

# Εισαγωγή στην Επιστήμη του Ιστού

Ενότητα 10: Μεγάλα δεδομένα και Διαδίκτυο στις Κοινωνικές Επιστήμες

Διδάσκων: Νίκος Τσέλιος

Τμήμα Επιστημών της Εκπαίδευσης και της Αγωγής στην Προσχολική Ηλικία



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Πατρών, Σχολή Κοινωνικών και Ανθρωπιστικών Επιστημών, Τμήμα Επιστημών της Εκπαίδευσης και Αγωγής στην Προσχολική Ηλικία, Νικόλαος Τσέλιος, «Εισαγωγή στην Επιστήμη του Ιστού». Έκδοση: 1.0.

Πάτρα 2014. Διαθέσιμο από τη δικτυακή διεύθυνση:

<https://eclass.upatras.gr/courses/PN1427/>

# Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 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».



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

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

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

# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα εκτός κι αν αναφέρεται διαφορετικά
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Πατρών**» έχει χρηματοδοτήσει μόνο τη 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



# Σκοποί ενότητας

- Να αναγνωρίσει ζητήματα μελέτης του Παγκόσμιου Ιστού μέσα από το πλαίσιο των Κοινωνικών Επιστημών
- Να αναγνωρίσει εργαλεία έρευνας
- Να εξοικειωθεί με ζητήματα που αφορούν στη συγκρότηση κοινωνικών σχέσεων με τη χρήση του Διαδικτύου

# Περιεχόμενα ενότητας

- ◎ Περιγραφή: Πραγματεύονται ζητήματα ανάλυσης κοινωνικών παραγόντων από δεδομένα αλληλεπίδρασης στο Διαδίκτυο.
- ◎ Λέξεις Κλειδιά: Μεγάλα Δεδομένα, Διαδίκτυο, Συλλογή δεδομένων, Ανάλυση δεδομένων

# Πως μπορούμε να μελετήσουμε καλύτερα τα φαινόμενα αυτά;

- ◎ Συνδυασμό ποσοτικών και ποιοτικών προσεγγίσεων
  - **Big data**: μεγάλα δεδομένα από πλήθος αλληλεπιδράσεων σημαντικού αριθμού χρηστών. Το ίδιο το μέσο διευκολύνει τη συλλογή τους
  - **Thick data**: δεδομένα από ποιοτικές προσεγγίσεις (συνεντεύξεις, εθνογραφία)
  - Διερεύνηση/καταγραφή αλλά και επεξήγηση φαινομένων. Τα όρια είναι δυσδιάκριτα (Stebbins, 2001)
  - Άνθηση υπολογιστικών μεθόδων και 'εισβολή' ειδικών από πεδία όπως Βιολογία, Φυσική και Ιατρική (Barabasi et al., 2002; Palla, Barabasi, & Vicsek, 2007)
- ◎ Όροι όπως Ψηφιακός Ανθρωπισμός /**Digital Humanities** και Digital Culture Studies (Jenkins, Ford, & Green, 2018) περιγράφουν την ανάπτυξη των νέων πρακτικών
- ◎ Παραδόξως, η τεράστια ποσότητα δεδομένων προς ανάλυση, αναδεικνύει τη σημασία των ορθών ποιοτικών αναλύσεων (Babones, 2016)

# Παραδείγματα εφαρμογών Big data 1/4

- ⦿ Διαφοροποιήσεις στη θεματολογία της επικοινωνία, συνδέονται με τον εντοπισμό απώλειας θέσης εργασίας (Llorente, Garcia-Herranz, Cebrian, & Moro, 2015)
- ⦿ Ανάλυση 'συλλογικού θυμικού', συνήθειες κοινοτήτων κλπ από ανάλυση tweets (Golder & Macy, 2011)
- ⦿ Ανάλυση συναισθημάτων που προκάλεσε μια ταινία και αυτόματη αξιολόγησή της (Thet, Na, & Khoo, 2010)
- ⦿ Ποσοτικοποιημένος εαυτός (Swan, 2013): Καταγραφή και ανάλυση δεδομένων κινητικότητας κ.α.
  - Δυνατότητα αναγνώρισης διάθεσης
- ⦿ Internet of Things: Συσκευές με αδιάλειπτη σύνδεση στο Διαδίκτυο (σε αυτοκίνητα, καταγραφικές κάμερες, ακόμα και σε ψυγεία)



