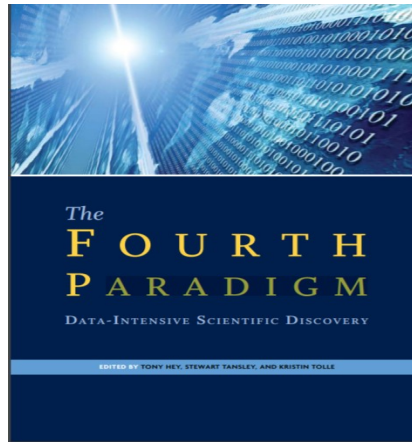


Μια εκτεταμένη σύνοψη του βιβλίου “Το Τέταρτο Πρότυπο”

Πατίρης Παναγιώτης



Επιβλέπων καθηγητής: Δρ. Μηνάς Δασυγένης
Πανεπιστήμιο Δυτικής Μακεδονίας
Τμήμα Μηχανικών Πληροφορικής & Τηλεπικοινωνιών
Εργαστήριο Ψηφιακών Συστημάτων και Αρχιτεκτονικής Υπολογιστών, <http://arch.icte.uowm.gr/>
Κοζάνη 2018



Πανεπιστήμιο Δυτικής Μακεδονίας

Το Τέταρτο Πρότυπο

Μία επιστημονική ανακάλυψη για την υπερπαραγωγή δεδομένων

Δημοσίευση από Tony Hey, Stewart Tansley and Kristin Tolle



Περιεχόμενα

- Η γη και το περιβάλλον
- Υγεία και ευεξία
- Επιστημονική υποδομή
- Ακαδημαϊκή επικοινωνία



Εισαγωγή



Υπερπαραγωγή δεδομένων

- Η εποχή μας μοιάζει με την εποχή που ανακαλύφθηκε η τυπογραφία, όταν η παραγόμενη γνώση και το ενδιαφέρον για αυτήν, δεν μπορούσε να ικανοποιηθεί με τα μέσα καταγραφής και διάδοσής της που υπήρχαν τότε.
- Σήμερα υπάρχει αδυναμία των υφιστάμενων μέσων να καταγράψουν αποτελεσματικά δεδομένα από την έρευνα μεμονωμένων ερευνητών, συνεχών μετρήσεων από αισθητήρες και άλλων στοιχείων που έτσι κινδυνεύουν να χαθούν.
- Επιτακτική είναι επίσης η ανάγκη πιο αποτελεσματικών τρόπων επεξεργασίας και αρχειοθέτησής των δεδομένων, ώστε να είναι προσβάσιμα διαχρονικά και διεπιστημονικά από όσους θα μπορούσαν να τα αξιοποιήσουν.



Ανάγκη για νέα εργαλεία διαχείρισης δεδομένων

- Η “ψηφιακή βιβλιοθήκη” θα αντιμετωπίσει αυτήν την πρόκληση.
- Απαιτείται η διάθεση πόρων και μέσων για τη δημιουργία των κατάλληλων εργαλείων που θα καλύψουν αυτήν την ανάγκη.
- Ήδη έχει γίνει αντιληπτή αυτή η ανάγκη και έχουν ξεκινήσει σχετικές προσπάθειες σε ορισμένους επιστημονικούς τομείς.
- Πρωτοποριακός υπήρξε ο ρόλος του James Gray στην κατανόηση της ανάγκης, αλλά και των βασικών αρχών της νέας αυτής επιστημονικής προσπάθειας.



Η γνώμη του Jim Gray για την eScience: Μια μεταμορφωμένη επιστημονική μέθοδος

Δημοσίευση από τον Tony Hey, τον Stewart Tansley και τον Kristin Tolle



Η πορεία της έρευνας

- Η πορεία της Έρευνας: Εμπειρικό στάδιο-θεωρητικό-υπολογιστικό.
- Υπερπαραγωγή δεδομένων, μικρό ποσοστό αξιοποίησής τους.
- Ανάγκη για νέα εργαλεία εξερεύνησης-αξιοποίησης της υπερπληθώρας δεδομένων.
- eScience: Η νέα επιστημονική ανακάλυψη-Το τέταρτο πρότυπο.



Τα τέσσερα στάδια της επιστήμης

- Εκατοντάδες χρόνια πριν η γνώση ήταν εμπειρική (περιγράφοντας φυσικά φαινόμενα).
 - Τα τελευταία εκατό χρόνια έγινε μετάβαση στον θεωρητικό τομέα (χρήση μοντέλων).
 - Τις τελευταίες δεκαετίες υπάρχει μια πιο υπολογιστική προσέγγιση (προσομοιώνοντας περίπλοκα φαινόμενα).
 - Πλέον επιτεύχθηκε η εξερεύνηση δεδομένων μέσω ενοποιημένων θεωριών, πειραμάτων προσομοιώσεων κτλ.
 - Δεδομένα συλλέγονται με κατάλληλα όργανα.
 - Επεξεργάζονται από κατάλληλα λογισμικά.
 - Αποθηκεύονται πληροφορίες σε υπολογιστές.
 - Επιστήμονες αναλύουν βάσεις δεδομένων μέσω στατιστικών.
-



Ανάγκη ψηφιακών εργαλείων για τη διαχείριση της υπερπαραγωγής δεδομένων

- Σήμερα τα δεδομένα προέρχονται από προσομοιώσεις ή από αυτόματα όργανα με αισθητήρες που παίρνουν μετρήσεις επί 24ωρου βάσεως, σε τεράστιους αριθμούς.
- Ο επιστήμονας πλέον δεν κοιτά μεμονωμένα ευρήματα, αλλά τα αποτελέσματα της λογισμικής επεξεργασίας τους και της συσχέτισή τους.
- Αυτή η διεργασία απαιτεί κατάλληλη κωδικοποίηση και παρουσίαση των αποτελεσμάτων της έρευνας με τη χρήση αλγορίθμων, ώστε να είναι εύκολα προσιτά.



Η περίπτωση μικρών προγραμμάτων

- Στα μεγάλα ερευνητικά προγράμματα διατίθενται πόροι για την ανάπτυξη του απαιτούμενου λογισμικού, με υψηλό όμως κόστος.
- Σε μικρότερα προγράμματα τα λογισμικά εργαλεία που υπάρχουν είναι ανεπαρκή.
- Απαιτείται η επεξεργασία ενός συστήματος διαχείρισης πληροφορίας “Laboratory Information Management Systems” (LIMS), που θα αποτελέσει τον “αγωγό” από τα πειραματικά δεδομένα και μετρήσεις σε δομημένες βάσεις πληροφοριών.



Επιστημονική επικοινωνία: η ψηφιακή βιβλιοθήκη

- Έρευνα που χρηματοδοτείται από δημόσιους πόρους πρέπει να δημοσιεύεται στο διαδίκτυο με ανοιχτή πρόσβαση, αλλά αυτό δεν είναι αρκετό.
 - Απαιτείται η συσχέτιση της κάθε δημοσίευσης με άλλες παρεμφερείς έρευνες ακόμη και από συγγενείς επιστημονικούς τομείς και με τη σχετική βιβλιογραφία, ώστε να υπάρχει ολοκληρωμένη πληροφόρηση σε ελάχιστο χρόνο.
 - Μετάβαση στην ελεύθερης πρόσβασης ψηφιακή βιβλιοθήκη, όπου με τα κατάλληλα λογισμικά εργαλεία θα αποδελτιώνονται και θα αρχειοθετούνται αξιολογημένες δημοσιεύσεις, ώστε να πληρούνται οι παραπάνω απαιτήσεις.
 - Η παραπάνω διαδικασία συνιστά ριζική αλλαγή σε σχέση με την υφιστάμενη κατάσταση και θα δώσει σημαντική ώθηση στην επιστήμη, όταν θα ολοκληρωθεί η εφαρμογή της.
-



Η ΓΗ ΚΑΙ ΤΟ ΠΕΡΙΒΑΛΛΟΝ



Εισαγωγή

DAN FAY Microsoft Research

- Σήμερα υπάρχει μία μετατόπιση από την προσομοίωση προς την υπολογιστική και την αξιοποίηση βάσεων δεδομένων.
- Οριακή δυνατότητα μελέτης πολύπλοκων φαινομένων όπως η κλιματική αλλαγή, το περιβάλλον, η βιολογία και η μελέτη του διαστήματος με τα διαθέσιμα υπολογιστικά εργαλεία.
- Υπάρχει ανάγκη για νέα εργαλεία για τη συλλογή, την επεξεργασία και την αρχειοθέτηση δεδομένων για την καλύτερη κατανόηση των φαινομένων και τη λήψη πιο τεκμηριωμένων αποφάσεων σε κρίσιμα θέματα.
- Στα άρθρα που ακολουθούν αναλύονται οι παραπάνω τάσεις στις επιστήμες της γης και του περιβάλλοντος.



Οι νόμοι του Gray: Η επικεντρωμένη σε βάσεις δεδομένων υπολογιστική στην επιστήμη (1/2)

Alexander S. Szalaay, The Johns Hopkins University José A. Blakeiey, Microsoft

- Οι νόμοι του Gray που κωδικοποιούν το πώς να προσεγγίσουμε την επεξεργασία των δεδομένων σε μεγάλη κλίμακα είναι οι παρακάτω:
 - Η επιστημονική υπολογιστική γίνεται ολοένα πιο πυκνή σε δεδομένα.
 - Επιλογή αρχιτεκτονικής επέκτασης (scale out) και όχι αντικατάστασης (scale up).
 - Προσαρμογή του υπολογισμού στα δεδομένα και όχι το αντίθετο καθώς αυτός είναι ο πιο αποτελεσματικός τρόπος προσέγγισης.



Οι νόμοι του Gray: Η επικεντρωμένη σε βάσεις δεδομένων υπολογιστική στην επιστήμη (2/2)

Alexander S. Szalaay, The Johns Hopkins University José A. Blakeiey, Microsoft

- Να λαμβάνονται ως βάση του σχεδιασμού οι 20 πιο πιθανές ερωτήσεις.
- Μετάβαση από την μια εργασία στην άλλη καθώς, η αρχιτεκτονική που καθοδηγείται από τα δεδομένα αλλάζει πολύ γρήγορα ιδίως αν περιλαμβάνει και διανομή δεδομένων.



Από την κλίμακα Tera στην κλίμακα Peta

- Μετάβαση όγκου δεδομένων από την tera κλίμακα στην peta.
- Ανάγκη ανάλογης προσαρμογής των υπολογιστικών μεθόδων.
- Εφαρμογή των καλύτερων πρακτικών με τη χρήση:
 - κατάλληλων αρχιτεκτονικών hardware και αντίστοιχων software.
 - υπολογιστικής νέφους.



Η αναδυόμενη επιστήμη των περιβαλλοντικών εφαρμογών

Jeff Dozier, University of California, Santa Barbara William B. Gall, Microsoft

- Η Επιστήμη της Γης και του Περιβάλλοντος πέρασε ήδη δύο στάδια:
 - Μέχρι τις αρχές τις δεκαετίας του 80' ήταν επιστημονικό πεδίο της γεωλογίας, της ατμοσφαιρικής χημείας και των οικοσυστημάτων.
 - Το επόμενο διάστημα έγινε αντιληπτή η στενή αλληλοσύνδεση αυτών των επιστημονικών τομέων και το θέμα άρχισε να μελετάται σαν ένα ενιαίο σύνολο, την επιστήμη της Γης.
- Έτσι έγινε δυνατό να μελετηθούν σύνθετα φαινόμενα προσανατολισμένα στο σύστημα, όπως η κλιματική αλλαγή που:
 - Ενσωματώνει πτυχές και γνώσεις από τις επιμέρους επιστήμες.
 - Αξιοποιεί στοιχεία από δορυφορικές μετρήσεις που αναλύονται με κατάλληλα μοντέλα.



Η επιστήμη της γης στο σημερινό τρίτο στάδιο (1/2)

- Στην αναδυόμενη τρίτη φάση, η γνώση που αποκτήθηκε για επιστημονική κατανόηση συμπληρώνεται με γνώση που αποκτήθηκε για δράση και λήψη αποφάσεων.
- Η περίπτωση της κλιματικής αλλαγής είναι χαρακτηριστική.
 - Στα πρώτα δύο στάδια βρέθηκαν οι απαντήσεις στα πώς και τα γιατί του φαινομένου.
 - Τώρα είναι η ώρα για την ποσοτικοποίηση των διαφόρων πτυχών του θέματος, ώστε να ληφθούν συγκεκριμένες αποφάσεις.
 - Αυτό αποτελεί το αντικείμενο μίας νέας επιστήμης, που ενσωματώνει τις επιμέρους επιστημονικές γνώσεις και δίνει ποσοτικοποιημένες απαντήσεις.



Η επιστήμη της γης στο σημερινό τρίτο στάδιο (2/2)

- Άλλη περίπτωση είναι και η διαχείριση των υδάτινων αποθεμάτων μίας λεκάνης απορροής όπου η συλλογή στοιχείων και οι καταγραφές πρέπει να είναι:
 - Δυναμικές, να λαμβάνουν υπόψη τις μεταβολές που συντελούνται.
 - Τα αποτελέσματα να μπορούν να προβάλλονται στο μέλλον.



Η επιστήμη των περιβαλλοντικών εφαρμογών

- Γεννιέται έτσι μία νέα επιστήμη -Η Επιστήμη Περιβαλλοντικών Εφαρμογών- που διαφέρει σημαντικά από την Βασική Επιστήμη Γης και περιβάλλοντος στα παρακάτω σημεία:
 - Οδηγείται από την ανάγκη για λήψη συγκεκριμένων αποφάσεων και όχι από επιστημονική περιέργεια.
 - Είναι περιορισμένη από εξωτερικούς παράγοντες που υπαγορεύουν τη λήψη μέτρων.



Τα χαρακτηριστικά της επιστήμης των περιβαλλοντικών εφαρμογών

- Είναι συνεχής και αναδραστική. Η ίδια η ανθρώπινη δράση μεταβάλλει ενίοτε τα δεδομένα και χρειάζεται επανεξέταση της κατάστασης.
 - Είναι χρήσιμη ακόμη και όταν η γνώση είναι ημιτελής, αφού πολλάκις δεν είναι δυνατό να αναβάλουμε τη δράση μέχρι να έχουμε ολοκληρωμένη γνώση του φαινομένου.
 - Είναι κλιμακωτή, ώστε να μπορούμε να εκτιμήσουμε τις συνολικές διαστάσεις των φαινομένων και σε μεγαλύτερη κλίμακα.
 - Είναι δυναμική και στέρεη, ώστε η χρήση της από αυτούς που παίρνουν αποφάσεις και συχνά είναι αδαείς (πολιτικοί) να μην ενέχει κινδύνους κατάχρησης .
 - Έχει μεγάλη συγκέντρωση δεδομένων με συμμετοχή συχνά και της κοινωνίας στη συλλογή στοιχείων.



Μέσα απόκτησης δεδομένων-Η περίπτωση μίας λεκάνης απορροής

- Για την απόκτηση των εκτεταμένων δεδομένων που απαιτούνται για την αναδυόμενη αυτή επιστήμη χρειάζεται η ανάπτυξη ενός συστήματος συλλογής και διαβίβασης μετρήσεων.
 - Για την περίπτωση της μελέτης μίας λεκάνης απορροής είναι:
 - Ακριβείς δορυφορικές παρατηρήσεις της χιονοκάλυψης.
 - Δίκτυο αισθητήρων χιονιού και υγρασίας.
 - Δεδομένα από κοινωνικές επιστήμες.
 - Κυβερνο-υποδομή για την ενσωμάτωση των δεδομένων και την μετάδοσή τους σε πραγματικό χρόνο, ώστε να μπορούν να ληφθούν αποφάσεις.
-



Επαναπροσδιορίζοντας την Οικολογική Επιστήμη με τη χρήση δεδομένων

James R. Hunt Πανεπιστήμιο της Καλιφόρνιας, Μπέρκλεϊ και το Κέντρο Νερού του Berkeley Dennis D.
Baldocchi Πανεπιστήμιο Καλιφόρνιας, Μπέρκλεϊ Catharine Van Ingen Microsoft Research



Προοπτικές της επιστήμης των περιβαλλοντικών εφαρμογών

- Αυτή η τρίτη φάση της Επιστήμης της Γης και του Περιβάλλοντος - η Επιστήμη Περιβαλλοντικών Εφαρμογών- εκτιμάται ότι θα αναπτυχθεί στις προσεχείς δεκαετίες χωρίς να σταματήσει η πρόοδος των συστατικών της επιμέρους βασικών επιστημών.
- Αντίθετα θα παντρέψει την πρόοδο των βασικών επιστημών με τις ανάγκες των ανθρώπων και των οργανισμών για καλύτερη γνώση και καλύτερα τεκμηριωμένες αποφάσεις.
- Το ίδιο το κοινό μπορεί να συνεισφέρει στην συλλογή δεδομένων με απλά όργανα όπως οι κάμερες των κινητών, ώστε να μοχλευτεί η απαιτούμενη μεγάλη συγκέντρωση δεδομένων.
- Η αναδυόμενη νέα επιστήμη θα συμβάλει σημαντικά στην ευημερία των 7 δις ανθρώπων του πλανήτη μας.



Η Εξέλιξη της Οικολογίας

- Οικολογία είναι η επιστήμη της ζωής και οι αλληλεπιδράσεις της με το φυσικό περιβάλλον.
 - Η έρευνα πλέον γίνεται σε μεγάλη κλίμακα.
 - Η διαχείριση των διαθέσιμων πόρων λαμβάνει κεφαλαιώδη σημασία.
 - Η υπολογιστική υποστήριξη αποτελεί κρίσιμο παράγοντα.
 - Το κλίμα σε παγκόσμιο επίπεδο και η μορφολογία της γης αλλάζουν.
 - Οι επιστήμονες πρέπει να ποσοτικοποιήσουν τη συσχέτιση ατμοσφαιρικών, γήινων και ωκεανικών διαδικασιών, ώστε να μελετήσουν κατάλληλα τα φαινόμενα.
-



Ανάγκη για σύνθεση δεδομένων (1/2)

- Φαινόμενα όπως ο επηρεασμός του περιεχομένου σε άνθρακα της συνολικής φυτικής ύλης από τα μετεωρολογικά φαινόμενα ή η υπερθέρμανση, για να μελετηθούν απαιτούν σύνθεση δεδομένων και μοντέλων που εκτείνονται τοπικά και χρονικά σε κλίμακα από το πολύ μικρό ως το πολύ μεγάλο.
- Αυτές οι νέες μελέτες σύνθεσης γίνονται εφικτές από:
 - Την εκτεταμένη χρήση αισθητήρων χαμηλού κόστους.
 - Την τηλεανίχνευση πχ μέσω δορυφόρου ακόμη και για δυσπρόσιτες περιοχές.
 - Τη συνδεσιμότητα στο Διαδίκτυο.
 - Τη διαθεσιμότητα υπολογιστών.



