει σε όλους.

Ονειρευόμαστε έναν κόσμο όπου:

- η τεχνολογία είναι διαθέσιμη σε όλους, δωρεάν.
- η ανακάλυψη δε σχετίζεται με πατέντες.
- οι ευκαιρίες για ανάπτυξη και ευδοκίμηση είναι διαθέσιμες σε όλους.



## SciFY — [www.scify.org](http://www.scify.org)

### Από πατέντες σε κοινή ωφέλεια

Η επιστήμη είναι ο αγώνας λίγων να κατακτήσουν ένα κομμάτι της γνώσης που ανήκει σε όλους.

Ονειρευόμαστε έναν κόσμο όπου:

- η τεχνολογία είναι διαθέσιμη σε όλους, δωρεάν.
- η ανακάλυψη δε σχετίζεται με πατέντες.
- οι ευκαιρίες για ανάπτυξη και ευδοκίμηση είναι διαθέσιμες σε όλους.
- το πρόβλημα ενός ατόμου είναι ευκαιρία συνεργασίας και προσφοράς.

## Και τι κάνουμε;

- Ενίσχυση πολυφωνίας (Κωδικό όνομα: NewSum)
- Πρόσβαση σε πολιτιστική κληρονομιά (Κωδικό όνομα: Touring Machine)
- Προσβασιμότητα υπολογιστή σε ανθρώπους με κινητικά προβλήματα (Κωδικό όνομα: ThinkFreedom)

Συνεργασίες με ΕΚΕΦΕ Δημόκριτος, ΟΠΑ, CSRI,  
Πανεπιστήμιο Αιγαίου...

## Και τι κάνουμε;

- Ενίσχυση πολυφωνίας (Κωδικό όνομα: NewSum)
- Πρόσβαση σε πολιτιστική κληρονομιά (Κωδικό όνομα: Touring Machine)
- Προσβασιμότητα υπολογιστή σε ανθρώπους με κινητικά προβλήματα (Κωδικό όνομα: ThinkFreedom)

Συνεργασίες με ΕΚΕΦΕ Δημόκριτος, ΟΠΑ, CSRI,  
Πανεπιστήμιο Αιγαίου... **και εσάς;**

## 5 Διδάγματα

- 1 Τα όμορφα πράγματα ξεκινούν απλά. Είναι σημαντικό να σκέφτεσαι από μηδενική βάση.

## 5 Διδάγματα

- 1 Τα όμορφα πράγματα ξεκινούν απλά. Είναι σημαντικό να σκέφτεσαι από μηδενική βάση.
- 2 Το διδακτορικό είναι δρόμος. Ανηφορικός. Αλλά είναι απλώς ένα κομμάτι της ζωής.

## 5 Διδάγματα

- 1 Τα όμορφα πράγματα ξεκινούν απλά. Είναι σημαντικό να σκέφτεσαι από μηδενική βάση.
- 2 Το διδακτορικό είναι δρόμος. Ανηφορικός. Αλλά είναι απλώς ένα κομμάτι της ζωής.
- 3 Πίστευε στον εαυτό σου, χωρίς έπαρση. Όλοι κάνουν λάθη. Ποτέ μην παραιτείσαι. Κάθε βήμα, σωστό ή λάθος, σε έφερε στο σήμερα.

## 5 Διδάγματα

- 1 Τα όμορφα πράγματα ξεκινούν απλά. Είναι σημαντικό να σκέφτεσαι από μηδενική βάση.
- 2 Το διδακτορικό είναι δρόμος. Ανηφορικός. Αλλά είναι απλώς ένα κομμάτι της ζωής.
- 3 Πίστευε στον εαυτό σου, χωρίς έπαρση. Όλοι κάνουν λάθη. Ποτέ μην παραιτείσαι. Κάθε βήμα, σωστό ή λάθος, σε έφερε στο σήμερα.
- 4 Οι συνεργάτες είναι καθοριστικοί. Παντού υπάρχουν αξιόλογοι συνεργάτες. Ψάξε τους.

## 5 Διδάγματα

- 1 Τα όμορφα πράγματα ξεκινούν απλά. Είναι σημαντικό να σκέφτεσαι από μηδενική βάση.
- 2 Το διδακτορικό είναι δρόμος. Ανηφορικός. Αλλά είναι απλώς ένα κομμάτι της ζωής.
- 3 Πίστευε στον εαυτό σου, χωρίς έπαρση. Όλοι κάνουν λάθη. Ποτέ μην παραιτείσαι. Κάθε βήμα, σωστό ή λάθος, σε έφερε στο σήμερα.
- 4 Οι συνεργάτες είναι καθοριστικοί. Παντού υπάρχουν αξιόλογοι συνεργάτες. Ψάξε τους.
- 5 Ταξίδεψε. Γνώρισε ανθρώπους και άφησε τις στατιστικές. Κοίτα ξανά τον τόπο σου γυρίζοντας.



## Και ζήσαν αυτοί καλά...

και εμείς μοιραζόμαστε, ονειρευόμαστε, πράττουμε,  
αγωνιζόμαστε, ευτυχούμε.