# Παραδείγματα εφαρμογών Big data 2/4

- ◎ Μέσα από την ανάλυση των Facebook like αναγνώριση χαρακτηριστικών όπως
  - sexual orientation, ethnicity, age, gender, religion, political views
  - psychological features : life satisfaction, intelligence, personality
  - tendency to use drugs, marital status of parents (Kosinski, Stillwell, & Graepel, 2013)
  - 88% ακρίβεια στην εκτίμηση σεξουαλικού προσανατολισμού, 95% στην αναγνώριση φυλετικής ομάδας
  - Δεδομένα από 58.000 που συμμετείχαν εθελοντικά με αντάλλαγμα δωρεάν ψυχομετρικά τεστ
  - Κάποια ευρήματα: Σειρά The Colbert Report σχετιζόταν με υψηλό δείκτη νοημοσύνης, προτίμηση σε μηχανές Harley Davidson με χαμηλό

# Παραδείγματα εφαρμογών Big data 3/4

- ⦿ Ανάλυση 86000 χρηστών από τους Youyou, Kosinski, and Stillwell (2015) έδειξε ότι:
  - Τα like είναι καλύτερος προβλεπτικός δείκτης από το δίκτυο των φίλων
  - Συχνά είναι καλύτερα και από αυτοαναφορικές κλίμακες (ερωτηματολόγια)
  - Demo στο [applymagicsauce.com](http://applymagicsauce.com)
- ⦿ Εξαιρετικές εφαρμογές στο marketing.
  - Πχ Andrew Pole (για την Target, αλυσίδα σουπερμάρκετ στις ΗΠΑ) έθεσε το ερώτημα: Μπορούμε να προβλέψουμε αν κάποια είναι σε 'ενδιαφέρουσα' ακόμα και αν δεν θέλει να το αποκαλύψει η ίδια;
  - Και αυτό γιατί συνιστούν μια εξαιρετικά κερδοφόρα ομάδα χρηστών / πελατών
  - Μέσα από την ανάλυση της διαφοροποίησης σε αγορές σε 25 προϊόντα μπορούσε να προβλέψει με μεγάλη ακρίβεια καθώς και την ημερομηνία γέννησης (Duhigg, 2012)
  - Πχ περίπου στην 20<sup>η</sup> εβδομάδα της κύησης πολλές άρχιζαν να συσσωρεύουν ασβέστιο, ψευδάργυρο και μαγνήσιο
  - Με τον τρόπο αυτό άρχισαν να στέλνουν εκπτωτικά κουπόνια για προϊόντα για μωρά.
  - Το 2010 ένας πατέρας διαμαρτυρήθηκε εντόνως για τα κουπόνια αυτά που κοινοποιούνταν στην κόρη του. Στην πορεία διαπιστώθηκε ότι ο αλγόριθμος είχε δίκιο και αυτός άδικο (Hill, 2012)

# Παραδείγματα εφαρμογών Big data 4/4

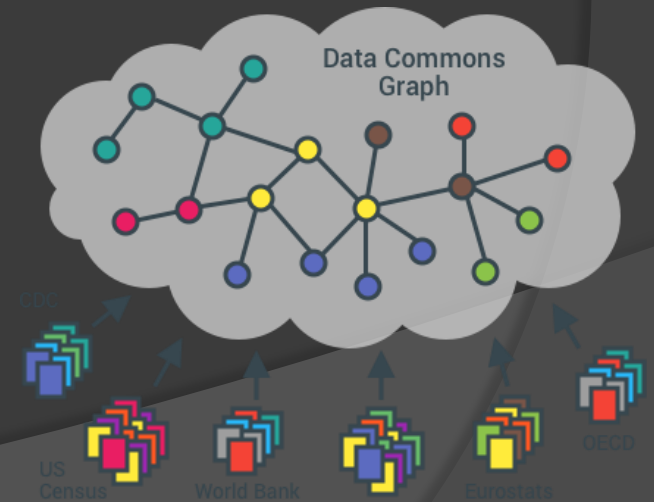
- ⦿ Από ανάλυση 35000 φωτογραφιών 91% ακρίβεια στην εκτίμηση του σεξουαλικού προσανατολισμού (Y. Wang & Kosinski, 2018)
- ⦿ Cambridge Analytica: behavioral microtargeting. Από δεδομένα συμπεριφοράς χρηστών στο FB + ψυχομετρικά τεστ
- ⦿ Στις εκλογές των ΗΠΑ αναγνωρίστηκαν 32 στυλ προσωπικότητας και προσαρμόστηκε το πολιτικό μήνυμα σε αυτές
  - Στις ΗΠΑ τα 2 μεγάλα κόμματα διατηρούν μεγάλες βάσεις με χρήσιμα στοιχεία των ψηφοφόρων τους
  - Με τον τρόπο αυτό ήταν πιο εύκολη η στόχευση των αναποφάσιστων ψηφοφόρων του 'αντίπαλου' κόμματος ώστε τουλάχιστον να μην ψηφίσουν την ημέρα των εκλογών
  - Αμφιλεγόμενες πρακτικές
  - Η αποκάλυψη των πρακτικών αυτών το 2018 ξεσήκωσε θύελλα διαμαρτυριών και σε βάρος του FB