Ανάγκη για σύνθεση δεδομένων (2/2)

- Στην οικολογία αντίθετα με άλλες επιστήμες τα δεδομένα παράγονται από μεμονωμένα γκρουπ επιστημόνων με διαφορετικές διαδικασίες και μεθόδους καταγραφής.
 - Για να επιτευχθεί η σύνθεση των δεδομένων χρειάζεται:
 - Ψηφιακή πρόσβαση σε βάσεις δεδομένων.
 - Κατάλληλη επεξεργασία των δεδομένων.
 - Επαλήθευσή αποτελεσμάτων.
 - Άλλες φορές η σύνθεση απαιτεί πρόσβαση σε δημόσιες ή άλλες βάσεις δεδομένων.
 - Η ενσωμάτωση όλων των πληροφοριών είναι μία πρόκληση για την επιστημονική ανάλυση.
-



Νέα δεδομένα για τους μελετητές και ανάγκη για συν δημοσιεύσεις (1/2)

- Η περαιτέρω επεξεργασία των δεδομένων οδηγεί στην ανάγκη της χρήσης της υπολογιστικής νέφους.
- Ανάγκη συνεργασίας μεταξύ επιστημόνων και συν δημοσιεύσεις.
- Τίθενται ηθικής φύσεως ζητήματα που πρέπει να απαντηθούν και σχετίζονται με την συνεισφορά δεδομένων που αποκτήθηκαν από κάποιους επιστήμονες με κόπο και προσπάθεια.
- Αν λυθούν αυτά τα θέματα θα μειωθούν πολύ τα διαδικαστικά προβλήματα των δημοσιεύσεων.



Νέα δεδομένα για τους μελετητές και ανάγκη για συν δημοσιεύσεις (2/2)

- Η πλημμύρα των οικολογικών δεδομένων και η ανάγκη για σύνθεση:
 - Σπάει τα στεγανά μεταξύ συγγενών επιστημονικών τομέων.
 - Θα συνδέσει παρατηρήσεις πεδίου με τα αποτελέσματα μοντέλων σε συνδυασμό:
 - Με την χρήση της υπολογιστικής νέφους.
 - Το διαμοιρασμό δεδομένων.
 - Την πρόκληση για μία κατάλληλη κυβερνο-υποδομή
 - Θα οδηγήσουν τα προσεχή χρόνια σε μία μία νέα διατομεακή επιστήμη.



Το όραμα του 2020 για την επιστήμη των Ωκεανών

John R. Delaney University of Washington Roger S. Barga Microsoft Research



Η σημασία των ωκεανών στη διαμόρφωση του κλίματος και της ζωής στη γη

- Οι ωκεανοί:
 - Καλύπτουν το 70% της επιφάνειας της γης.
 - Αποτελούν μία τεράστια δεξαμενή ύλης και ενέργειας.
 - Διαμορφώνουν το κλίμα.
 - Επηρεάζουν άμεσα την παραγωγή τροφής, τις πλημμύρες και την ξηρασία στη γη.
 - Αποτελούν το τελικό αποδέκτη των αποβλήτων.
 - Αποτελούν όμως και εν δυνάμει κίνδυνο για τσουνάμι, σεισμούς και τυφώνες.
 - Είναι συνεπώς κρίσιμης σημασίας για την ύπαρξή μας η καλύτερη κατανόησή τους.
 - Η πρόκληση είναι να μεγιστοποιηθούν τα οφέλη και να περιοριστούν οι κίνδυνοι από ένα σύστημα στο οποίο ανά πάσα στιγμή λαμβάνουν χώρα χιλιάδες περίπλοκες βιολογικές, χημικές και φυσικές διεργασίες σε πολλές κλίμακες χρόνου και διαστήματος.
-



Η εξέλιξη της ωκεανογραφικής έρευνας (1/2)

- Τα περασμένα χρόνια η ωκεανογραφική έρευνα βασίσθηκε σε στοιχεία από πλοία υποβρύχια και δορυφόρους.
- Σήμερα η εξερεύνηση και η εκπαίδευση θα βασισθεί στην εμφάνιση και τη σύγκλιση πολλών ταχέως εξελισσόμενων νέων καινοτόμων τεχνολογιών.
- Ορισμένες από αυτές τις δυνατότητες θα επιτρέψουν την τηλεανίχνευση κρίσιμων αλλαγών που λαμβάνουν χώρα σε επιλεγμένους όγκους ωκεανού.



Η εξέλιξη της ωκεανογραφικής έρευνας (2/2)

- Η ταχεία αναδιάταξη των βασικών συστοιχιών αισθητήρων που συνδέονται με το Διαδίκτυο θα μας επιτρέψει:
 - Να καταγραφούν.
 - Να απεικονιστούν.
 - Να τεκμηριωθούν.
 - Να μετρηθούν φαινόμενα όπως η έκρηξη ηφαίστειων, τα μεγάλα μεταναστευτικά σχήματα, οι μεγάλες υποβρύχιες καταστροφές, καταιγίδες και μια σειρά άλλων πολύπλοκων φαινομένων, που ήταν σε μεγάλο βαθμό απρόσιτα για την επιστημονική μελέτη.
-



Το τέταρτο πρότυπο στην ωκεανογραφική έρευνα (1/2)

- Η μελέτη των ωκεανών απαιτεί πληθώρα δεδομένων και αποτελεσματικό τρόπο επεξεργασίας τους, πράγμα που ως τώρα δεν έχει καταστεί δυνατόν.
 - Για να αποκτήσει κάποια προβλεπτική αξία η μελέτη των ωκεανών θα πρέπει να καταστεί εφικτή η ποσοτικοποίηση φαινομένων, όπως γιγάντιες καταιγίδες και η έκρηξη ηφαιστείων.
 - Η σύγκριση των προβλέψεων των μοντέλων με τα πραγματικά φαινόμενα θα μας επιτρέψει να κάνουμε τις απαραίτητες προσαρμογές, ώστε τα μοντέλα να γίνουν πραγματικά χρήσιμα.
 - Απαιτείται η συγκέντρωση στοιχείων, όπως ροών, από ένα ευρύ δίκτυο αισθητήρων σε πραγματικό χρόνο και για πολύ μεγάλα διαστήματα.
 - Τα δεδομένα πρέπει να συλλέγονται, να αρχειοθετούνται, να απεικονίζονται και να συγκρίνονται αμέσως με τις προσομοιώσεις μοντέλων.
-



Το τέταρτο πρότυπο στην ωκεανογραφική έρευνα (2/2)

- Η κατάλληλη αρχειοθέτηση των δεδομένων και η δυνατότητα πρόσβασης σε αυτά όλων, ταυτόχρονα, με την λήψη ή και σε μεταγενέστερο χρόνο, θα αυξήσει τις ερευνητικές δυνατότητες και θα αποδώσει καλύτερα αποτελέσματα.
- Για τους επιστήμονες που θα εργάζονται στο νέο αυτό περιβάλλον της πληθώρας των δεδομένων θα χρειάζονται νέα εργαλεία για την επεξεργασία, διαχείριση και αρχειοθέτηση.



Η πρωτοβουλία για το ωκεανογραφικό παρατήριο

- Πρόκειται για ένα 25ετές πρόγραμμα συνεργασίας ΗΠΑ και Καναδά που περιλαμβάνει μετρήσεις από παράκτιο εξοπλισμό, από δορυφόρους, αλλά και από ένα ευρύ δίκτυο ηλεκτροπτικά καλωδιωμένου (συνδεδεμένου) συστήματος παρατηρήσεων στο βορειοανατολικό Ειρηνικό Ωκεανό.
- Αυτό το ευρύτατο δίκτυο αισθητήρων-οργάνων και ρομπότ στις ΗΠΑ και το Καναδά θα αποτελεί ένα τεράστιο φυσικό εργαστήριο με τεράστιες δυνατότητες ανεύρεσης δεδομένων, επεξεργασίας και αρχειοθέτησής τους.
- Στο πρόγραμμα θα συνεργαστούν πανεπιστήμια από τις ΗΠΑ και το Καναδά.
- Τα αποτελέσματα θα είναι εντυπωσιακά για τους επιστήμονες, τους φοιτητές και το διδακτικό προσωπικό, αλλά και για το κοινό όπως φαίνεται και στην παρακάτω εικόνα, ενώ προβλέπονται ακόμη πιο εντυπωσιακά αποτελέσματα μέχρι το 2020.



Παρατηρητήριο του μέλλοντος



Επιστήμονες ή πολίτες της επόμενης γενιάς. Αυτή η εικονική φωτογραφία παρουσιάζει ένα χταπόδι βαθύ ωκεανού, γνωστό ως *Grimptoteuthis*, και ένα τμήμα ενός υποθαλάσσιου υδροθερμικού συστήματος στο Juan de Fuca Ridge. Τέτοιες απεικονίσεις βίντεο 3-D HD σε πραγματικό χρόνο θα είναι ρουτίνα μέσα σε 5 χρόνια.



“Φέρνοντας το Νυχτερινό ουρανό πιο κοντά”: Ανακαλύψεις από το κατακλυσμό δεδομένων

Alyssa A. Goodman Harvard University Curtis G. Wong Microsoft Research



Εισαγωγή

Στην παρουσίαση αυτή εξετάζονται τρεις αλληλοσυσχετιζόμενες ερωτήσεις:

- Οι τάσεις που διαπιστώθηκαν και αναμένονται, στην ανάπτυξη συλλογής εικόνων και δεδομένων από τηλεσκόπια.
- Πώς θα καταστεί δυνατόν να αντιμετωπιστεί η πρόκληση για την εύρεση των κρίσιμων για την επιστήμη στοιχείων μέσα σε ένα κυκεώνα δεδομένων.
- Ποιες οπτικοποιήσεις και αναλυτικές ευκαιρίες έχει το μέλλον.



Τάσεις στην αύξηση των δεδομένων (1/2)

- Οι δύο σταθμοί στην αστρονομία, τα τηλεσκόπια και η χρήση του ηλεκτρομαγνητικού φάσματος πέραν του ορατού, δεν οδήγησαν στην παραγωγή τόσο όγκου δεδομένων, όσο στην ψηφιακή εποχή.
- Οι συσκευές συζευγμένου φορτίου, Charge-coupled devices (CCDs), απεδείχθησαν πολύ πιο αποτελεσματικές και παραγωγικές από προηγούμενες τεχνικές στη συλλογή δεδομένων.
- Η επεξεργασία ανάλυση και αποθήκευση των δεδομένων με ψηφιακά εργαλεία μετέτρεψε την αστρονομία από μία επιστήμη παρατήρησης σε ψηφιακή και υπολογιστική επιστήμη.



Τάσεις στην αύξηση των δεδομένων (2/2)

- Η παραγωγή δεδομένων πλέον είναι στην κλίμακα peta και υποτίθεται ότι η λύση της διαχείρισής τους θα δοθεί μέσω διαδικασιών, όπως η υπολογιστική νέφους, αλλά και τάσεων, όπως η αξιοποίηση ερασιτεχνών με ευρύτερη και πιο ελεύθερη πρόσβαση στα δεδομένα.
 - Η αξιοποίηση ήδη αρχειοθετημένων δεδομένων μπορεί να γίνει μέσω νέων υπολογιστικών εργαλείων, όπως το Παγκόσμιο τηλεσκόπιο της Microsoft (WWT).
 - Με το WWT κάποιος μπορεί να περιηγηθεί στον ουρανό, να χρησιμοποιήσει μεγάλο εύρος από μήκη κύματος, να έχει για κάθε σημείο του ουρανού όλο το ιστορικό της έρευνας που έχει γίνει και ένα σωρό άλλες πληροφορίες και θα αποτελέσει στο μέλλον μέρος την Online αστρονομίας για επιστήμονες, αλλά και για ενδιαφερόμενους πολίτες.
-



Αυξάνοντας την υπολογιστική ισχύ

- Οι πολυπύρηντοι επεξεργαστές στην προσεχή δεκαετία θα αυξήσουν την υπολογιστική ισχύ δύο με τρεις τάξεις μεγέθους σε σχέση με τη σημερινή.
 - Η αυξημένη υπολογιστική ισχύς θα βοηθήσει στην διαχείριση της επερχόμενης πλημμυρίδας δεδομένων.
 - Θα επιτρέψει τον άμεσο εντοπισμό δεδομένων και συμβάντων και την διοχέτευσή τους σε δίκτυα επιστημόνων και πολιτών για την επιβεβαίωσή τους.
 - Η αύξηση της ροής δεδομένων θα δημιουργήσει την ανάγκη για πιο εξειδικευμένα και αποτελεσματικά εργαλεία software για την πιο αποτελεσματική διαχείρισή τους.
-



“Διασωληνώνοντας” τη Γη: Δίκτυα αισθητήρων επόμενης γενιάς και περιβαλλοντική επιστήμη

Michael Iehning Nicholas Dawes Mathias Bavay WSL Institute for Snow and Avalanche Research SLF

Marc Pariange École Polytechnique Fédérale de Lausanne

Suman Nath Feng zhao Microsoft Research



Εισαγωγή

- Οι νέες περιβαλλοντικές προκλήσεις και η κλιματική αλλαγή κάνουν επιτακτική την ανάγκη για μία νέα προσέγγιση.
 - Τα συστήματα παρακολούθησης δεν έχουν την απαιτούμενη ανάλυση, ώστε να τροφοδοτήσουν επαρκώς τα προγνωστικά μοντέλα.
 - Οι αλληλεπιδράσεις μεταξύ των παραγόντων που διαμορφώνουν το κλίμα είναι τόσο πολλές που για να μελετηθούν και να μοντελοποιηθούν χρειάζονται δεδομένα της τάξεως των terabytes.
 - Ανάγκη συνεργασίας των περιβαλλοντολόγων με τους πληροφορικούς για την ανάπτυξη τόσο των αισθητήρων όσο και των βάσεων δεδομένων με επαρκή τρόπο.
-



Η περιβαλλοντική επιστήμη της επόμενης γενιάς

- Θα έχει τα εξής χαρακτηριστικά:
 - Οι προβλέψεις για τις κλιματικές αλλαγές και το καιρό θα αντιστοιχούν με τις αλληλεπιδράσεις γης-ατμόσφαιρας.
 - Η μέση επίδραση μίας ανομοιογενούς επιφάνειας στην ατμόσφαιρα μπορεί να είναι πολύ διαφορετική από αυτή που προκύπτει από το μέσο όρο ενός παράγοντα της επιφάνειας (πχ της υγρασίας).
- Την επόμενη γενιά της περιβαλλοντικής επιστήμης θα χαρακτηρίζει η χρήση ευρύτατων δικτύων ασύρματων κυρίως αισθητήρων υψηλής ανάλυσης και αξιοπιστίας, οι οποίοι θα διαβιβάζουν τα δεδομένα σε εργαστήρια για περαιτέρω επεξεργασία και αρχειοθέτηση.
- Οι ασύρματοι αισθητήρες θα λύσουν πολλά προβλήματα αλλά απαιτούν:
 - Ενεργειακή αυτονομία.
 - Ιδιαίτερη ανθεκτικότητα για λειτουργία σε αντίξοες συνθήκες.
 - Να υπάρχει δυνατότητα εντοπισμού και αντιμετώπισης εσφαλμένων δεδομένων.



Διαχείριση και εξερεύνηση μεγάλου όγκου δεδομένων από αισθητήρες

- Ο τεράστιος όγκος (πολλών terabyte) υψηλής ανάλυσης δεδομένων από αισθητήρες απαιτεί την αξιόπιστη διαχείριση, επεξεργασία και αρχειοθέτηση τους υπό μορφή πλέον μεταδεδομένων.
- Οι επιστήμονες μπορούν για το σκοπό αυτό να αξιοποιούν μέχρι ένα βαθμό εργαλεία από βάσεις δεδομένων που χρησιμοποιούνται στο εμπόριο.
- Απαιτείται η ανάπτυξη εξειδικευμένων λογισμικών εργαλείων για τη διαχείριση του όγκου των δεδομένων από τα δίκτυα αισθητήρων.
- Με τη χρήση των κατάλληλων μοντέλων και σε συνδυασμό με άλλα δεδομένα από διαφορετικές πηγές θα προκύψει μία “αφομοίωση” των δεδομένων και τελικά μία γρήγορη και έγκυρη πρόβλεψη φαινομένων, όπως πχ πλημμύρες.
- Για να είναι χρήσιμη μια τέτοια πρόβλεψη για λήψη αποφάσεων και για το κοινό, θα πρέπει να παρουσιάζεται και να μεταδίδεται με ένα προσιτό τρόπο.



Παγκόσμια περιβαλλοντική παρακολούθηση

- Στο μέλλον ευρύτατα συστήματα παρακολούθησης θα επιτρέψουν:
 - Τη συνεργασία πολλών επιστημόνων.
 - Το διαμοιρασμό δεδομένων από εκατομμύρια αισθητήρες για να καταλάβουν καλύτερα τις παγκόσμιες περιβαλλοντικές τάσεις.
 - Αυτή η παγκόσμια διάσταση θα απαιτήσει νέο τρόπο διαχείρισης δεδομένων.
 - Θα πρέπει η επεξεργασία των δεδομένων να γίνεται στην πηγή.
 - Επίσης θα πρέπει τα δεδομένα να συμπιέζονται και να ομογενοποιούνται για να είναι πιο εύκολα διαχειρίσιμα.
 - Τα υπολογιστικά εργαλεία που θα αναπτυχθούν, θα πρέπει να χρησιμοποιηθούν από τους αντίστοιχους επιστήμονες. Για αυτό θα πρέπει να είναι εύκολα στη χρήση, δυναμικά και προσαρμοσμένα στο αντικείμενο.
-



ΥΓΕΙΑ ΚΑΙ ΕΥΞΕΙΑ



Εισαγωγή (1/2)

Simon mercer | Microsoft Research

- Το παρόν κεφάλαιο ασχολείται με το κρίσιμο για όλους θέμα της υγείας και των φαρμάκων.
- Γίνεται μία ιστορική παρουσίαση των προόδων της ιατρικής επιστήμης και των εξελίξεων που επίκεινται, καθώς και των επιμέρους εξελίξεων στους διάφορους τομείς.
- Η διάδοση των νέων τεχνολογιών κάνει εφικτή την τήρηση ατομικών φακέλλων υγείας και την γρήγορη διάγνωση.
- Η κατανόηση της λειτουργίας του εγκεφάλου με τα 160 τρις συνάψεις είναι μία μεγάλη πρόκληση για την επιστήμη και αναμένεται να συνδυαστεί με τις εξελίξεις στο θέμα της τεχνητής νοημοσύνης.



Εισαγωγή (2/2)

Simon mercer | Microsoft Research

- Η εκτιμώμενη απόκτηση προσβάσιμων ψηφιακών ατομικών φακέλλων υγείας με πλήρη στοιχεία για ένα μεγάλο μέρος του πληθυσμού της γης, θα δημιουργήσει μία τεράστια βάση δεδομένων που θα αξιοποιηθεί κατάλληλα και θα δώσει ώθηση στην καθιέρωση των καλύτερων πρακτικών στην επιλογή κλινικών πρακτικών.
- Θα αναπτυχθούν κατάλληλα μοντέλα που θα αξιοποιήσουν το περιεχόμενο των βάσεων δεδομένων προς την κατεύθυνση πρόβλεψης της συμπεριφοράς έναντι συγκεκριμένων ερεθισμάτων, του τρόπου που σκεφτόμαστε, καθώς και των μηχανισμών των ασθενειών.