## Ένα γραφικό παραμύθι ν-γραμμάτων

Γιώργος Γιαννακόπουλος <sup>1,2</sup>

<sup>1</sup>Ε.Κ.Ε.Φ.Ε. Δημόκριτος, ΙΠΤ  
Εργαστήριο Μηχανικής Γνώσης και Λογισμικού  
ggianna@iit.demokritos.gr

<sup>2</sup>SciFY A.M.K.E.

Μάιος 2012

## N-gram Graph – Similarity 1

- $T_i$  maps a set of graphs  $\mathbb{G}^r$
- Size Similarity of  $G_1, G_2$ :  $SS = \frac{\min(|G_1|, |G_2|)}{\max(|G_1|, |G_2|)}$
- Containment Similarity: Each common edge adds  $\frac{1}{\min(|G_1|, |G_2|)}$  to a sum.
- Value Similarity: Using weights, every common edge adds

$$\frac{\min(w_e^i, w_e^j)}{\max(w_e^i, w_e^j)}$$

$$SS$$

- Normalized Value Similarity, SS factored out:  
 $NVS = \frac{VS}{SS}$
- Overall Similarity for  $n \in [L_{\min}, L_{\max}]$ : Weighted sum of rank similarity.

# N-Gram Graph Generic Operators

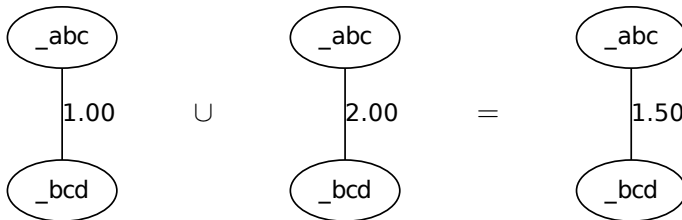
- Merging or Union  $\cup$  (Update  $U$ )
- Intersection  $\cap$
- Delta Operator (*All-Not-In* operator)  $\Delta$
- Inverse Intersection Operator  $\nabla$
- Similarity function  $\text{sim}$
- Degradation  $\searrow$

# Representing Sets of Graphs

## A representative graph for a set

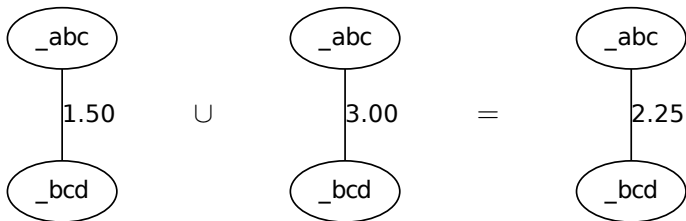
- is similar to functionality to the centroid of vectors
- **cannot** be represented using the **merging** operator (unless trivial)
- **can** be represented using the **update** operator
- for non-common edges the effect is **non-linear**

# Merge vs. Update 1



$$\frac{1 + 2}{2} = 1.5$$

## Merge vs. Update 2



$$\frac{1.5 + 3}{2} = 2.25$$

But what if we want  $\frac{1+2+3}{3} = 2?$

## Merge vs. Update 3

$$\text{updatedValue} = \text{oldValue} + l \times (\text{newValue} - \text{oldValue}) \quad (1)$$

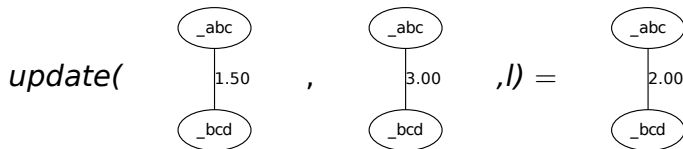
where  $0 \leq l \leq 1$  is the learning factor

### Representative (or *Centroid*) Graph

Use update operator with learning factor:  $\frac{1}{\text{instanceCount}}$ , where *instanceCount* is the number of instances that will be described by the graph *after* the update.

## Merge vs. Update 4

$$l = \frac{1}{3}$$

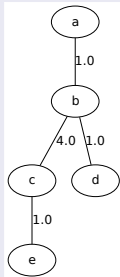
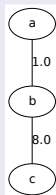


$$1.5 + \frac{1}{3} \times (3 - 1.5) = \frac{3}{2} + \frac{1}{3} \times \frac{3}{2} = \frac{4}{2} = 2$$



# Ομοιότητα τιμής (Value Sim.)

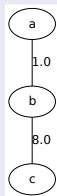
## Παράδειγμα



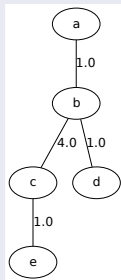
Αποτέλεσμα:  $\frac{1.0}{\frac{1.0}{4}} + \frac{4.0}{\frac{8.0}{4}} = \frac{1}{4} + \frac{1}{8} = 0.375$

# Ομοιότητα μεγέθους (Size Similarity)

## Παράδειγμα



$$|G_1| = 2$$



$$|G_2| = 4$$

Αποτέλεσμα:  $\frac{2}{4} = 0.5$