# Μοντελοποίηση κοινωνικών φαινομένων

- Πολλά φαινόμενα ακολουθούν την κατανομή Gauss (πχ ευφυΐα, ηλικία, ύψος)
- Εκθετικές κατανομές: Zipf (γλωσσολόγος στο Harvard). Παρατήρησε ότι οι πλέον συχνά χρησιμοποιούμενες λέξεις παρουσιάζουν συγκεκριμένη κατανομή
- Η πλέον συχνά χρησιμοποιούμενη έχει διπλάσια συχνότητα εμφάνισης από τη 2η, τριπλάσια από την 3<sup>η</sup> (Reed, 2001)
- Pareto law: το 20% των ανθρώπων κατέχει το 80% του εισοδήματος
- 1% των χρηστών δημιουργεί το 99% του περιεχομένου (Hargittai & Walejko, 2008)
- Στη Wikipedia το 1/1000 των χρηστών δημιουργεί το 44% του περιεχομένου
- Αντίστοιχα για τη δημοφιλία χρηστών (πχ Facebook likes)

# Παραδείγματα εφαρμογών Big data

- Συχνά οργανισμοί όπως το FB δεν δίνουν τα δεδομένα τους σε άλλους κοινωνικούς επιστήμονες. Αυτό υπονομεύει την εξέλιξη των επιστημών αυτών.
- Στην αντίθετη κατεύθυνση το [datacommons.org](https://datacommons.org)
  - Συγκεντρώνει στοιχεία από διαφορετες πηγες
  - Ευρωπαϊκή και Αμερικανική στατιστική υπηρεσία
  - Παγκόσμια Τράπεζα
  - ΟΟΣΑ
  - Wikipedia



# Google Trends

- Ανάλυση δημοφιλών αναζητήσεων από G. Search
- Απεικόνιση εξέλιξής τους στο χρόνο
- Τα δεδομένα μπορούν να εξαχθούν και να οπτικοποιηθούν και στο [datawrapper](#)

See what was trending in 2020 - Greece

Ταχύτερα Αυξανόμενες Αναζητήσεις	Διασημότητες	Εκλιπούσες Διασημότητες
1 eclass	1 Σωτήρης Τσιόδρας	1 Kobe Bryant
2 Κρούσματα	2 Νίκος Χαρδαλιάς	2 Κώστας Βουτσάς
3 Κορωνοϊός	3 Donald Trump	3 Diego Maradona
4 e-me	4 Ιωάννα Τούνη	4 Κική Δημουλά
5 webex	5 Joe Biden	5 Naya Rivera

Αθλητικά γεγονότα	Επαναλαμβανόμενα γεγονότα	Επικαιρότητα
1 NBA	1 Black Friday	1 Κορωνοϊός
2 ΠΑΟΚ - Ολυμπιακός	2 25η Μαρτίου	2 Εκλογές ΗΠΑ
3 Άρσενal - Ολυμπιακός	3 Πάσχα	3 Αγία Σοφία
4 Παναθηναϊκός - ΠΑΟΚ	4 Αγίου Βαλεντίνου	4 Έβρος
5 Premier League	5 Πανελλήνιες	5 Επίθεση με βιτριόλι

Κοντά μου...	Πού...	Πότε...
1 Φαρμακείο κοντά μου	1 Πού έγινε σεισμός	1 Πότε ανοίγουν τα σχολεία
2 Βενζινάδικο κοντά μου	2 Πού είναι υποχρεωτική η μάσκα	2 Πότε ανοίγουν τα μαγαζιά
	3 Πού εντοπίζονται...	3 Πότε ανοίγουν τα...