Η Μοναδικότητα της υγειονομικής περίθαλψης και η εποχή της σημασιολογικής ιατρικής

Michael Gillam Craig Feied Jonathan Hanndler Eliza Moody Microsoft

Ben Seheiderman Catherine Plaisant University of Maryland

Mark Smith Medstar Health Institutes for Innovation John Dickason Private practice



Ιστορική αναδρομή

- Η μετάβαση από την επιστημονική ανακάλυψη στην καθημερινή εφαρμογή της, ιστορικά, απαιτεί μεγάλο χρονικό διάστημα.
- Για την περίπτωση πχ του σκορβούτου αυτό το διάστημα ήταν 264 χρόνια και παρά την τεράστια βελτίωση που έχει επέλθει παραμένει ακόμη στα 17 με μεγάλο κόστος σε ανθρώπινες ζωές, αλλά και στην οικονομία.
- Ακόμη η γνώση που ένας γιατρός πρέπει να έχει αυξάνει με εκθετικούς ρυθμούς.
- Ένας γενικός γιατρός πρέπει να είναι ενήμερος για 10.000 ασθένειες 3000 φαρμακευτικές αγωγές και 1100 εργαστηριακά τεστ ή διαφορετικά για να κρατιέται ενήμερος ένας γιατρός έχει εκτιμηθεί πως πρέπει να διαβάζει 21 ώρες το 24ωρο και όλα αυτά αυξάνονται πολύ γρήγορα.



Θα μπορούσε η υιοθέτηση των ιατρικών ανακαλύψεων να γίνει άμεση;

- Ιστορικά το διάστημα μεταξύ της ιατρικής ανακάλυψης και της εφαρμογής της μικραίνει συνεχώς και ιδιαίτερα τα τελευταία 150 χρόνια που συμπίπτουν με τη βιομηχανική επανάσταση.
- Μάλιστα η γραφική παράσταση της σχέσης μεταξύ των δύο μεγεθών δείχνει μία τάση το διάστημα αυτό να μηδενίζεται γύρω στα 2025, δηλαδή η εφαρμογή των ιατρικών ανακαλύψεων να είναι άμεση.
- Έρευνες δείχνουν ότι το έτος 2045 θα είναι το έτος της “Μοναδικότητας” δηλαδή το έτος που η ικανότητα των υπολογιστών θα ξεπεράσει αυτή των ανθρώπων και η δυνατότητα τους να αυτοβελτιώνονται θα οδηγήσει σε μία έκρηξη ευφυΐας.
- Σήμερα η διάδοση της ιατρικής επιστημονικής γνώσης περιλαμβάνει πολλές διαδικασίες όπως διαλέξεις, συγγράμματα, φυλλάδια κλπ πράγμα που αποτελεί ήδη αναχρονισμό καθώς θα πρέπει να γίνει ψηφιακή και άμεση, η γνώση, να ρέει χωρίς εμπόδια από την έρευνα στην εφαρμογή.



Οι υποσχέσεις για ένα “Μοναδικό” κόσμο μετά την υγειονομική περίθαλψη (1/2)

- Πολλές φορές έχει αποδειχτεί εκ των υστέρων ότι ένα εγκεκριμένο αρχικά φάρμακο είχε τελικά παρενέργειες που στοίχισαν πολύ στην υγεία και τις ζωές των ασθενών.
- Σήμερα ακόμη η απόδειξη της επικινδυνότητας ενός φαρμάκου δεν ακολουθείται από την άμεση απόσυρσή του καθώς υπάρχει καθυστέρηση στην ενημέρωση.
- Σε ένα κόσμο μετά το σημείο της “μοναδικότητας” αυτό θα γίνεται στιγμιαία.
- Για να γίνει αυτό εφικτό τα επιστημονικά άρθρα θα είναι εφοδιασμένα με κώδικες και αλγορίθμους, έτσι ώστε, η παγκόσμια επιστημονική κοινότητα να μπορεί να μετέχει στην αξιολόγησή τους και αυτό θα οδηγεί τάχιστα σε ένα συμπέρασμα και τα επικίνδυνα φάρμακα θα αποσύρονται άμεσα από την αγορά.



Οι υποσχέσεις για ένα “Μοναδικό” κόσμο μετά την υγειονομική περίθαλψη (2/2)

- Η εύκολη ροή της επιστημονικής γνώσης με την πρόσβαση σε τράπεζες πληροφοριών θα απαλλάξει τους γιατρούς από τα κλασικά εγχειρίδια και θα οδηγήσει σε πιο εξατομικευμένες και επιτυχείς θεραπείες.
- Η στιγμιαία ροή της γνώσης από τα ερευνητικά κέντρα στην πρακτική εφαρμογή της ιατρικής θα λειτουργήσει και αναδραστικά (πχ με την παραγωγή εμβολίων ή αντισωμάτων για νεοεμφανιζόμενες ασθένειες).



Πρόοδος προς την Μοναδικότητα της υγιεινής περίθαλψης

- Οι μεγάλες εταιρείες πληροφορικής δημιουργούν νέφη (cloud) δεδομένων από κλινικά στοιχεία ασθενών. Οι ίδιοι οι ασθενείς αποφασίζουν για το εύρος πρόσβασης σε αυτά τα στοιχεία.
 - Δημιουργούνται ειδικές πλατφόρμες που ομογενοποιούνε δεδομένα διάφορης προέλευσης και με αυτά τροφοδοτούν τα νέφη δεδομένων.
 - Τα νέφη δεδομένων απευθύνονται και στους ίδιους τους ασθενείς σε όλο τον κόσμο.
 - Με τα νέα πληροφορικά εργαλεία που δημιουργούνται δικτυώνονται ασθενείς, ερευνητές και πάροχοι υγείας.
 - Η άμεση μετάδοση της ιατρικής γνώσης δεν είναι μόνο πολύτιμη, επιθυμητή και εφικτή στις μέρες μας, αλλά ακόμη και αναπόφευκτη.
-



Παροχή υγειονομικής περίθαλψης στις αναπτυσσόμενες χώρες: Προκλήσεις και πιθανές λύσεις



Εισαγωγή

- Για τις αναπτυσσόμενες χώρες υπάρχουν αρκετά εμπόδια για την παροχή υγειονομικής περίθαλψης, όπως το οικονομικό κόστος, η έλλειψη εκπαιδευμένου υγειονομικού προσωπικού, το γενικότερο πρόβλημα της δύσκολης πρόσβασης σε βάσεις ιατρικών δεδομένων, γεωγραφικοί περιορισμοί κλπ.
 - Απλές πρακτικές που θα είχαν πολύ σημαντικά αποτελέσματα, όπως οι εμβολιασμοί, η φροντίδα μετά το τοκετό, η παροχή αποτελεσματικής περίθαλψης σε παιδιά με συμπτώματα πνευμονίας, προχωρούν με βραδύ ρυθμό.
 - Ο στόχος είναι η παροχή αξιόπιστων οδηγιών και η διάγνωση στο σημείο που αυτά χρειάζονται με αξιοποίηση εκπαιδευμένου προσωπικού.
 - Το εργαλείο για την επίτευξη του παραπάνω στόχου είναι ο συνδυασμός μίας ιατρικής πλατφόρμας της NxOpinion Knowledge Manager (NxKM) σε συνδυασμό με κινητά τηλέφωνα.
-



Χαρακτηριστικά της πλατφόρμας NxKM

- Η πλατφόρμα έχει την δυνατότητα να στηρίζει (scale up) μέχρι και εκατομμύρια χρήστες και αυτό έχει επαληθευτεί στην Ινδία, στο Κογκό και αλλού.
 - Έχει τη δυνατότητα να εργάζεται σε πολλές χώρες και για πολλές ασθένειες (scale out) και μπορεί να εξειδικεύει ανά περιοχή καθώς και να παρέχει εξατομικευμένη διάγνωση (scale in) αφού περιλαμβάνει πάνω από 1000 ασθένειες με πάνω από 6000 συμπτώματα.
 - Έχει αξιόλογη ικανότητα να κάνει επιτυχημένες διαγνώσεις. Μπορεί να ρωτά επιπλέον ερωτήσεις για να κάνει πιο εντοπισμένη διάγνωση. Μπορεί να εντοπίζει την πιθανότερη διάγνωση.
 - Όλες αυτές οι δυνατότητες είναι πολύ χρήσιμες για λιγότερο έμπειρους χρήστες του συστήματος που χρησιμοποιούνται σε υπό ανάπτυξη χώρες, όπως οι health extension workers (HEWs).
-



Δίνοντας υγειονομική φροντίδα εκεί που χρειάζεται- Το τελευταίο μίλι

- Τα κινητά τηλέφωνα είναι το κατάλληλο όχημα για την σύνδεση της πλατφόρμας με τους τελικούς χρήστες με τη μεσολάβηση των HEWs αφού η χρήση των κινητών στις υπό ανάπτυξη χώρες είναι ήδη πολύ εκτεταμένη και διευρύνεται ταχύτατα.
- Επίσης με το σύστημα αυτό μπορεί να δημιουργηθούν ατομικοί φάκελλοι υγείας σε ηλεκτρονική μορφή.
- Ο τρόπος αυτός συλλογής στοιχείων μπορεί να λειτουργήσει αναδραστικά προς την κατεύθυνση της βελτίωσης της πλατφόρμας ιδιαίτερα κατά την εκδήλωση επιδημιών ή πανδημιών όπου η ανταπόκριση πρέπει να είναι άμεση.
- Το πρόβλημα της μικρής δυνατότητας αποστολής στοιχείων μέσω κινητών αντιμετωπίζεται με πολλούς τρόπους (πχ τα γενικά στοιχεία έχουν περαστεί εκ των προτέρων στο σύστημα, γίνεται χρήση ερωτήσεων πολλαπλής επιλογής κλπ).
- Επίσης αντιμετωπίζονται τα προβλήματα που οι διαφορετικοί τύποι κινητών μπορεί να δημιουργήσουν.



Δυνατότητες και προοπτικές για το μέλλον

- Είναι σημαντική η δυνατότητα διαμοιρασμού δεδομένων που υπάρχουν σε διάφορους φορείς, όπως ασφαλιστικές εταιρείες, φαρμακευτικές εταιρείες, δημόσιοι οργανισμοί κλπ, πρακτική που ακολουθεί η πλατφόρμα NxKM
 - Απομένουν ακόμη σημαντικά πράγματα να γίνουν, ώστε να βελτιωθεί η αποτελεσματικότητα και η αξιοπιστία της πλατφόρμας ιδίως στο διαγνωστικό τομέα.
 - Η εκπαίδευση και η ενημέρωση των HEWs πρέπει να γίνει πιο μεθοδικά.
 - Πρέπει να βελτιωθεί η πιστοποίηση του ασθενούς που λαμβάνει την θεραπεία και αυτό μπορεί να γίνει με αναγνώριση φωνής ή προσώπου ή με τη χρήση GPS.
 - Η διαχείριση και αξιοποίηση του τεράστιου όγκου δεδομένων που συλλέγονται με παράλληλη τήρηση των αρχών σχετικά με τα προσωπικά δεδομένα είναι μία εξαιρετικά προκλητική προοπτική.
 - Τέλος πολύ σημαντικό θέμα είναι η συνεργασία οργανισμών, εταιρειών και κρατικών δομών. Η δημιουργία μίας παγκόσμιας βάσης δεδομένων θα βοηθήσει σημαντικά, τόσο την παροχή υγειονομικής περίθαλψης, όσο και τη σχετική έρευνα.
-



Ανακαλύπτοντας το διάγραμμα καλωδίωσης του εγκεφάλου

Jeff W. Lichtman r. Clay Reidd Hasspeter Pfistelr Harvard University Michael F. Coehn Microsoft Research



Εισαγωγή (1/2)

- Η κατανόηση του τρόπου κατασκευής και λειτουργίας του εγκεφάλου παραμένει ακόμη ένα μυστήριο.
 - Τα νευρικά κύτταρα του εγκεφάλου συνδέονται μεταξύ τους σε συνάψεις με κυκλώματα απεριόριστης πολυπλοκότητας και υποθέτουμε ότι αυτό συνδέεται με τη διαδικασία μάθησης απλών είτε εξαιρετικά πολύπλοκων εννοιών.
 - Τόσο η δομή όσο και η λειτουργία του εγκεφάλου αποτελούν περίπλοκες υποθέσεις. Έτσι ο εγκέφαλος αποτελείται από δισεκατομμύρια νευρώνες και συνολικά οι συνάψεις ξεπερνούν τα 160 τρισεκατομμύρια με πυκνότητα 300 εκατομμύρια ανά κυβικό χιλιοστό. Κάθε νευρώνας παίρνει μηνύματα από εκατοντάδες άλλους και τα διαβιβάζει ομοίως.
 - Από τα παραπάνω φαίνεται ότι η κατασκευή των συναπτικών διαγραμμάτων είναι τεχνικά εξαιρετικά δύσκολη υπόθεση.
-



Εισαγωγή (2/2)

- Η λειτουργική μαγνητική τομογραφία έδωσε κάποια τρισδιάστατη εικόνα της λειτουργίας του εγκεφάλου μέχρι το επίπεδο κυβικού χιλιοστού που προφανώς είναι ανεπαρκής.
- Η μόνη μεθοδολογία που μπορεί να χρησιμοποιηθεί για τη μελέτη των συναπτικών συνδέσεων είναι το ηλεκτρονικό μικροσκόπιο (EM) με ανάλυση μέχρι nm, επαρκή για την περίπτωση, με λήψη τομών πάχους δεκάτων του nm που προστιθέμενες στη συνέχεια μπορούν να ανασυνθέσουν μεγαλύτερους όγκους.
- Ένα κυβικό χιλιοστό του ανθρώπινου εγκεφάλου θα απαιτούσε για να απεικονιστεί 1 peta bytes (10 εις την δεκάτη πέμπτη) και ο ανθρώπινος εγκέφαλος έχει όγκο ένα εκατομμύριο κυβικά χιλιοστά.



Κάποιες επιτυχίες σήμερα

- Η περιγραφείσα πολυπλοκότητα του εγκεφάλου δεν σημαίνει ότι δεν γίνονται προσπάθειες χάρις στις νέες τεχνικές και τεχνολογίες με τη συνεργασία βιολόγων, μηχανικών, και επιστημόνων πληροφορικής.
- Ήδη έχουν ξεκινήσει προσπάθειες με EM και σύντομα θα καταστεί δυνατή η πλήρης απεικόνιση συνάψεων σε ένα κύβο 300 μm με τα δεδομένα να ανέρχονται σε 10 terabytes.
- Παραμένει βέβαια το πρόβλημα της ιχνηλάτησης μεμονωμένων νευρώνων. Αυτό γίνεται προσπάθεια να αντιμετωπισθεί με κάποιες φωσφορίζουσες πρωτεΐνες σε γενετικά τροποποιημένο ποντίκι παράγοντας έτσι 90 χρώματα ή συνδυασμούς χρωμάτων που απεικονίζουν τους νευρώνες και τις συνάψεις.
- Η εργασία με EM προϋποθέτει λήψεις σε πολύ λεπτές φέτες (σε nano μεγέθη) και στη συνέχεια ενοποίηση των λήψεων. Η εργασία επιταχύνεται πολύ με μία επινόηση -ALTUS- που αυτοματοποιεί τη διαδικασία.
- Η έρευνα συνεχίζεται και πολλές από τις δυσκολίες αντιμετωπίζονται με την πρόοδο που σημειώνεται τόσο στη βιολογία όσο και στην τεχνολογία.



Προς ένα Υπολογιστικό μικροσκόπιο για τη νευροβιολογία

Eric Horvits Microsoft Research William Kristan University of California, San Diego



Εισαγωγή

Παρά τις προόδους στην νευροβιολογία δεν έχει γίνει ακόμα κατανοητός ο τρόπος αλληλεπίδρασης των νευρώνων, ώστε να προέλθουν οι ζωικές λειτουργίες της μάθησης, της απόφασης, της ενέργειας. Ελπίδες προς την κατεύθυνση δίνουν οι υπολογιστικές μέθοδοι και η μηχανική μάθηση με τη δημιουργία επεξηγηματικών μοντέλων.



Νέα εργαλεία για νευροβιολόγους (1/2)

- Οι νευροβιολόγοι μέχρι τώρα προκειμένου να μελετήσουν τις ιδιότητες της μεμβράνης των νευρώνων εργάστηκαν με μικροσκοπικά ηλεκτρόδια υάλου που τα εισήγαγαν εντός των κυττάρων Έτσι βρήκαν δεδομένα σχετικά με το δυναμικό στη μεμβράνη μεμονωμένων κυττάρων ή λίγων κυττάρων.
- Ήδη αυτό αλλάζει με τη χρήση εξοπλισμού που μπορεί να καταγράφει τη δραστηριότητα μεγάλου αριθμού νευρώνων, που με την αξιοποίηση υπολογιστικών διαδικασιών μπορεί να γίνει κατανοητή η αρχιτεκτονική, η λειτουργία και όλος ο μηχανισμός του νευρικού συστήματος.



Νέα εργαλεία για νευροβιολόγους (2/2)

- Χρησιμοποιούνται γρήγορες απεικονιστικές μέθοδοι με χρήση βαφών και φωτοπολλαπλασιαστών για να ιχνηλατήσουν τα επίπεδα καλίου και τα δυναμικά των μεμβρανών των νευρώνων.
 - Έτσι γίνεται εφικτή η ταυτόχρονη μελέτη μεγάλου αριθμού νευρώνων σε μεγάλη έκταση χωροχρονικά και αυτό εκτιμάται ότι θα αποδώσει σημαντικά αποτελέσματα.
 - Εκτιμάται ότι στο μέλλον οι νευροβιολόγοι θα αξιοποιήσουν εργαλεία που να λειτουργούν σαν υπολογιστικά μικροσκόπια τα οποία θα αξιοποιούν την δυνατότητα μηχανικής μάθησης για να τεστάρουν υποθέσεις.
 - Εκτεταμένη υπολογιστική ανάλυση θα αξιοποιηθεί για μοντελοποίηση και οπτικοποίηση των τεράστιων αριθμών δεδομένων που θα προκύπτει από την αλληλεπίδραση των νευρώνων, των συνάψεων και των κυκλωμάτων.
-



Υπολογιστική και πολυπλοκότητα

- Υπάρχουν πολλές δυνατότητες αξιοποίησης της υπολογιστικής δυναμικής από τους νευροβιολόγους:
 - Η στατιστική ανάλυση.
 - Η αξιοποίηση της μηχανικής μάθησης.
 - Η χρήση υπολογιστικών μεθόδων για τη διερεύνηση μεγάλων αριθμών δεδομένων.
 - Η εξαγωγή συμπερασμάτων για την εφαρμοσιμότητα διαφόρων μοντέλων.
- Με τη χρήση υπολογιστικών εργαλείων μπορούν οι νευροβιολόγοι να βοηθηθούν στην καθοδήγηση της έρευνάς τους.



Το Φάσμα της επιτήδευσης

- Η μελέτη των νευρικών συστημάτων διαφόρων ζώων θα βοηθήσει στη μελέτη του νευρικού συστήματος κατά την εμβρυογένεση καθώς και στην σύγκριση των νευρικών συστημάτων διαφόρων ειδών.
 - Οι νευροβιολογοί εξετάζουν το νευρικό σύστημα ασπόνδυλων και σπονδυλωτών μέσα σε ένα εύρος πολυπλοκότητας.
 - Ο ανθρώπινος εγκέφαλος έχει 100 δισεκατομμύρια νευρώνες και ενεργοποιούνται 100 τρισεκατομμύρια συνάψεις. Ο εγκέφαλος του C.ELEGANT έχει 302 νευρώνες μόνο ενώ της ιατρικής βδέλλας μόνο 400 νευρώνες.
 - Οι νευροβιολόγοι πιστεύουν ότι η αποκωδικοποίηση της λειτουργίας των νευρώνων σε απλούς οργανισμούς θα βοηθήσει στην κατανόηση πιο σύνθετων συστημάτων, όπως του ανθρώπου, αφού εκτιμούν ότι κάποιες λειτουργικές αρχές τηρούνται σε όλους τους τύπους.
 - Αναπτύσσονται αναλυτικές και συνθετικές μέθοδοι προς την κατεύθυνση της υλοποίησης του οράματος του ηλεκτρονικού μικροσκοπίου που θα μπορέσει να μελετήσει τη δραστηριότητα των νευρώνων και τη σχέση τους με τη συμπεριφορά.
-



Η Ανάδυση του ηλεκτρονικού Μικροσκοπίου

- Ξεκίνησε η δημιουργία εργαλείων και παρατηρητών που αλληλεπιδρούν και επιτρέπουν στους επιστήμονες να αλλάζουν τις υποθέσεις και τις παραμέτρους και να βλέπουν οπτικοποιημένα τα αποτελέσματα.
 - Αναπτύχθηκε το MSR ηλεκτρονικό μικροσκόπιο με το οποίο οι χρήστες οπτικοποιούν την λειτουργία των νευρώνων για μία χρονική περίοδο, μπορούν να αλλάζουν τις παραμέτρους λειτουργίας, να επισημαίνουν συγκεκριμένα σύνολα νευρώνων κλπ.
 - Βρίσκεται σε αρχικό στάδιο η διαδικασία και θα πρέπει να συνεργαστούν πολλοί επιστήμονες, να βρεθούν καινούργια εργαλεία, ώστε οι υπολογιστικές μέθοδοι να βοηθήσουν πραγματικά τους νευροβιολόγους να ρίξουν φως σε ένα σκοτεινό ως τώρα πεδίο.
-



Μία προσέγγιση σε ένα ενοποιημένο μοντέλο πυκνής σε δεδομένα υγειονομικής περίθαλψης

Iain Buchan University of Manchester John Winn Chris Bishop Microsoft Research



Εισαγωγή

- Η ραγδαία αύξηση των δεδομένων υγειονομικής περίθαλψης επηρεάζεται από τρεις παράγοντες που εξετάζονται παρακάτω και οδηγούν στην πρότασή μας για ένα ενοποιημένο μοντέλο που θα μπορεί να τα αξιοποιήσει.
 - Τα ηλεκτρονικά αρχεία υγείας (EHRs).
 - Η βιοτεχνολογία.
 - Τα αποτελέσματα της επιστήμης.



Ηλεκτρονικά αρχεία υγείας (EHRs)

- Τα EHRs χρησιμοποιούνται ήδη σε αρκετή έκταση και βοηθούν πέραν του ατομικού επιπέδου στη διαχείριση της δημόσιας υγείας, στην υποστήριξη της έρευνας κλπ.
- Η ενθάρρυνση της χρήσης EHRs θέλει προσοχή, γιατί μπορεί να αυξήσει στατιστικά κάποια δεδομένα, που απλώς δεν είχαν καταγραφεί πριν.
- Η διεύρυνση της χρήση EHRs οδηγεί στην συγκέντρωση τεράστιου αριθμού δεδομένου, που για να διαχειριστεί και να αξιοποιηθεί με επιτυχία πρέπει να αναπτυχθούν τα κατάλληλα υπολογιστικά εργαλεία.



Βιοτεχνολογία

- Η βιοτεχνολογία έχει πυροδοτήσει μία έκρηξη στην ιατρική έρευνα και στην παραγωγή δεδομένων, όπως η μελέτη του γονιδιώματος.
 - Η παραγωγή δεδομένων στο νέο πλαίσιο ερευνών είναι ευρύτερης γκάμας και λιγότερο επιλεκτική από τις συμβατικές πειραματικές μεθόδους.
 - Η ενσωμάτωση διαφόρου κλίμακας δεδομένων από μοριακό επίπεδο σε επίπεδο πληθυσμού αποτελεί μία πρόκληση.
 - Εφόσον οι παραπάνω προϋποθέσεις πραγματοποιηθούν το επόμενο βήμα είναι τα δεδομένα να γίνουν προσβάσιμα σε γιατρούς και ασθενείς για προσωπική χρήση.
-

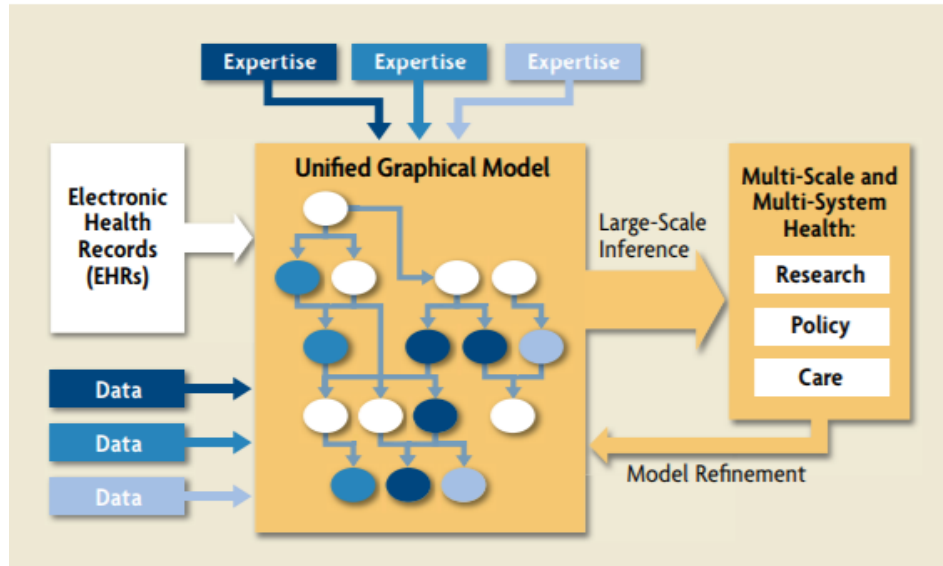


Επιστημονικά αποτελέσματα

- Υπάρχει υπερπαραγωγή ιατρικών επιστημονικών δεδομένων και όχι μόνο η παρακολούθησή τους είναι δύσκολη, αλλά και η αξιοποίησή τους είναι δυσχερής και αυτό επηρεάζει αρνητικά εκτός από τους γιατρούς και τους ασθενείς και τους ερευνητές.
 - Υπάρχει ανάγκη να συνδυαστεί η εξόρυξη ενδείξεων με υπολογιστικά μοντέλα για την εξερεύνηση δεδομένων περίθαλψης και έρευνας.
 - Η ως τώρα τάση της ιατρικής έρευνας να προσδιορίζει τον κύριο παράγοντα που επιδρά σε ένα θέμα υγείας λειτούργησε ικανοποιητικά.
 - Όμως επειδή τα θέματα υγείας είναι σύνθετα μπορεί να υπάρξει, μία πιο σύνθετη προσέγγιση, με κατάλληλη υπολογιστική αξιοποίηση της πληθώρας των δεδομένων που ήδη διατίθενται.
-



Ένα Ενοποιημένο μοντέλο (1/2)



Προτείνεται ένα ενοποιημένο μοντέλο που αξιοποιεί το πλήθος των δεδομένων χωρίς να χάνει την πραγματική πολυπλοκότητα της υγείας.



Ένα Ενοποιημένο μοντέλο (2/2)

- Η προσέγγισή βασίζεται στην μηχανική μάθηση με εμπλουτισμό με νέα εργαλεία κατάλληλα για την περίπτωση. Για να γίνει κατανοητός ο καινοτομικός χαρακτήρας της προσέγγισης λαμβάνονται δύο σενάρια:
 - Στο σενάριο ένα συμβατικός επιδημιολόγος εξετάζει-ορίζει τους γενετικούς και περιβαλλοντικούς παράγοντες που μπορεί να επηρεάσουν την εμφάνιση άσθματος σε παιδιά. Με τη νέα μηχανικής μάθησης προσέγγιση δεν προκαθορίζονται οι παράγοντες, αλλά ο επιδημιολόγος ξεκινά κάνοντας ένα πρότυπο γράφημα του προβλήματος συμβουλευτεί τη βιβλιογραφία και συναδέλφους του για τον εμπλουτισμό του μοντέλου και καταλήγει σε ένα πιο ρεαλιστικό σχήμα.
 - Στο σενάριο δύο αξιολογούνται τα αποτελέσματα θεραπείας της χρόνιας στηθάγχης. Η κλασική μεθοδολογία περιλαμβάνει την εξέταση της θεραπείας που δόθηκε, της σχετικής βιβλιογραφίας, την εξέταση των στατιστικών δεδομένων, ώστε να δημιουργηθεί ένα σενάριο και να ελεγχθεί η βασιμότητά του. Η εναλλακτική προσέγγιση μηχανικής μάθησης θα έχτιζε ένα γραφικό μοντέλο της υποτιθέμενης θεραπείας, μέσω διαβούλευσης, αναφοράς στη βιβλιογραφία και θα συνέκρινε αυτό με τα αποτελέσματα από τους ασθενείς.



Το τέταρτο πρότυπο, προσωποποίηση υγείας

- Τα ενοποιημένα μοντέλα σίγουρα θα επηρεάσουν τα θέματα υγείας στο μέλλον. Είναι λοιπόν το τέταρτο πρότυπο;
 - Το πρώτο πρότυπο θεωρείται αυτό που καθιέρωσε ο Ιπποκράτης πριν 2000 χρόνια. Στο δεύτερο πρότυπο τα ιατρικά δεδομένα διαμοιράζονται μεταξύ μίας ομάδας κλινικών που ο καθένας τα κοιτάει από τη σκοπιά του. Στο τρίτο πρότυπο η περίθαλψη βασίζεται σε ενδείξεις, που συνδέουν επαγγελματίες της υγείας και ιατρικά αρχεία. Το τελευταίο στάδιο είναι εν εξελίξει με ψηφιοποίηση των αρχείων και αξιοποίηση των υπολογιστών.
 - Προσδοκείται το τέταρτο πρότυπο ιατρικών αρχείων που συγκεντρώνουν στοιχεία από πολλές πηγές με έντονη συμμετοχή του προσώπου που αφορά, που θα αποτελέσει την ηλεκτρονική παράσταση της υγείας ενός ατόμου, το “avatar” υγείας που θα είναι ένα στατιστικό εργαλείο για τη διάγνωση και τη θεραπεία, αλλά και τη σύνδεση με τα δίκτυα γιατρών και άλλων ειδικών.
 - Τα “avatar” αυτά μπορούν να επεκταθούν σε επίπεδο κοινότητας για προστασία της υγείας σε συλλογικό επίπεδο.
-



Οπτικοποίηση σε αλγεβρικά διαδικαστικά μοντέλα βιολογικών συστημάτων

Luca Cardelli Microsoft Research

Corrado Priami Microsoft Research - University of Trento Centre for Computational and Systems Biology



Εισαγωγή

- Κορυφαίοι βιολόγοι κάνουν λόγο για την ανάγκη ανάπτυξης κατάλληλης γλώσσας για την περιγραφή των βιολογικών φαινομένων. Αυτή η γλώσσα θα μπορεί να χρησιμοποιεί μηχανιστικούς κώδικες και να κάνει προσομοιώσεις των βιολογικών συστημάτων.
 - Η επιστήμη των υπολογιστών μπορεί να κάνει τις παραπάνω λειτουργίες με τους αλγόριθμους και τις γλώσσες προγραμματισμού.
 - Οι υπολογιστικές διαδικασίες δίνουν ελπίδες για την μελέτη των βιολογικών συστημάτων που από τη φύση τους είναι εξαιρετικά περίπλοκα.
 - Η αιτιότητα (εύρεση της αιτίας) είναι πολύ σημαντικός παράγοντας σε σύνθετα αλληλεπιδρώντα συστήματα και αυτό συμβαίνει ιδιαίτερα στα βιολογικά συστήματα.
 - Η αιτιότητα μπορεί κάλλιστα να μελετηθεί με αλγοριθμικές γλώσσες σε συνδυασμό με μαθηματικά μοντέλα.
-



Οπτικοποίηση της γλώσσας

- Η καθιέρωση μίας τυπικής γλώσσας στην βιολογία πρέπει να είναι εύχρηστη από βιολόγους και ένας τρόπος να το κάνουμε αυτό είναι να οπτικοποιηθεί, πράγμα δυσκολότερο από το να οπτικοποιηθούν δεδομένα ή αποτελέσματα προσομοιώσεων.
- Η οπτικοποίηση πρέπει να επιτρέπει παρεμβάσεις με αλλαγή παραμέτρων, ώστε να ελέγχονται διάφορα σενάρια.
- Έχουν γίνει προσπάθειες προς αυτήν την κατεύθυνση, αλλά πολλά προβλήματα μένει ακόμη να λυθούν.



Ανάλυση

- Πέραν της κατασκευής του μοντέλου πολύ χρήσιμη είναι και η ανάλυση του μοντέλου με τη χρήση πολλών τεχνικών.
- Μερικές τέτοιες τεχνικές μπορεί να σχετίζονται με το μαθηματικό σκελετό του μοντέλου, όπως οι διαφορικές εξισώσεις, άλλες με την περιγραφή του μοντέλου.
- Η βιολογία εξετάζει τις σχέσεις γονιδίων, γονιδιώματος και πρωτεϊνών. Οι σχέσεις που προκύπτουν είναι τεράστιες σε όγκο και κάνουν τη βιολογία μία επιστήμη με μεγάλη πυκνότητα δεδομένων.
- Οι έρευνες σε μοριακό επίπεδο έδωσαν ακόμη μεγαλύτερο αριθμό δεδομένων που για γίνει εφικτή η αποτελεσματική διαχείριση, αξιοποίηση και αρχειοθέτηση τους, χρειάζεται η ανάπτυξη κατάλληλων μοντέλων, που θα βοηθήσουν σημαντικά την βιολογία στα επόμενα βήματά της



Οπτικοποίηση της Ανάλυσης

- Μοντέλα που μπορούν να εκτελούνται χρειάζονται οπτικοποίηση, ώστε να είναι διαδραστικά και αντιδραστικά.
- Εκτός της εκτέλεσης και άλλες διαδικασίες χρειάζονται οπτικοποίηση, όπως η ανάλυση σε ανεξάρτητες παρτίδες πρέπει να αντικατασταθεί από πιο διαδραστικές και εξερευνητικές διαδικασίες.
- Όλα τα παραπάνω ενισχύουν την ανάγκη για μία τυπική και εκτελέσιμη γλώσσα που θα μοντελοποιήσει τη βιολογία και θα καθιερώσει ένα νέο τρόπο έρευνας, *in silico*, δηλαδή μέσω υπολογιστή, κατά αναλογία του *in vitro* /*in vivo* (σε δοκιμαστικό σωλήνα /σε ζώντα οργανισμό), που μπορεί να αποτελέσει τη νέα εποχή για τη βιολογία.



ΕΠΙΣΤΗΜΟΝΙΚΗ ΥΠΟΔΟΜΗ



Εισαγωγή

- Στο κεφάλαιο αυτό περιγράφεται με δραματικό τρόπο η επίδραση της υπερπαραγωγής δεδομένων στην επιστημονική έρευνα.
 - Εξετάζονται θέματα όπως:
 - Αν στις νέες συνθήκες θα επικρατήσει η ανάγκη για διαμοιρασμό δεδομένων και συνεργασία.
 - Αν η παραγωγή επιστημονικής γνώσης θα είναι αντίστοιχη της ζήτησης από τον ιδιωτικό και το δημόσιο τομέα.
 - Η έκρηξη γενετικών δεδομένων.
 - Η σημασία των πολυπύρηνων επεξεργαστών.
 - Η σημασία υπολογιστικών μοντέλων ροής εργασίας δια τη διαχείριση της έκρηξης δεδομένων.
 - Η ανάγκη οπτικοποίησης.
 - Πώς όλα αυτά θα συντελέσουν στο να δημιουργηθεί μία σφριγηλή υποδομή για την επιστημονική έρευνα.
-



Ένας νέος δρόμος για την επιστήμη;

Mark R. Abbot Oregon State University



Η προοπτική της επιστημονικής έρευνας

- Η επιστημονική έρευνα έχει βέβαια παραγάγει πολύ σημαντικά αποτελέσματα φτάνει όμως σε κάποια κρίσιμα σημεία, όπου κάποια πράγματα θα πρέπει να επανεξεταστούν.
- Τέτοια θέματα είναι η εσωστρέφεια σε κάθε επιστημονικό πεδίο, η αναντιστοιχία της παραγόμενης γνώσης με τη ζήτηση από το δημόσιο και τον ιδιωτικό τομέα, η αξιολόγηση της επιστημονικής έρευνας στα επιστημονικά ιδρύματα από την ακαδημαϊκή κοινότητα, χωρίς να λαμβάνεται υπόψη η γνώμη της υπόλοιπης κοινωνίας κλπ.
- Πλέον θα πρέπει η επιστημονική έρευνα να δίνει συγκεκριμένες απαντήσεις σε κρίσιμα κοινωνικοοικονομικά ερωτήματα που θέτει, είτε η επιχειρηματική κοινότητα, είτε οι πολιτικοί. Τέτοια θέματα είναι πχ τα θέματα της κλιματικής αλλαγής και της αειφορίας.
- Η βασική επιστημονική έρευνα θα συνεχιστεί σε αυτούς τους τομείς, αλλά τα πρακτικά ερωτήματα που πρέπει να απαντηθούν είναι διεπιστημονικά και εμπεριέχουν την εκτίμηση κινδύνου και ακόμη την πιθανότητα για έκτακτα γεγονότα και την εκτίμηση των συνεπειών τους.