# Άλλες εφαρμογές

- Εκτίμηση φτώχειας από φωτογραφίες από δορυφόρο (Jean et al., 2016)
- Ανάλυση / κατηγοριοποίηση posts από social media με τεχνικές ανάλυσης περιεχομένου (Fu, Guo, & Land, 2018)
- Πρόβλεψη βίας σε φυλακές (Bacak & Kennedy, 2018)
- Εμπλουτισμός ποιότητας αποτελεσμάτων ερωτηματολογίων από δεδομένα συμπεριφοράς (Fu, Guo, & Land, 2018)
- Ανάλυση συναισθημάτων (sentiment analysis) από δεδομένα (Joseph, Wei, Benigni, & Carley, 2016)

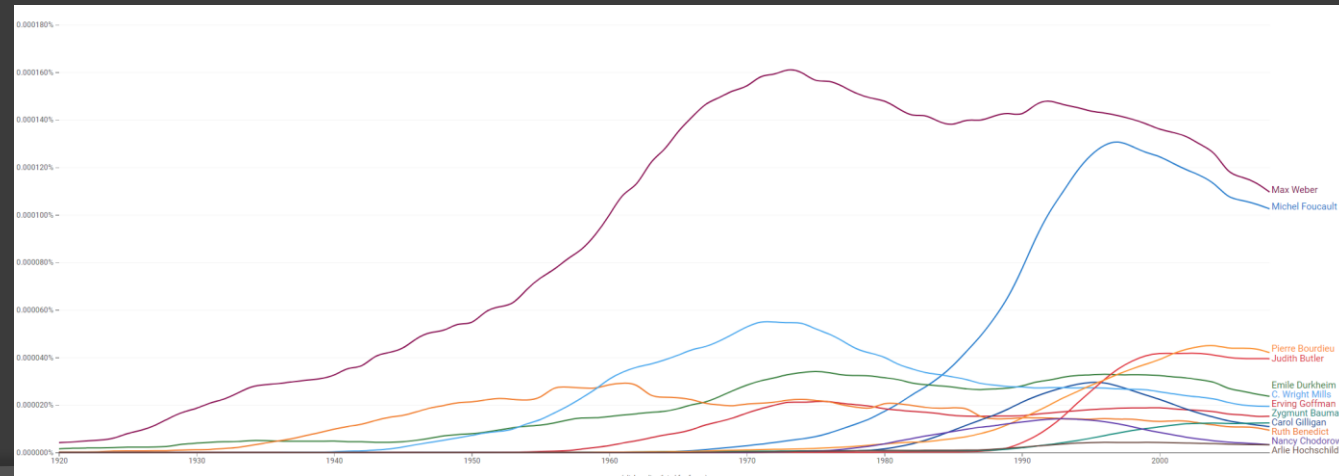
# Άλλες τεχνικές

- ⦿ Social network analysis
- ⦿ Online polls
- ⦿ Culturomics
- ⦿ Scraping



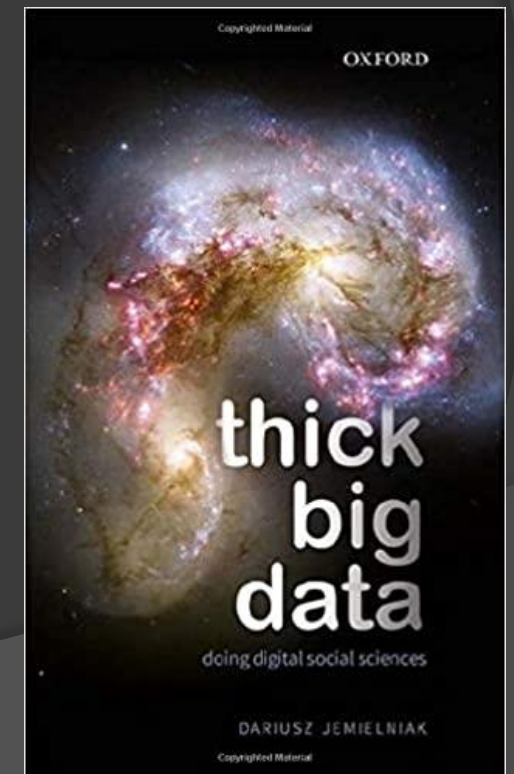
# Culturomics

- Ψηφιοποίηση μεγάλου όγκου βιβλίων
- Ανάλυση με ποσοτικές τεχνικές διάφορων ζητημάτων
- Πχ εκτίμηση επιρροής κοινωνιολόγων
- Εξέλιξη ενδιαφέροντος για το πεδίο αλληλεπίδρασης ανθρώπου υπολογιστή



# Σχετικό βιβλίο

- Jemielniak, D. (2020). *Thick big data: Doing digital social sciences*. Oxford University Press.



# Τέλος Ενότητας



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