Πρόσφατη ιστορία

- Η τεχνολογία πληροφορικής της τελευταίας 25ετίας στήριζε μία αντίληψη έρευνας όπου η ροή εργασίας περιλάμβανε τη συλλογή δεδομένων, την αποθήκευση, επεξεργασία και ανάλυσή τους, την οπτικοποίησή τους και τη δημοσίευσή τους.
 - Χαρακτηριστικό παράδειγμα είναι το πληροφοριακό σύστημα της NASA (EOSDIS) που δημιουργήθηκε στο τέλος της δεκαετίας του 80. Είχε σχεδιαστεί με μεγάλες δυνατότητες για να αποτελέσει ένα εργοστάσιο δεδομένων και ο προβληματισμός ήταν, αν θα ήταν συγκεντρωτικού ή αποκεντρωτικού χαρακτήρα Στο τέλος το πραγματικό ερώτημα ήταν ποιος θα είχε το πρώτο λόγο, η επιστημονική κοινότητα ή η NASA - εργολάβος.
-



Σήμερα

- Παρά τις δυνατότητες υπερυπολογιστών και της υπολογιστικής νέφους η λογική παραμένει η ίδια όπως στο EOSDIS.
- Ο νόμος του Moore (διπλασιασμός τρανζιστορς ανά διετία) εφαρμόστηκε επί 25 χρόνια από την καθιέρωση του EOSDIS. Ο ραγδαίος ρυθμός παραγωγής δεδομένων πιέζει για την μεταβολή του παραδοσιακού τρόπου διεξαγωγής της έρευνας.
- Το μεμονωμένο δεδομένο δεν έχει πλέον χρησιμότητα, και η αναφορά σε πυκνότητα δεδομένων δεν σχετίζεται μόνο σε αποθηκευμένα δεδομένα, αλλά έχει να κάνει και με δεδομένα σε δίκτυα αισθητήρων, αποτελέσματα έρευνας σε εργαστήρια κλπ.
- Η πολυπλοκότητα των νέων συνόλων δεδομένων και η διαφορετικότητα της ροής δεδομένων κάνει το παραδοσιακό σχήμα υπολογιστικής ανεπαρκές για τη σύγχρονη έρευνα.



Επιπτώσεις για την επιστήμη (1/2)

- Η επιστήμη της πληροφορικής και των τηλεπικοινωνιών (IT) έχει επηρεάσει την έρευνα με δύο τρόπους.
- Με την εμπορευματοποίηση της υπολογιστικής και της αποθήκευσης. Έτσι αξιοποιούνται υπηρεσίες που έχουν μεγάλο δυναμικό και μικρό περιθώριο κέρδους για back up, email και άλλες εφαρμογές, όπως η υπολογιστική νέφος.
- Με δραστική εξατομίκευση. Ο επιστήμονας έχοντας διαθεσιμότητα σε τεράστια μεγέθη αποθήκευσης και υπολογιστικής μπορεί να φτιάξει το δικό του νέφος. Δημιουργείται έτσι ένα δίκτυο κατανεμημένης γνώσης με τεράστιες δυνατότητες.
- Σήμερα η επιστημονική ανακάλυψη δεν συντελείται μέσω της επαλήθευσης υποθέσεων.



Επιπτώσεις για την επιστήμη (2/2)

- Οι τεράστιοι όγκοι δεδομένων, οι περίπλοκες και δυσνόητες σχέσεις, οι έντονες και μεταβαλλόμενες μορφές συνεργασίας μεταξύ των επιστημονικών πεδίων και οι νέοι τύποι εκδόσεων σχεδόν σε πραγματικό χρόνο, διαμορφώνουν τη νέα επιστημονική μέθοδο και αυτό είναι ιδιαίτερα αισθητό στο τομέα της κλιματικής αλλαγής.
 - Στο μέλλον οι αλλαγές σε σχέση με το σήμερα θα είναι τεράστιες. Το νέο σχήμα έρευνας θα είναι διεπιστημονικό, θα λειτουργεί σε δίκτυα συνεργασίας και διαμοιρασμού δεδομένων, θα είναι λιγότερο προβλέψιμο και οριοθετημένο, θα είναι πιο ευέλικτο και θα επιτρέπει την γρήγορη καινοτομία.
 - Το νέο σχήμα θα έρθει σε αντίθεση με τη σημερινή δομή της έρευνας από άποψη υλοποίησής της, χρηματοδότησης, κοινωνικού οφέλους.
 - Το νέο σχήμα θα ενώσει ταλαντούχους ανθρώπους από όλο τον κόσμο και θα μπορέσει να δώσει ολοκληρωμένες βιώσιμες λύσεις σε πολλά προβλήματα, όπως αυτό της κλιματικής αλλαγής.
-



Πώς να γίνει πραγματικότητα

- Με τα σημερινά δεδομένα η τεχνολογία των βάσεων δεδομένων δεν αξιοποιεί την καινοτομία.
- Σήμερα σημασιολογικοί ιστοί και οι οντολογίες προτείνονται σαν εργαλεία ανακάλυψης γνώσης και συνεργασίας, αλλά λόγω της εγγενούς πολυπλοκότητά τους είναι αμφίβολο αν θα υιοθετηθούν από την επιστημονική κοινότητα.
- Νέα εργαλεία IT μπορούν να φιλτράρουν τα σημερινά ογκώδη δεδομένα σε ένα διαχειρίσιμο επίπεδο και να παρέχουν οπτικοποίηση, ώστε να γίνουν πιο κατανοητά και να κτιστούν συνεργασίες.
- Η αρχιτεκτονική για τη υπολογιστική υψηλής πυκνότητας δεδομένων θα πρέπει να είναι αποκεντρωτική σε κάθε κόμβο του συστήματος, θα επιτρέπει την καινοτομία στα άκρα του συστήματος και θα μπορέσει να αποτελέσει ένα “οικοσύστημα” γνώσης.



Πέρα από το “τσουνάμι”: Αναπτύσσοντας την υποδομή για την ανταπόκριση στα δεδομένα των επιστημών της ζωής

Christopher Southan Graham Cameron EMBL-European Bioinformatics Institute



Εισαγωγή

- Η έκρηξη της παραγωγής δεδομένων στις επιστήμες ζωής μπορεί να γίνει αντιληπτή από την περίπτωση του European Molecular Biology Laboratory Nucleotide Sequence Database1 (EMBL-Bank) όπου η υποβολή ατομικών βάσεων από τη παγκόσμια επιστημονική κοινότητα αυξάνεται κατά 200% ετησίως.
 - Η EMBL-Bank φιλοξενείται στο μέλος της European Bioinformatics Institute (EBI), που εδράζεται στο Cambridge, UK και είναι ένα κέντρο έρευνας και παροχής υπηρεσιών βιοπληροφορικής όπου συνεργάζονται, βιοπληροφορικοί, ερευνητές βιολόγοι, αναλυτές βιολόγοι και όπου θέματα όπως η εξόρυξη δεδομένων, η αποθήκευση δεδομένων και τα σχετικά αντιμετωπίζονται σε καθημερινή βάση.
-



Η πρόκληση της αλληλούχισης της επόμενης γενιάς

- Η καθιέρωση από το 2005 νέων εργαλείων αποτύπωσης της αλληλουχίας έδωσε τρομακτική ώθηση στην παραγωγή γενετικών πληροφοριών και έφερε την βιοπληροφορική στην πρώτη γραμμή της ανάπτυξης υποδομών για την αποθήκευση, ανάλυση, ερμηνεία και οπτικοποίηση δεδομένων της τάξης των petabytes.
 - Ο ρυθμός κατάθεσης δεδομένων ανέβηκε στην τάξη των 30 terabytes για το πρώτο εξάμηνο, σχεδόν στο 30% του συνολικού περιεχομένου της EMBL-Bank.
 - Αντίστοιχα το EBI για να ανταποκριθεί ανέβασε την αποθηκευτική του ικανότητα στα 5000 terabytes (5peta) το 2009 και συνεχίζει να αναζητά τρόπους για να ανταποκριθεί στην εκθετικά αυξανόμενη ζήτηση.
 - Η απόρριψη επεξεργασθέντων πρωτογενών δεδομένων είναι μία δραστική λύση που θα μειώσει αποτελεσματικά τις απαιτήσεις, παρά το ότι είναι κάτι που οι ερευνητές γενικά δεν το θέλουν.
-



Η επιτάχυνση της αλληλούχησης γονιδιωμάτων

- Η επιτάχυνση της παράστασης της αλληλουχίας γονιδιωμάτων βελτιώνει την ικανότητά μας να καταλάβουμε τη βιολογία και την εξέλιξη.
 - Πολλά γονιδιώματα έχουν ήδη αποκωδικοποιηθεί και η πρόοδος είναι ταχύτατη.
 - Εκτιμάται ότι όλα τα γονιδιώματα των κατοικίδιων ζώων θα έχουν σύντομα αποκωδικοποιηθεί και αυτό θα βοηθήσει στην κατανόηση της εξέλιξής τους, αλλά και θα επιταχύνει την γενετική βελτίωση στην εκτροφή των ζώων και την παραγωγή τροφίμων.
 - Αντίστοιχες μελέτες σε ανθρώπους βελτίωσαν την αντίληψή μας για διάφορες ασθένειες.
 - Αυτό ενθάρρυνε τις προσπάθειες για την εύρεση της αλληλουχίας σε τουλάχιστον 1000 ανθρώπινα γονιδιώματα και αυτό οδηγεί σε τεράστια παραγωγή δεδομένων που θα ανέλθουν σε 500 terabytes για το συνολικό πρόγραμμα.
-



Ο λόγος της διαχείρισης δεδομένων είναι η μετατροπή τους σε νέα γνώση

- Ήδη από καιρό διαπιστώνεται ότι στις επιστήμες ζωής η ικανότητά μας για εξόρυξη δεδομένων είναι πίσω από την ικανότητά μας να τα δημιουργούμε.
 - Έτσι δημιουργείται πιεστική ανάγκη για νέες μεθόδους που θα επιτρέψουν την πλήρη αξιοποίηση γενετικών πληροφοριών που αποθηκεύονται σε βάσεις δεδομένων.
 - Η πρόσβαση στο EBI από την βιολογική κοινότητα αυξάνει εντυπωσιακά με 3.5 εκ επισκέψεις την ημέρα που αντιστοιχούν σε περισσότερους από μισό εκ ανεξάρτητους χρήστες το μήνα.
 - Η διαφαινόμενη αύξηση στην εξόρυξη δεδομένων ενισχύεται από το EBI με την καθιέρωση ευκολότερων τρόπων αναζήτησης πληροφοριών.
-



Τα ευρωπαϊκά σχέδια για την διεύρυνση των υποδομών (1/2)

- Το EBI ανταποκρίνεται για την ώρα στην αυξανόμενη απαίτηση τόσο των δημιουργών όσο και των χρηστών, αλλά η διαφαινόμενη αλλαγή κλίμακας για τις επιστήμες ζωής απαιτεί μακροπρόθεσμο σχεδιασμό.
- Αυτό το στόχο θα το πραγματοποιήσει το πρόγραμμα ELIXIR με το οποίο θα καθιερωθεί μία αξιόπιστη υποδομή που θα μεγιστοποιήσει την πρόσβαση σε βιολογικές πληροφορίες που ήδη διανέμονται σε περισσότερες από 500 βάσεις δεδομένων σε όλη την Ευρώπη.



Τα ευρωπαϊκά σχέδια για την διεύρυνση των υποδομών (2/2)

- Το πρόγραμμα ELIXIR θα αντιμετωπίσει προβλήματα διαχείρισης δεδομένων, αποτελεσματικής χρηματοδότησης και θα αναπτύξει νέα βιοπληροφορικά εργαλεία.
- Η ανάπτυξη του ELIXIR θα σηματοδοτήσει μία νέα φάση, όπου η υπολογιστική υψηλής πυκνότητας δεδομένων θα δώσει μεγάλη ώθηση στις επιστήμες ζωής.
- Τα προσδοκώμενα αποτελέσματα θα είναι σημαντικά στην κατανόηση της ζωής και της εξέλιξης, στη συνθετική και συστηματική βιολογία και τέλος εκτός της καθαρής επιστήμης και σε πιο εφαρμοσμένους επιστημονικούς τομείς, όπως στην παραγωγή φαρμάκων, στη βιοτεχνολογία, τη δημόσια υγεία, τη γεωργία και τις περιβαλλοντικές επιστήμες.



Πολυπύρρηνοι επεξεργαστές και επιστημονικές ανακαλύψεις

James Larus Dennis Gannon Microsoft Research



Παράλληλη υπολογιστική (1/2)

- Η ακόρεστη δίψα των ερευνητών για ολοένα μεγαλύτερους υπολογισμούς μπορεί μόνο να ικανοποιηθεί με τον παραλληλισμό - δηλαδή την εκτέλεση περισσότερων από μία λειτουργία ταυτόχρονα. Αυτό παρουσιάζει δυσκολίες στο επίπεδο του λογισμικού που ακόμη δεν έχουν πλήρως ξεπεραστεί.
- Ο παραλληλισμός υλοποιήθηκε με την παράλληλη λειτουργία επεξεργαστών, χωρίς όμως κοινή μνήμη αφού αυτό απεδείχθη δύσκολο, με την επικοινωνία μεταξύ των μηχανών να γίνεται με αποστολή μηνυμάτων. Παράλληλες σειρές από Pc συνδεδεμένες σε ETHERNET αποτελούν μία περίπτωση παραλληλισμού.



Παράλληλη υπολογιστική (2/2)

- Νεώτερη προσπάθεια αποτελούν οι πολυπύρρηνοι επεξεργαστές.
- Ο εκφρασθείς με το νόμο του Moore διπλασιασμός των τρανζίστορς ανά διετία έδωσε μία ετήσια αύξηση της τάξεως του 40% από τη δεκαετία του 70 που έφτασε σε κάποιο όριο λόγω προβλημάτων από την κατανάλωση ενέργειας.
- Για το ξεπέρασμα του προβλήματος επινοήθηκαν οι πολυπύρρηνοι επεξεργαστές που αθροιστικά έχουν μεγαλύτερες δυνατότητες.



Προκλήσεις του παράλληλου προγραμματισμού

- Η νέα γενιά υπολογιστών έχει τα ίδια προβλήματα λογισμικού όπως και παλαιότερα σε παράλληλα συστήματα αφού το λογισμικό είναι φτιαγμένο για σειρές υπολογιστών και δεν θα τρέξει γρηγορότερα σε πολυπύρηνους επεξεργαστές.
- Για να αξιοποιηθούν οι δυνατότητες των νέων υπολογιστών το λογισμικό θα πρέπει να μπορεί να σπάσει ένα πρόβλημα σε μικρότερα, τα οποία να επιλυθούν ανεξάρτητα και να συνενώσει τα επιμέρους αποτελέσματα σε μία και μόνη απάντηση.
- Αυτό το παράλληλο λογισμικό δεν έχει ακόμη κατορθωθεί να αναπτυχθεί πλήρως και πολλά προβλήματα θα πρέπει να αντιμετωπισθούν και όταν ευοδωθούν οι προσπάθειες που είναι σε εξέλιξη, η παράλληλη υπολογιστική θα αποτελέσει ένα νέο επιστημονικό πρότυπο (Paradigm).



Απαντώντας στις προκλήσεις (1/2)

- Η παλαιότερη και πιο μεθοδική προσπάθεια παράλληλου προγραμματισμού είναι ο παράλληλος προγραμματισμός δεδομένων, όπου μία λειτουργία ή ακολουθία λειτουργιών εφαρμόζεται ταυτόχρονα σε όλα τα στοιχεία σε μία συλλογή δεδομένων.
- Ο παραλληλισμός δεν είναι εμφανής στον προγραμματιστή και κάθε υπολογιστική πράξη είναι ανεξάρτητη από άλλες ταυτόχρονες σε άλλα δεδομένα και ειδικός εξοπλισμός υποστηρίζει αυτόν τον τρόπο προγραμματισμού.
- Δυστυχώς αυτό το είδος παραλληλισμού δεν μπορεί να εφαρμοστεί σε όλα τα προβλήματα.
- Μία πρόσφατη ιδέα είναι η συναλλακτική μνήμη - transactional memory (TM) ένας μηχανισμός συντονισμού του διαμοιρασμού δεδομένων σε ένα πολυπύρηνο υπολογιστή. Είναι μία πολλά υποσχόμενη πρακτική που όμως δεν μπορεί να εφαρμοστεί άμεσα γιατί υπάρχουν πολλά τεχνικά προβλήματα.



Απαντώντας στις προκλήσεις (2/2)

- Μία άλλη δελεαστική ιδέα είναι η χρήση λειτουργικών γλωσσών προγραμματισμού που είναι στο στάδιο της εξέτασης.
- Τελικά η επινόηση των πολυπύρηνων υπολογιστών έφερε στην επιφάνεια το θέμα του παράλληλου προγραμματισμού σαν μία μεγάλη πρόκληση που όταν αντιμετωπιστεί θα δώσει τεράστια ώθηση στις υπολογιστικές δυνατότητες.



Ο Παραλληλισμός και το νέφος

Dennis Gannon -Dan Reed Microsoft Research



Η σχέση της υπολογιστικής επιστήμης με τις άλλες επιστήμες

- Την περασμένη δεκαετία η επιστημονική έρευνα μέσω υπολογιστικής ανεδείχθη σαν ο τρίτος πυλώνας της έρευνας μετά το πείραμα και τη θεωρία.
 - Η υπολογιστική επιστήμη σταδιακά περιέλαβε τόσο τον υπολογισμό (computation) όσο και την ανάλυση δεδομένων.
 - Σήμερα ένα, κυριολεκτικά, τσουνάμι δεδομένων από δίκτυα αισθητήρων που υπάρχουν από τα βάθη των ωκεανών μέχρι το διάστημα και άλλους μηχανισμούς μας κατακλύζει με κίνδυνο να μας υπερκεράσει.
 - Αυτή η μετάβαση από τη σπανιότητα στην πληθώρα δεδομένων μας αναγκάζει να αξιοποιήσουμε νέα κατάλληλα υπολογιστικά εργαλεία για να τα διαχειριστούμε.
 - Επίσης τώρα δεν δίνουμε σημασία σε μεμονωμένα δεδομένα αλλά στα αποτελέσματα της ανάλυσης των δεδομένων που πολλές φορές είναι διεπιστημονική.
-



Η ανάλυση νέφους μία απάντηση στο τσουνάμι των δεδομένων

- Η μεγάλη ανάπτυξη της υπολογιστικής δυναμικότητας έδωσε τεράστια ώθηση στην επιστήμη, την μηχανική και στην οικονομία της γνώσης.
- Η υπολογιστική νέφους είναι η τελευταία εξέλιξη στην υπολογιστική επιστήμη, αφού επιτρέπει σε ομάδες να φιλοξενήσουν, να επεξεργαστούν και να αναλύσουν τεράστιους όγκους πολυτομεακών δεδομένων.
- Η επεξεργασία και αποθήκευση, δεδομένων σε μεγάλα κέντρα δεδομένων (data centers) δημιουργεί οικονομίες κλίμακας και μεγάλη εξοικονόμηση πόρων σε κατασκευή, εξοπλισμό, λειτουργία και συντήρηση, ενθαρρύνει τη συνεργασία και το διαμοιρασμό δεδομένων και επιτρέπει την διατήρηση δεδομένων για μεγάλα διαστήματα.
- Η ύπαρξη σήμερα τεράστιων κέντρων δεδομένων και υπερυπολογιστών είναι ένα δίδυμο που αλληλοσυμπληρώνεται και δίνουν εντυπωσιακά αποτελέσματα.



Υπολογιστική νέφους και παραλληλισμός (1/2)

Ο παραλληλισμός μπορεί να αξιοποιηθεί στην υπολογιστική νέφους με δύο τρόπους:

- Για την ανθρώπινη πρόσβαση αφού οι εφαρμογές νέφους είναι σχεδιασμένες να είναι προσβάσιμες από τον ιστό (web) και είναι οργανωμένες σαν δύο ή περισσότερες στοιβάδες διαδικασιών παρέχοντας ευελιξία. Η πρώτη είναι η διεπιφάνεια με το χρήστη που παίρνει τα αιτήματά του τα διαχειρίζεται και τα διαβιβάζει στην δεύτερη στοιβάδα. Η δεύτερη στοιβάδα -ο εργάτης- εκτελεί την αναλυτική εργασία που ζητά ο χρήστης.
- Ο δεύτερος τρόπος αφορά την φύση της ανάλυσης δεδομένων που αναλαμβάνει η εφαρμογή. Σε αναλύσεις τεράστιων ποσοτήτων δεδομένων είναι πιο αποτελεσματικό τα δεδομένα να διαμοιράζονται σε πολλούς επεξεργαστές και τα αποτελέσματα να επανενώνονται και να στέλνονται στο χρήστη.
- Τρία είδη τεχνολογίας στηρίζουν αυτό το είδος παραλληλισμού
 - Η Google έχει μία εσωτερική εκδοχή.
 - Η Yahoo! Έχει μία ανοιχτού τύπου εκδοχή γνωστή ως Hadoop.



Υπολογιστική νέφους και παραλληλισμός (2/2)

- Η Microsoft έχει ένα μηχανισμό map-reduce γνωστό με το όνομα DryadLINQ
 - Η Microsoft Windows Azure στηρίζει ένα συνδυασμό πολλών χρηστών με ένα παραλληλισμό ανάλυσης δεδομένων όντας πιο ευέλικτη και επιτρέποντας την ταυτόχρονη χρήση από πολλούς.
 - Αυτό το είδος παραλληλισμού (το map-reduce) είναι απαραίτητο σε πολλές μεγάλης κλίμακας επιστημονικές αναλύσεις πχ την σύγκριση του DNA δείγματος με αυτό χιλιάδων γνωστών οργανισμών.
 - Επίσης η αποτελεσματική αναζήτηση στον παγκόσμιο ιστό έγινε εφικτή με τον παραλληλισμό σε μαζικά κέντρα δεδομένων.
 - Τέλος μπαίνοντας στην εποχή των τεράστιων δεδομένων η υπολογιστική νέφους με τον ενδογενή της παραλληλισμό σε πολλά επίπεδα αποτελεί μία αποτελεσματική τεχνολογία για προώθηση της επιστημονικής γνώσης.
-



Οι επιπτώσεις των εργαλείων ροής εργασίας σε μία δεδομενοκεντρική έρευνα

Carole Goble University of Manchester David De Roure University of Southampton



Εισαγωγή (1/2)

- Βρισκόμαστε σε μία φάση της επιστημονικής έρευνας όπου οι υποθέσεις μπορούν πλέον και να δημιουργηθούν με συνδυασμό και εξόρυξη από υφιστάμενα ήδη δεδομένων.
- Η παραγωγή δεδομένων σε πολλούς επιστημονικούς τομείς, όπως στη βιολογία αυξάνεται αλματωδώς με ρυθμούς της τάξεως *tera* για ένα απλό πείραμα και ανάλογη τεράστια αύξηση παρατηρείται στον αριθμό των βάσεων δεδομένων.
- Ομοίως αυξάνει και η πολυπλοκότητα των δεδομένων καθιστώντας έτσι τη συνολική διαχείρισή τους δύσκολη από επιστήμονες μάλιστα που δεν έχουν τόση εξειδίκευση στον τομέα της πληροφορικής.



Εισαγωγή (2/2)

- Η χρήση ροών εργασίας (workflows) παρέχει:
 - Συστηματικά και αυτόματα μέσα ανάλυσης δεδομένων.
 - Ένα εργαλείο ελέγχου της διαδικασίας, ώστε τα αποτελέσματα να επανελέγχονται, να αναθεωρούνται και να προσαρμόζονται.
 - Οπτική επαφή της διαδικασίας για διευκόλυνση των χρηστών.
 - Μία ολοκληρωμένη πλατφόρμα πρόσβασης που οι επιστήμονες πληροφορικής δεν χρειάζεται να εξειδικεύονται σε κάθε επιμέρους θέμα.
- Οι ροές εργασίας γίνονται έτσι πρότυπο για την υλοποίηση της επιστήμης σε μεγάλη κλίμακα, με τη εφαρμογή αγωγών προπαρασκευής και ανάλυσης δεδομένων καθώς και αποτελώντας το μέσο απόληξης γνώσης.



Ορισμός των ροών εργασίας (1/2)

- Ροή εργασίας είναι μία λεπτομερής περιγραφή μίας επιστημονικής εργασίας δομημένη σε βήματα που συντονίζει πολλαπλά έργα. Κάθε έργο παριστά μία υπολογιστική εργασία που το αποτέλεσμα της αποτελεί εισερχόμενο για άλλο έργο σύμφωνα με μία τοπολογία που ενορχηστρώνει τη ροή των εργασιών.
- Τα συστήματα ροών εργασίας γενικά έχουν τρία συστατικά: Μία πλατφόρμα εκτέλεσης, μία οπτική δομή σχεδιασμού και μία εργαλειοθήκη ανάπτυξης
- Η πλατφόρμα εκτελεί την ροή εργασίας για λογαριασμό των εφαρμογών λαμβάνοντας ειδική μέριμνα για θέματα, όπως η ανάκτηση δεδομένων σε περίπτωση αστοχίας, η βελτιστοποίηση της μνήμης και της αποθηκευτικής ικανότητας, ο χειρισμός δεδομένων, η ασφάλεια και η παρακολούθηση της πολιτικής πρόσβασης.
- Η δομή σχεδιασμού παρέχει μία οπτικοποιημένη εφαρμογή για κατασκευή και διαμοιρασμό ροών εργασίας καθώς και προετοιμασίας των επιμέρους συστατικών που θα αποτελέσουν τα επιμέρους βήματα. Είναι στην ουσία ένα απλοποιημένο εργαλείο προγραμματισμού που επιτρέπει στους επιστήμονες να το φτιάξουν όπως θέλουν και όποτε θέλουν χωρίς να προσφεύγουν σε ειδικούς.



Ορισμός των ροών εργασίας (2/2)

- Τέλος η εργαλειοθήκη διευρύνει τις δυνατότητες του συστήματος σε διάφορες εφαρμογές.
- Κάθε σύστημα ροής εργασίας έχει τα δικά του χαρακτηριστικά ανάλογα με την σκοπούμενη χρήση και υπάρχουν πολλά τέτοια συστήματα ένα εκ των οποίων, αρκετά πολύπλοκο, φέρει το ελληνικό όνομα “Ταβέρνα”.



Χρήση των ροών εργασίας

- Οι ροές εργασίας απαλλάσσουν τον επιστήμονα από κοπιώδεις εργασίες άσχετες με το επιστημονικό του αντικείμενο και του επιτρέπει να διαχειρίζεται δεδομένα σε μεγάλους όγκους με ακρίβεια, αποτελεσματικότητα και επαναληπτικότητα.
- Σε ένα ανώτερο επίπεδο οι ροές εργασίας αποτελούν μία έκφραση του “στεγνού εργαστηρίου” “in silico” (κατά αναλογία του in vitro/in vivo).
- Είναι ιδανικές για τη συλλογή και συγκέντρωση δεδομένων μέσα από κατανεμημένα σύνολα δεδομένων όπως πχ η συλλογή και συγκέντρωση βιολογικών δεδομένων από τράπεζες.
- Τελικά οι ροές εργασίας απλοποιούν τα πράγματα για τους χρήστες δίνοντάς τους εύχρηστα εργαλεία.



Οι ροές εργασίας έκαναν εφικτή την επικεντρωμένη στα δεδομένα επιστήμη (1/2)

- Έχουν ευκολία στη χρήση και ευελιξία, δίνουν δυνατότητες επανάληψης, χρήσης εικονικών δεδομένων, αλλαγής των παραμέτρων, σύγκρισης και αξιολόγησης των αποτελεσμάτων.
- Παρέχουν διαφάνεια και πρόσβαση στη διαδικασία, μπορούν να ανασχεδιάζονται και αναπροσαρμόζονται.
- Ο σχεδιασμός των ροών εργασίας ενίοτε είναι δύσκολος για τους μη ειδικούς, αφού αποτελεί μία μορφή προγραμματισμού. Συνεπώς ο σχεδιασμός από εξειδικευμένους συλλογών από ροές εργασίας που μπορούν να επαναχρησιμοποιούνται για άλλους σκοπούς, επιτρέπει στους μη ειδικούς να χρησιμοποιούν τα πιο εξελιγμένα πρωτόκολλα.



Οι ροές εργασίας έκαναν εφικτή την επικεντρωμένη στα δεδομένα επιστήμη (2/2)

- Προωθούν τη συνεργασία μεταξύ επιστημόνων, με το διαμοιρασμό δεδομένων και υπηρεσιών πολλές φορές σε διεπιστημονικό πλαίσιο.
- Δίνουν τεχνικές για την υποστήριξη του νέου προτύπου της επικεντρωμένης στα δεδομένα επιστήμης.
- Υπάρχουν και ορισμένα προβλήματα προς αντιμετώπιση, όπως η ύπαρξη μεγάλου αριθμού μη ελεγχόμενων πλατφορμών, δυσκολίες στην χρήση, αστοχίες σε επιμέρους στοιχεία των πλατφορμών κλπ.



Σημασιολογική eScience: Κωδικοποιώντας τη σημασία στην ψηφιακά ενισχυμένη επιστήμη της επόμενης γενιάς

Peter Fox James Hendler Rensselaer Polytechnic Institute



Escience: Το νέο στάδιο εξέλιξης της επιστήμης

- Οι παραδοσιακές τεχνολογίες διαχείρισης δεδομένων δεν ήταν σχεδιασμένες για τους όγκους που σήμερα πρέπει να διαχειρίζονται και που σε ορισμένα προγράμματα είναι της τάξεως των petabytes.
 - Η ψηφιακή ή ηλεκτρονική διευκόλυνση της επιστήμης, η eScience, είναι πλέον ουσιώδης και πολύ διαδεδομένη.
 - Η επιστήμη των μεγάλων όγκων δεδομένων πρέπει πλέον να γίνει πιο προσβάσιμη, πιο ανοικτή και πιο φιλική σε άξιους επιστήμονες που δεν ανήκουν στο κλάδο της πληροφορικής.
 - Στο νέο τοπίο ανακύπτουν ερωτήματα σχετικά με το τρόπο αξιοποίησης των δεδομένων στα οποία υπάρχει πλέον πρόσβαση.
-



Εργαλεία για την αξιοποίηση της πρόσβασης σε πληθώρα δεδομένων (1/2)

- Καθώς αυξάνει ο όγκος και η πολυπλοκότητα των δεδομένων οι επιστήμονες χρειάζονται νέες ικανότητες που βασίζονται σε νέες σημασιολογικές προσεγγίσεις που πλέον κερδίζουν έδαφος σε πολλούς επιστημονικούς τομείς, όπως η οικολογία, οι επιστήμες υγείας κλπ.
- Με την αξιοποίηση σημασιολογικών υποδομών θα διευκολυνθεί η επιστήμη, η λογική διερεύνηση υποθέσεων, ανάλυση και διαχείριση δεδομένων προς χρήση επιστημόνων, φοιτητών, αλλά και μη ειδικών.
- Η αυξανόμενη διαθεσιμότητα δεδομένων στον ιστό, έχει οδηγήσει στην ανάπτυξη κατάλληλων εφαρμογών και γλωσσών.



Εργαλεία για την αξιοποίηση της πρόσβασης σε πληθώρα δεδομένων (1/2)

- Όλες οι ικανότητες που τώρα απαιτούνται για την eScience, όπως η διαχείριση δεδομένων, η ανάπτυξη εντοχών και εκτέλεση ροών εργασίας, χρειάζονται σημασιολογική παράσταση για να γίνει η eScience πλήρως πυκνή σε δεδομένα.
- Η ανάγκη για περισσότερη σημασιολογία στην eScience προκύπτει από την ολοένα περισσότερο κατανομημένη και διεπιστημονική φύση της έρευνας.
- Γίνεται φανερό η ανάγκη συνεργασίας μεταξύ των λοιπών επιστημόνων και των επιστημόνων πληροφορικής, καθώς επίσης μεταξύ μηχανικών πληροφορικής, προμηθευτών δεδομένων και διαχειριστών δεδομένων.



Αξιοποίηση υπηρεσιών του ιστού (Web) για την απόκτηση δεδομένων και πληροφοριών

- Η χρήση υπηρεσιών του web για την απόκτηση πληροφοριών και δεδομένων διευρύνεται συνέχεια καθώς επίσης και γλώσσες σήμανσης που βασίζονται στο web, ακόμη αναδρομικά δεδομένα ή δεδομένα σε πραγματικό χρόνο.
 - Οι υπηρεσίες αυτές πλέον παρέχονται σε πολλούς τομείς και τα αντίστοιχα μητρώα είναι ιδιαίτερα δημοφιλή, όμως χρειάζονται περαιτέρω βελτίωση.
 - Καθώς η ζήτηση για πληροφορίες μεγαλώνει και σε κύκλους και πέραν των επιστημόνων μεγαλώνει και η ανάγκη για ευκολότερη πρόσβαση σε αυτές.
 - Ενώ οι επιστήμονες βασίζονται συνεχώς περισσότερο στο Web για αναζήτηση δεδομένων, δεν έχει ακόμη δημιουργηθεί μία σχετική ατζέντα για την διερεύνηση των αναδυόμενων τάσεων από τις σημασιολογικές τεχνολογίες και την αλληλεπίδραση με τους σημασιολογικούς ερευνητές του web.
-



Πρόσκληση για τέσσερα σημεία δράσης για τη δημιουργία της ατζέντας

- Οι επιστήμονες ερευνητές πρέπει να συνεργαστούν με τους συναδέλφους στην επιστήμη των υπολογιστών και την πληροφορική για να αναπτύξουν ειδικές απαιτήσεις για το πεδίο και να εφαρμόσουν και να αξιολογήσουν τις γλώσσες, τα εργαλεία και τις εφαρμογές που αναπτύσσονται για τη σημασιολογική eScience.
- Οι επιστημονικές και επαγγελματικές κοινότητες πρέπει να παράσχουν τις ρυθμίσεις στις οποίες μπορεί να υλοποιηθεί η απαιτούμενη πλούσια αλληλεπίδραση μεταξύ των απαιτήσεων της επιστήμης και των ικανοτήτων της πληροφορικής και πρέπει να αναγνωρίσουν τη σημασία αυτού του έργου για την επαγγελματική εξέλιξη.
- Οι οργανισμοί χρηματοδότησης πρέπει να στοχεύουν όλο και περισσότερο στην οικοδόμηση κοινοτήτων πρακτικής, με έμφαση στους τύπους διεπιστημονικών ομάδων ερευνητών και επαγγελματιών που απαιτούνται για την προώθηση και τη διατήρηση των σημασιολογικών προσπαθειών της eScience.
- Όλες οι πλευρές - επιστήμονες, κοινωνίες και χρηματοδότες πρέπει να διαδραματίσουν έναν ρόλο στη δημιουργία διακυβέρνησης γύρω από ελεγχόμενα λεξιλόγια, ταξινομήσεις και οντολογίες που μπορούν να χρησιμοποιηθούν σε επιστημονικές εφαρμογές για να εξασφαλίσουν την κυκλοφορία και την εξέλιξη της γνώσης που κωδικοποιούνται στη σημασιολογία.



Οπτικοποίηση για μία επιστήμη εντατικών δεδομένων

Charles Hansen Chris R.- Johnson Valerio Pascucci -Claudio T. Silva University of Utah



Εισαγωγή

- Τα τελευταία χρόνια υπάρχει ένα big-bag δεδομένων και πληροφοριών το 90% των οποίων είναι σε ψηφιακή μορφή.
- Η μεγάλη πρόκληση του 21ου αιώνα θα είναι η κατανόηση και η αποτελεσματική χρήση του τεράστιου αυτού όγκου δεδομένων.
- Η οπτικοποίηση των δεδομένων μεταξύ των άλλων επιτρέπει ανίχνευση αναμενόμενων ή μη αποτελεσμάτων, επιτρέπει τη θεμελίωση θεωρητικών μοντέλων, παρέχει συγκρίσεις πραγματικών και δεδομένων από προσομοιώσεις, διευκολύνει την ερμηνεία αποτελεσμάτων.
- Η οπτική ανάλυση δεδομένων είναι στην πραγματικότητα μία επιστημονική ανακάλυψη που όμως απαιτούνται να γίνουν ακόμη πολλά βήματα.



VISUS: Μία σταδιακή προσέγγιση στην εξερεύνηση δεδομένων

- Πρόσφατα οι επιστήμονες με τη χρήση υπερυπολογιστών προσομοίωσαν φυσικά φαινόμενα παράγοντας τεράστιους όγκους δεδομένων που πρέπει να οπτικοποιηθούν και να αναλυθούν, ώστε να κατανοηθούν τα φαινόμενα που συνδέονται με αυτά.
 - Η εργασία αυτή απαιτεί ανάπτυξη αλγόριθμων με υψηλή λειτουργικότητα.
 - Τεχνικές ταχείας ροής δεδομένων και εκτός πυρήνα υπολογιστική αντιμετωπίζουν αυτά τα προβλήματα, όπως πχ με το σχέδιο VISUS η διαχείριση των δεδομένων είναι πιο ευέλικτη και δεν εξαρτάται από τα χαρακτηριστικά του χρησιμοποιούμενου εξοπλισμού.
 - Παραμένουν ακόμη πολλά προβλήματα προς επίλυση.
-



Vistrails: Προέλευση και εξερεύνηση δεδομένων

- Η εξερεύνηση των δεδομένων είναι μία κρίσιμη επιστημονική εργασία που απαιτεί την οπτικοποίησή τους, την ανακάλυψη συσχετισμών, την συνεργασία και δοθέντος του όγκου και της πολυπλοκότητας των δεδομένων νέα εργαλεία θα πρέπει να αναπτυχθούν.
 - Η προέλευση των δεδομένων είναι επίσης κρίσιμης σημασίας για την αξιοπιστία τους, την αναπαραγωγή τους, την τεκμηρίωση υποθέσεων κλπ.
 - Το vistrails είναι ένα σύστημα οπτικοποίησης, εξόρυξης και ολοκλήρωσης δεδομένων, ανοικτής πηγής (open source) που ανιχνεύει την προέλευση και μπορεί να συνδυαστεί με υπάρχουσες βιβλιοθήκες.
 - Το σύστημα αυτό έχει την ικανότητα της ανίχνευσης της προέλευσης σε ροές εργασίας (workflows).
 - Η γνώση της προέλευσης και η απάντηση σε ερωτήματα όπως ποιος δημιούργησε τα δεδομένα και πότε, ποιος τα τροποποίησε, πότε και με ποία διαδικασία είναι κρίσιμης σημασίας για την αξιοπιστία των δεδομένων και το vistrails είναι αξιόπιστο και ευέλικτο σύστημα που επιτελεί τις παραπάνω λειτουργίες.
-



Τεχνικές οπτικοποίησης της ροής

- Τρισδιάστατες ποσοτικά και ποιοτικά ακριβείς εκτιμήσεις φαινομένων ροής απαιτούνται σε πολλούς τομείς της επιστήμης και της μηχανικής, όπως στην αεροναυτική, στις εξατμίσεις αυτοκινήτων, στην αλληλεπίδραση μεταξύ βιοηλεκτρικών πεδίων και της επιφάνειας οργάνων
- Η προσομοίωση των ροών με οπτικοποίηση των αποτελεσμάτων είναι μία πολύ καλή τεχνική για την κατανόηση των φαινομένων και κατάλληλοι αλγόριθμοι έχουν σχεδιαστεί για το σκοπό αυτό.
- Μία πρόσφατη τάση είναι η χρήση GPUs (Graphics Processing Units) για την οπτική παράσταση ροών σε καμπυλωμένες επιφάνειες λόγω της ικανότητας της μεθόδου να παρέχει λεπτομερή αναπαράσταση της ροής, χωρίς σαφή παραμετροποίηση της επιφάνειας.
- Επειδή όμως με την παραπάνω τεχνική δημιουργούνται ατέλειες στην οπτικοποίηση αναπτύχθηκε ένα νέο σχέδιο, τα διαγράμματα ροής (Flow charts) με τα οποία η οπτικοποίηση φθάνει σε υψηλά επίπεδα τεχνικής.



Μελλοντικές προκλήσεις για οπτικοποίηση πυκνών δεδομένων

- Δραστικές πρόοδοι στις τεχνικές οπτικοποίησης πρέπει να γίνουν ενόψει της τεράστιας σε όγκο και πολυπλοκότητα παραγωγής δεδομένων.
 - Η δραστική προσομοίωση και οπτικοποίηση φαινομένων πρέπει να βασιστεί σε δυνατούς αλγόριθμους με γερά μαθηματικά θεμέλια, όπως η τοπολογία, οι πιθανότητες και η στατιστική.
 - Οι αλγόριθμοι θα πρέπει να μπορούν να φιλτράρουν τον τεράστιο όγκο των δεδομένων και να δίνουν μία περίληψη οπτικοποίησης.
 - Η παράσταση και η ποσοτικοποίηση της αβεβαιότητας με οπτικές μεθόδους θα βοηθήσει σημαντικά τους επιστήμονες.
 - Το πεδίο της οπτικής ανάλυσης έχει ακόμη σημαντικά σημεία προς διερεύνηση και βελτίωση και προς αυτήν την κατεύθυνση πρέπει να γίνουν περαιτέρω προσπάθειες.
-



**Μια πλατφόρμα για όλα όσα γνωρίζουμε:
Δημιουργία μιας καθοδηγούμενης από τη γνώση
ερευνητικής υποδομής**

Σάββας Παραστατίδης Microsoft



Εισαγωγή

- Τα υπολογιστικά συστήματα έχουν γίνει ζωτικό μέρος του σύγχρονου επιστημονικού περιβάλλοντος και η επιστημονική κοινότητα χρησιμοποιεί όρους, όπως eScience για να αποδώσει τα χαρακτηριστικά της ψηφιακής εποχής μας.
 - Πράξεις μέσω των υπολογιστών όπως η συλλογή επεξεργασία και αρχειοθέτηση δεδομένων, ο διαμοιρασμός δεδομένων και η συνεργασία αποτελούν πλέον μία καθημερινότητα.
 - Αυτή η διευκόλυνση που έφεραν οι υπολογιστές οδήγησε σε μία έκρηξη δεδομένων, η διαχείριση των οποίων μας φέρνει εμπρός σε προκλήσεις, που όπως ανέλυσαν και οι Jim Gray and Alex Szalay φέρνουν πιο κοντά μία νέα μεθοδολογική προσέγγιση για την επιστήμη-το τέταρτο πρότυπο.
-



Το τέταρτο πρότυπο

- Η ανάδειξη του ιστού (web) σαν μηχανισμού διαμοιρασμού δεδομένων και συνεργασίας έχει σπάσει πολλά στεγανά στον τρόπο διεξαγωγής της έρευνας.
 - Η υπολογιστική νέφους και η νέα γενιά υπολογιστικών συστημάτων πυκνών δεδομένων όπως τα DISC, Google's MapReduce, Hadoop, 4 and Dryad κατάφεραν να διαχειριστούν τεράστιες ποσότητες δεδομένων
 - Οι ως τώρα αναπτυχθείσες υποδομές έκαναν το πρώτο βήμα λύνοντας τα προβλήματα αποθήκευσης, επεξεργασίας, διαχείρισης δεδομένων.
 - Το επόμενο βήμα θα είναι η αυτόματη μέσω υπολογιστών συλλογή, επεξεργασία, συσχετισμός ερμηνεία δεδομένων από διάφορες πηγές όπως το διαδίκτυο, σκληροί δίσκοι ερευνητών.
 - Δεν υπάρχει ακόμη ένας μηχανισμός που αυτόματα να μπορεί να συλλέξει και να επεξεργαστεί τη γνώση σε ένα παγκόσμιο επίπεδο πράγμα που θα είναι το επόμενο βήμα, μετά το τέταρτο πρότυπο, για την επιστήμη.
-



Επιστημονικές υποδομές προσανατολισμένες στη γνώση

- Η εργασία από την κοινότητα του σημασιολογικού ιστού (semantic web) ανέπτυξε τεχνολογίες για προτυποποίηση δεδομένων, την παράσταση πληροφοριών και την ανταλλαγή σημασιολογιών μέσα στο πλαίσιο ενός επιστημονικού τομέα όπως η Web Ontology Language, ή OWL.
- Το επόμενο βήμα θα είναι η αυτόματη συλλογή, επεξεργασία δεδομένων και προς την κατεύθυνση αυτή έχουν ξεκινήσει κάποιες προσπάθειες όπως πχ η Wolfram Alpha, myGrid και myExperiment.
- Η ανθρωπότητα θα είχε πολύ να ωφεληθεί από μία παγκόσμια υποδομή συγκέντρωσης, διαμοιρασμού και διαχείρισης της γνώσης, χωρίς αυτό να σημαίνει ότι ο κάθε μεμονωμένος ερευνητής θα χάσει την ανεξαρτησία του στην τεκμηρίωση και παρουσίαση της έρευνάς του.
- Μία τέτοια πλατφόρμα θα αποτελούσε αστείρευτη πηγή δεδομένων για τους ερευνητές, θα παρείχε υπολογιστικές υπηρεσίες και όλα αυτά με αυτοματοποιημένες διαδικασίες με τη χρήση των κατάλληλων εργαλείων, χωρίς πάντως να αξιοποιεί την τεχνητή νοημοσύνη.



Βρίσκοντας τη γνώση

- Η διεπιστημονική έρευνα είδε μεγάλη ώθηση ως αποτέλεσμα της eScience και της κυβερνο-υποδομής ενώ η τεχνολογία βοήθησε με την στήριξη της συνεργασίας και του διαμοιρασμού δεδομένων σε διεπιστημονικό επίπεδο.
 - Η ερευνητική υποδομή θα στηρίζει την αυτόματη συλλογή δεδομένων από διάφορα πεδία που διενεργήθηκε σε διάφορους χρόνους και με την τεκμηρίωση με την οποία εξήχθησαν.
 - Έτσι στο μέλλον ένας ερευνητής θα μπορούσε να θέσει το ερώτημα “Εντοπίστε 100.000 μόρια που είναι παρόμοια με τους γνωστούς αναστολείς της πρωτεάσης του HIV, στη συνέχεια υπολογίστε τις ηλεκτρονικές τους ιδιότητες και τοποθετείστε τα σε μεταλλάξεις ιογενούς μορφής”. Από το παράδειγμα φαίνεται η σημασία της καθιέρωσης σημασιολογικών web τεχνολογιών.
 - Τα επόμενα βήματα θα είναι η χαλάρωση της στεγανότητας μεταξύ των επιστημονικών πεδίων και αυτό θα δώσει περαιτέρω ώθηση στην επιστημονική έρευνα.
-



Πρόσκληση για δράση

- Σήμερα οι πλατφόρμες που προσφέρουν υπολογιστικές εφαρμογές του MapReduce δηλαδή η Hadoop και η Dryad κάνουν εύκολη τη διαχείριση δεδομένων σε μεγάλη κλίμακα.
- Στο μέλλον θα χρειαστούν πλατφόρμες και κυβερνο-υποδομή (cyberinfrastructure) για την συλλογή δεδομένων και αξιοποίηση της συγκεντρωμένης σε παγκόσμιο επίπεδο γνώσης.
- Οι επιστημονικές κοινότητες θα πρέπει να συνεργαστούν για την καθιέρωση λεξιλογίων για την αλίευση πληροφοριών και γεγονότων από τον κάθε ξεχωριστό τομέα και αυτά τα λεξιλόγια, θα αλληλοσυσχετιστούν μέσω των κατάλληλων υποδομών, ώστε να δημιουργηθεί ένα εύκολα προσβάσιμο παγκόσμιο δίκτυο γεγονότων και πληροφοριών.
- Οι επιστημονικές υποδομές στο μέλλον θα είναι πολύ πιο ευέλικτες και θα οδηγούνται από τη γνώση και όχι όπως σήμερα από τα δεδομένα.



ΑΚΑΔΗΜΑΪΚΗ ΕΠΙΚΟΙΝΩΝΙΑ



Εισαγωγή

Lee Dirks | Microsoft Research

Στο κεφάλαιο αυτό εκτίθενται οι απόψεις πάνω στα παρακάτω θέματα

- Πως το τέταρτο πρότυπο εφαρμόζεται στην ακαδημαϊκή επικοινωνία.
- Πως με την επίδραση των νέων τεχνολογιών μετασχηματίζονται οι επιστημονικές δημοσιεύσεις.
- Το ότι οι ακαδημαϊκοί θα πρέπει να αντιληφθούν τα εν δυνάμει πλεονεκτήματα της τεχνολογίας στην ακαδημαϊκή επικοινωνία.
- Το γεγονός ότι ενίοτε οι επιστήμονες με τη συμπεριφορά τους παρεμποδίζουν την πρόοδο στην επιστημονική έρευνα.
- Την διασύνδεση παρά τους φραγμούς που ακόμη υπάρχουν μεταξύ των διαφόρων επιστημονικών τομέων.



Το τέταρτο πρότυπο του Jim Gray και η Κατασκευή του επιστημονικού φακέλλου

Clifford Lynch Coalition for Networked Information



Από το τρίτο στο τέταρτο επιστημονικό πρότυπο

- Το τρίτο πρότυπο διεξαγωγής της επιστημονικής έρευνας, αυτό των μεγάλης κλίμακας προσομοιώσεων, αναδείχτηκε από τις εργασίες του John Newman στα μέσα του 20ου αιώνα.
- Το τρίτο πρότυπο αν και συνέβαλε σε μεγάλο βαθμό στην επιστημονική έρευνα περιήλθε σε αδιέξοδο όταν έχοντας να κάνει με τεράστιο αριθμό δεδομένων και χαοτικές συμπεριφορές δεν μπόρεσε να φθάσει σε αποτελέσματα, όπως πχ στην μακροπρόθεσμη πρόβλεψη του καιρού.
- Το τέταρτο πρότυπο που πρότεινε ο Jim Gray, αυτό της επιστήμης των πυκνών δεδομένων-επιτρέπει στα τρία προηγούμενα να αλληλεπιδράσουν και να ενισχύσουν το ένα το άλλο όπως η θεωρία προδιέγραφε τον πειραματισμό και αντίστροφα.



Τι είναι και τι σκοπό έχει ο επιστημονικός φάκελλος (δημοσίευση) (1/2)

- Περιλαμβάνει στοιχεία ανεξάρτητες επιστημονικές εκδόσεις, παραπομπές, δεδομένα που υποστηρίζουν τα ευρήματα της δημοσίευσης.
 - Αυτός ο φάκελλος αρχειοθετείται και διανέμεται σε βιβλιοθήκες, αρχεία και μουσεία σε όλο τον κόσμο.
 - Με τον καιρό ογκώδη δεδομένα από πειράματα τηρήθηκαν σε ξεχωριστά αρχεία, φωτογραφήθηκαν ή βιντεοσκοπήθηκαν για να γίνει εφικτή η διαχείρισή τους.
 - Ο επιστημονικός φάκελλος πρώτο και κύριο στόχο έχει να επικοινωνήσει μέσα στον χώρο και στο χρόνο ευρήματα, υποθέσεις και συμπεράσματα από ένα πρόσωπο σε άλλα.
 - Έχει ακόμη σκοπό να καθιερώσει κοινή ορολογία, να συσχετίσει παρόμοιες εργασίες, να αναπτύξει επιστημονικά πεδία, να προωθήσει τη συνεργασία.
-



Τι είναι και τι σκοπό έχει ο επιστημονικός φάκελλος (δημοσίευση) (2/2)

- Πρέπει να είναι αξιόπιστος σε μικρή και μεγάλη κλίμακα.
- Ένα άλλο χαρακτηριστικό του επιστημονικού φακέλλου είναι η αναπαραγωγικότητα των αποτελεσμάτων και προς τούτο πρέπει να περιλαμβάνει λεπτομερείς περιγραφές των πειραμάτων και των εξοπλισμών.



Το τέταρτο επιστημονικό πρότυπο (1/2)

- Με την πρότυπο της υπολογιστικής πυκνών δεδομένων ο επιστημονικός φάκελλος και ο υποστηρικτικός αυτού μηχανισμός επικοινωνίας και δημοσίευσης έχει αποκτήσει μία νέα μορφή.
 - Τα δεδομένα και το λογισμικό με το οποίο παρήχθησαν είναι σημαντικά συστατικά του φακέλλου και θα πρέπει να μπορούν να οδηγήσουν στην επανάληψη των υπολογισμών. Η δημοσίευση πλέον αποτελεί ένα παράθυρο όχι μόνο για την κατανόηση των αποτελεσμάτων αλλά και για την επέκταση.
 - Θα υπάρχει συνδυασμός με σχετικές δημοσιεύσεις και άλλες συσχετίσεις σε ψηφιακό περιβάλλον.
 - Εντούτοις σημειώνονται εξελίξεις με απροσδιόριστες ακόμη συνέπειες όπως η συλλογή δεδομένων παραπομπών που γίνεται αυτόνομα, επικαιροποιείται εμπλουτίζεται με νέους υπολογισμούς και χρησιμοποιείται από νέες δημοσιεύσεις για παραπομπές.
-



Το τέταρτο επιστημονικό πρότυπο (2/2)

- Ο επιστημονικός φάκελλος πλέον αποκτά μία νέα οντότητα που μπορεί να επεκταθεί μέχρι τα όρια που βάζουν φραγμοί όπως πατέντες και δικαιώματα.
- Στο επίπεδο της υπολογιστικής των πυκνών δεδομένων η ανάγνωση των επιστημονικών φακέλλων μπορεί να γίνει σε δύο κλίμακες.
 - Τη μικρή, απλώς ανάγνωση με δυνατότητα χρήσης υπολογιστικών εργαλείων για εμφάθυνση.
 - Σε μεγάλη κλίμακα από ακαδημαϊκούς που αξιοποιούν υπολογιστικά εργαλεία που θα αναδείξουν δημοσιεύσεις με ενδιαφέρον που θα αξιολογηθούν με πειραματισμούς ή προσομοιώσεις



Τα κείμενα σε ένα κόσμο επικεντρωμένο στα δεδομένα

Paul Ginspag Cornell University



Αναζήτηση σε κείμενα και βάσεις δεδομένων

- Η αναζήτηση μέσα σε κείμενα ή βάσεις δεδομένων ξεκίνησε στα μέσα του 90 με τις μεγάλες μηχανές αναζήτησης όπως το Google και το Yahoo.
- Η αναζήτηση σε επιστημονικά κείμενα με ένα αποτελεσματικό τρόπο ήταν ένα από τα οράματα του Jim Gray στο τέλος της δεκαετίας του '90 όταν ακόμη υπήρχαν αρκετές τεχνικές δυσκολίες.
- Πολλές από τις ερωτήσεις που αφορούσαν την ακαδημαϊκή υποδομή, όπως σχετικά με τη δομή που έπρεπε να έχουν τα επιστημονικά άρθρα, θέματα ποιοτικού ελέγχου αναδύθηκαν την προηγούμενη δεκαετία χωρίς να τύχουν καθαρής απάντησης.
- Η περίπτωση της PubMed Central βάσης δεδομένων (με βιολογικά στοιχεία) όπου σε κάποιο άρθρο υπάρχουν συνδέσεις σε γονίδια, γονιδιώματα, πρωτεΐνες κλπ, σε σχετικές βάσεις δεδομένων ή άλλα άρθρα, είναι ένα χαρακτηριστικό παράδειγμα για ανάλογες περιπτώσεις για το μέλλον.
- Άλλα παραδείγματα είναι η βάση χημικών δεδομένων Molecular BioSystems όπου μαρκάρει όρους μέσα στο κείμενο και τους συνδέει με άλλες βάσεις δεδομένων που έχουν σε κατάλογο τους όρους αυτούς, η πλήρως αυτοματοποιημένη βάση EMBL.



Προοπτικές για βελτίωση των τεχνικών αναζήτησης σε κείμενα

- Σημαντικές βελτιώσεις θα μπορούσαν να γίνουν στο μέλλον, όπως η καταγραφή σε μία ετικέτα των τεχνικών όρων του άρθρου, μία στατιστική του άρθρου, ανάλυση των παραπομπών, ορολογία, συσχέτιση με άλλα ομοειδή άρθρα κλπ.
- Μεγάλη σημασία θα έχει η δόμηση των άρθρων με ψηφιακό τρόπο ώστε να μπορεί να γίνει περίληψη και κωδικοποίηση των στοιχείων τους και να διαβάζονται από μηχανές.
- Επί του παρόντος η αναγνώριση οντοτήτων σε αδόμητα κείμενα γίνεται κατευθείαν αλλά δεν προχωράει η αναζήτηση συσχετίσεων.
- Σε μελλοντικά κείμενα κατάλληλα δομημένα η αναζήτηση θα παρέχει γνώση με συσχέτιση, με ανάλυση παραπομπών, με αυτόματη ανάλυση εικόνας και με πολλές άλλες δυνατότητες που οι χρήστες θα μπορούν να αξιοποιήσουν.



Συμμετοχή εκδοτών και συγγραφέων στην προσβασιμότητα των κειμένων σε έξυπνη αναζήτηση

- Στο μέλλον τα κείμενα θα γίνουν πιο πρόσφορα σε αναζήτηση με την εφαρμογή εργαλείων από τους εκδότες και τους συγγραφείς καθώς και αυτόματων εργαλείων σε νέα ή αρχειοθετημένα κείμενα.
- Ίσως στην αρχή αυτή η βελτίωση των κειμένων να μη γίνει εύκολα αφού θα απαιτεί πρόσθετη προσπάθεια. Όμως ο ανταγωνισμός με τα κείμενα που θα έχουν αυτές τις δυνατότητες θα αναγκάσει και όσους διστάζουν να συμμορφωθούν και η τυποποίηση της απαιτούμενης προσπάθειας θα διευκολύνει σταδιακά την προσαρμογή.
- Ένα ερώτημα για το μέλλον είναι που θα φιλοξενούνται οι σημασιολογικές παρατηρήσεις και πως θα αποθηκεύονται.



Ευκολότερη πρόσβαση στο ευρύτερο κοινό

- Για να αποφύγουμε προβλήματα με δεδομένα και έγγραφα πρέπει να υπάρχει μία σημασιολογική στοιβάδα πάνω από τις βάσεις δεδομένων που θα κάνει δυνατή την πρόσβαση στα δεδομένα και από γενικού τύπου μηχανές αναζήτησης.
- Έτσι τα δεδομένα θα είναι ευρύτερα προσβάσιμα, όχι μόνο από τους πανεπιστημιακούς.
- Τέτοια συστήματα έχουν ήδη αναπτυχθεί και οι προοπτικές για την εξέλιξή τους είναι ευοίωνες.



Αναζήτηση σε κείμενα σε διάφορους επιστημονικούς τομείς

- Η εξόρυξη πληροφοριών από κείμενα και η ανάπτυξη εργαλείων επεξεργασίας φυσικών γλωσσών που αυτόματα ανακαλύπτουν οντότητες και τις συνδέουν με οντολογίες διαφόρων τομέων έχει προχωρήσει σε πολλούς επιστημονικούς τομείς, όπως στις επιστήμες ζωής, στη χημεία και τα μαθηματικά κλπ.
- Παραδόξως στην φυσική υπάρχει υστέρηση, παρόλο την πρωτοπορία του τομέα σε θέματα ψηφιακής προσαρμογής, ίσως διότι δεν υπάρχουν κυρίαρχα εργαστήρια με εξέχουσες βιβλιοθήκες που θα στήριζαν μία τέτοια προσπάθεια.
- Στο τομέα της βιολογίας η αναζήτηση σε κείμενα δεν αναμένεται να αναπτυχθεί αφού η κατεύθυνση είναι στην εξερεύνηση δεδομένων.



Δεδομένα ή κείμενα (1/2)

- Το επιστημονικό άρθρο σίγουρα θα διατηρήσει τη σημασία του στο να πείσει τον αναγνώστη για την ορθότητα της υπόθεσης που υποστηρίζει ο επιστήμονας, ενώ η πρόσβαση σε βάσεις δεδομένων θα έχει βοηθητικό-επιβεβαιωτικό ρόλο.
- Είναι γεγονός ότι η σχέση θεωρίας-δεδομένων είναι διαφορετική στους διάφορους επιστημονικούς τομείς αφού μερικοί τομείς κυριαρχούνται από τα δεδομένα (πχ η βιολογία) και άλλοι από την θεωρία (πχ τα μαθηματικά).
- Υποστηρίζεται (σε άρθρα και της παρούσης έκδοσης) ότι η μαζική παραγωγή και επεξεργασία δεδομένων θα εκτοπίσει σε πολλά σημεία την κλασική επιστημονική μεθοδολογία, άποψη που παραπέμπει σε μία μηχανιστική θεώρηση της επιστήμης.



Δεδομένα ή κείμενα (2/2)

- Η επιστήμη αξιοποιεί τις συσχετίσεις που δίνει η ανάλυση των δεδομένων για να κατανοήσει βαθύτερα τα φαινόμενα και να προβλέψει.
- Συνεπώς ούτε να υπερεκτιμούμε τον ρόλο των δεδομένων πρέπει ούτε να υποτιμούμε το ρόλο των κειμένων.
- Συμπερασματικά οι νέες εξελίξεις στο τομέα διαχείρισης δεδομένων και κειμένων, δίνει τις δυνατότητες για γρήγορη αντιμετώπιση πολλών προβλημάτων και την ταχύτατη επιστημονική πρόοδο.



“Επιβιβαστείτε” : Προς ένα φιλικό στις μηχανές σύστημα ακαδημαϊκής επικοινωνίας

Herbert Van De Sompel Los Alamos National Laboratory

Carl Lagoze Cornell University



Εισαγωγή

- Το σημερινό σύστημα ακαδημαϊκής επικοινωνίας είναι μία αντιγραφή του βασισμένου στις δημοσιεύσεις, αφού δεν εφαρμόζονται ακόμη πλήρως οι νέες ψηφιακές δυνατότητες (2009).
- Τρεις είναι οι κύριες ελλείψεις του συστήματος.
 - Συστημικές ελλείψεις.
 - Οικονομικοί περιορισμοί.
 - Τεχνικά εμπόδια που μπαίνουν στην υποδομή πληροφορίας.
- Παράλληλα ξεκίνησαν πρωτοβουλίες για την Ανοιχτή Πρόσβαση που σχετίζεται άμεσα με την ακαδημαϊκή επικοινωνία.
- Μία άλλη τάση εκδηλώθηκε προς την κατεύθυνση των μηχανικά επεξεργάσιμες πληροφορίες.



Αυξάνοντας τον ακαδημαϊκό φάκελλο με ένα μηχανικά επεξεργάσιμο υπόστρωμα

- Το πρόβλημα με την πλημμύρα των δημοσιεύσεων που είναι δύσκολα διαχειρίσιμο από τους ερευνητές βρήκε μία λύση με μηχανικούς παράγοντες που κάνουν αυτή τη δουλειά, ιδιαίτερα στις περιπτώσεις διεπιστημονικών θεμάτων.
- Ενδεικτικά στις επιστήμες ζωής τα θέματα έχουν συνάφεια με την γενετική, την βιοχημεία και την οργανική χημεία και η διερεύνηση των δημοσιεύσεων στους γειτονικούς αυτούς τομείς ενέχει δυσκολία και η βοήθεια που ήδη παρέχεται από μηχανικούς παράγοντες είναι σημαντική.
- Προηγμένες τεχνικές ανάλυσης κειμένου εντοπίζουν αυτόματα οντότητες και σχέσεις μεταξύ τους και παριστούν ομοιόμορφα επιστημονική γνώση σε βιβλιογραφία που πριν μόνο σε ανθρώπους ήταν προσβάσιμη.
- Πολλές σχετικές πρωτοβουλίες είναι σε εξέλιξη και το αποτέλεσμα τους θα βοηθήσει τόσο τους πανεπιστημιακούς δασκάλους όσο και τους φοιτητές να έρθουν σε μεγαλύτερη επαφή με τη γνώση.



Διαδικασίες και δεδομένα στον ακαδημαϊκό φάκελλο

- Μέχρι πρόσφατα σε μία δημοσίευση τα δεδομένα απλώς τεκμηριώναν την υπόθεση που διατυπωνόταν όντας σε δεύτερη μοίρα.
 - Ήδη αυτό αλλάζει και τα δεδομένα έρχονται σε πρώτη μοίρα καθώς οι πανεπιστημιακοί συζητούν και εφαρμόζουν όλες τις φάσεις: καταχώριση, πιστοποίηση, ευαισθητοποίηση, αρχειοθέτηση και ανταμοιβή.
 - Οι παραπομπές σε μία δημοσίευση παριστούν την συνδρομή υφιστάμενης.
 - Γνώσης στη δημιουργία νέας. Με αυτή την έννοια το ακαδημαϊκό γράφημα παραπομπών (δηλαδή πόσοι χρησιμοποιούν τις δικές σου δημοσιεύσεις ως παραπομπές) είναι ένας δείκτης δυναμικής και διασύνδεσης επιστημονικών πεδίων.
 - Εντούτοις το παραπάνω γράφημα μπορεί να οδηγεί σε σφάλματα, ειδικά στην αυτόματη εξαγωγή του και ενδεχομένως σε μεροληπτικά συμπεράσματα και αυτό καταδεικνύει την αμφιλεγόμενη αξία του.
-



Κοιτάζοντας προς το μέλλον

- Ο ακαδημαϊκός φάκελλος στο μέλλον θα αποτελέσει ένα συνδυασμό από τον παραδοσιακό τρόπο εργασίας με την μηχανική υποστήριξη στην αναζήτηση γνώσης και τα μεταδεδομένα.
- Για να υπάρξει πλήρης αξιοποίηση αυτών των υποστηρικτικών εργαλείων θα πρέπει η διαδικασία να βασιστεί στην κυβερνο-υποδομή που θα αναπτυχθεί στον ιστό (web) εξασφαλίζοντας έτσι βιωσιμότητα και διαφάνεια και ελεύθερη προσβασιμότητα από την επιστημονική κοινότητα.
- Ίσως σημαντικότερα από τα τεχνικά προβλήματα στην δημιουργία της κυβερνο-υποδομής (cyberinfrastructure) είναι η ενσωμάτωσή της μέσα στην κοινότητα που την χρησιμοποιεί αντιμετωπίζοντας το θέμα σαν κοινωνικό-τεχνικό πρόβλημα η συνολική επίλυση του οποίου θα μας οδηγήσει στην επόμενη φάση της ακαδημαϊκής επικοινωνίας.



Το μέλλον της πολιτικής για τα δεδομένα

Anne Fitzgerald -Brian Fitzgerald -Kylie Pappalardo

Queensland University of Technology



Εισαγωγή

- Η εποχή μας χαρακτηρίζεται από μία επανάσταση στον τρόπο συλλογής, παραγωγής, διαμοιρασμού και διανομής της πληροφορίας.
- Η ανάγκη για ελεύθερη και γρήγορη πρόσβαση σε επιστημονικά δεδομένα είναι κοινώς αποδεκτή και έχει επίσημα διατυπωθεί διεθνώς σε πολλές ευκαιρίες.
- Θα πρέπει όμως και σε εθνικό επίπεδο να εφαρμοστούν στην πράξη οι πολιτικές αυτές.
- Μία τέτοια διαπίστωση από διεθνή οργανισμό για την ανάγκη της ελεύθερης πρόσβασης σε επιστημονικά δεδομένα έγινε με το Global Earth Observation System of Systems (GEOSS) που ενεργοποιήθηκε το 2005 από το Group on Earth Observations (GEO).



Διεθνείς Πολιτικές

- Μία από τις πρώτες πράξεις του GEO ήταν η διατύπωση των αρχών του διαμοιρασμού δεδομένων και της ελεύθερης πρόσβασης σε αυτά.
- Έτσι ο GEO διατυπώνει την ανάγκη για πλήρη και ανοιχτή ανταλλαγή δεδομένων και μεταδεδομένων (επεξεργασμένων δεδομένων) με το ελάχιστο δυνατόν κόστος και στον ελάχιστο δυνατόν χρόνο με την υιοθέτηση των κατάλληλων ρυθμίσεων σε διεθνές και εθνικό επίπεδο.
- Οι αρχές αυτές της ελεύθερης πρόσβασης σε επιστημονικά δεδομένα είχαν ήδη καθιερωθεί σε παλαιότερες διεθνείς συμφωνίες, όπως οι Bermuda Principles (1996), η διακήρυξη του Βερολίνου για την ανοιχτή πρόσβαση στην γνώση Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003) αλλά και πολύ παλαιότερα με τη συνθήκη της Ανταρκτικής Antarctic Treaty (1959).
- Όλα όμως τα παραπάνω πρέπει να ενσωματωθούν στις εθνικές νομοθεσίες για να έχουμε πρακτικό αποτέλεσμα.



Εθνικές πολιτικές (1/2)

- Για να έχουμε μία συνεκτική εθνική νομοθεσία που να εγγυάται την ανοιχτή πρόσβαση στα δεδομένα θα πρέπει να υιοθετηθούν οι διεθνείς αρχές και να υπάρξει σαφής και πρακτική για το τι θα πρέπει οι ερευνητές να κάνουν σε ένα ερευνητικό σχέδιο.
- Στις ΗΠΑ και στην Ευρώπη υπάρχει το κατάλληλο για το θέμα θεσμικό πλαίσιο.
- Συγκεκριμένα στις ΗΠΑ οι εγκύκλιοι (OMB) Circular A-130 και η συμπληρωματική της Circular A-16, ρυθμίζουν λεπτομερώς τις διαδικασίες ελεύθερης και ανεμπόδιστης πρόσβασης σε κρατικές επιστημονικές πληροφορίες αλλά και προβλέπει τρόπους για την διευκόλυνση αυτής της πρόσβασης



Εθνικές πολιτικές (2/2)

- Στην Ευρώπη αντίστοιχα με τις οδηγίες για την επαναχρησιμοποίηση πληροφοριών του Δημόσιου τομέα (2003) (the PSI Directive) και επίσης την ειδική οδηγία για την δημιουργία υποδομής για τη χωρική πληροφορία (2007) INSPIRE Directive δόθηκε ικανοποιητική λύση στο θέμα.
- Ακόμη διεθνείς οργανισμοί όπως ο ΟΟΣΑ έχουν καθιερώσει σε γενικές γραμμές τις αρχές που πρέπει να διέπουν την ελεύθερη πρόσβαση σε δεδομένα και αυτό δεσμεύει τα μέλη του, ηθικά βέβαια, για να ακολουθήσουν μία ανάλογη πολιτική και σε εθνικό επίπεδο.
- Ορισμένες χώρες, όπως η Αυστραλία, δεν έχουν καθιερώσει σε εθνικό επίπεδο την κατάλληλη νομοθεσία, όπως περιγράφηκε παραπάνω και αυτό αποτελεί ένα πρόβλημα που θα πρέπει να βρει τη λύση του για να γίνει διαχειρίσιμη η πλημμυρίδα των δεδομένων.



“Είδα” την μετατόπιση του προτύπου και είμαστε εμείς

John Wilbanks Creative Common



Η αλλαγή κοσμοθεωριών και προτύπων στο σύγχρονο κόσμο

- Ο Thomas Khun στο βιβλίο του “δομή της επιστημονικής επανάστασης ” περιέγραψε γλαφυρά τις φάσεις ανάπτυξης, εμπέδωσης και κατάρπτωσης των κοσμοθεωριών και προτύπων.
- Στην πρώτη φάση ένα σύνολο ιδεών που εξηγεί με εύλογο τρόπο τα ως τότε φαινόμενα γίνεται παραδεκτό από όλους και κυρίαρχο-κοσμοθεωρία, πρότυπο- όπως πχ η θεωρία του αιθέρα, η γεωκεντρική ερμηνεία του ηλιακού συστήματος κλπ.
- Στην δεύτερη φάση όλοι σπεύδουν να ενισχύσουν την αυθεντία του προτύπου για να έχουν ηθικά ή υλικά κέρδη και αυτό δημιουργεί μία ασπίδα προστασίας από αμφισβητήσεις.
- Στην τρίτη φάση, της πτώσης, κάποια γεγονότα, φαινόμενα ή πειράματα δεν εξηγούνται πλέον με βάσει το πρότυπο που ίσχυε και έτσι έρχεται η στιγμή για μετατόπιση στο επόμενο πρότυπο ή κοσμοθεωρία. Η ύπαρξη του αιθέρα κατέρρευσε αφού ποτέ δεν αποδείχτηκε και ο γεωκεντρισμός καταρρίφθηκε από τον Γαλιλαίο.



Το τέταρτο επιστημονικό πρότυπο του Jim Gray (1/2)

- Σύμφωνα με τον Jim Gray η ικανότητά μας να διαχειριστούμε και να αξιοποιήσουμε σήμερα την πλημμύρα των δεδομένων αποτελεί το νέο, τέταρτο πρότυπο για την επιστήμη παράλληλα με τα προηγούμενα τρία, τον εμπειροκρατισμό, τη θεωρία και την προσομοίωση.
 - Αλλά δεν πρέπει να υποτιμούμε τα τρία προηγούμενα πρότυπα που έδωσαν τεράστια ώθηση στην επιστήμη και το καθένα ήταν προϋπόθεση για τα επόμενα, χωρίς να καταργείται η σημασία και η αξία του.
 - Με την έννοια που δίνει ο Khun δεν έχουμε μία πραγματική μετάβαση σε άλλο πρότυπο, απλώς τα δεδομένα αλλάζουν τον τρόπο που εφαρμόζεται ο εμπειρισμός, η θεωρία και η προσομοίωση.
 - Με την έννοια αυτή πρέπει να αλλάξουμε πρότυπο εμείς σαν επιστήμονες όχι ο τρόπος της επιστημονικής ανακάλυψης.
 - Η πληθώρα των δεδομένων η γρήγορη και αποτελεσματική αξιοποίησή τους και διάχυσή τους θα βοηθήσουν τελικά σε ανατροπές ή αλλαγές στις θεωρίες.
-



Το τέταρτο επιστημονικό πρότυπο του Jim Gray (2/2)

- Η καλύτερη διαχείριση της πλημμύρας των δεδομένων θα είναι μέσω δικτύων, όπου πολλοί επιστήμονες θα συνεργάζονται αθροίζοντας τις ικανότητές τους.
- Τα δίκτυα αυτά θα είναι ανοιχτού τύπου, δομημένα σε στοιβάδες με ίδια λογική όπως φτιάχτηκε ο ιστός και το διαδίκτυο που αποτελούν εξαιρετικά παραδείγματα επιτυχημένων δομών κυρίως λόγω της φιλοσοφίας κατασκευής και λειτουργίας τους.
- Με αυτή την λογική τα δεδομένα δεν είναι το τέταρτο πρότυπο αλλά η τέταρτη στοιβάδα πυκνών δεδομένων σε ανοιχτού τύπου δίκτυα, με ανοιχτού τύπου λογισμικό.



Ανοιχτές πηγές

- Τα κίνητρα είναι μία βασική έννοια της οικονομικής επιστήμης και βάσει αυτής η εργασία χωρίς όφελος είναι ακατανόητη.
 - Όμως αυτό συμβαίνει και τα παραδείγματα συλλογικής εργασίας στο διαδίκτυο χωρίς άμεσο κίνητρο είναι πολλά και αποτελεί ένα πραγματικό νέο πρότυπο κατά την έννοια της ανάλυσης του Khun.
 - Εκείνο που έπρεπε να μας απασχολεί είναι πια είναι τα εμπόδια προς αυτήν την κατεύθυνση και τι πρέπει να κάνουμε για την άρση αυτών των εμποδίων.
 - Εμπόδια δεν είναι μόνο αυτά που απορρέουν από τους περιορισμούς για την προστασία της πνευματικής ιδιοκτησίας αλλά και τεχνικές αιτίες, όπως η δυσκολία μαζικής αποστολής δεδομένων, επανασύνδεσης δεδομένων κλπ.
 - Με μία φράση θα μπορούσαμε να πούμε αντισταθείτε στην αντίσταση.
-



Τι πρέπει να γίνει

- Χρειάζονται επενδύσεις για να ενισχυθεί η υποδομή για την συνεργασία και τη συλλογική εργασία στο διαδίκτυο σε θέματα, όπως η ανάπτυξη ανοιχτών προδιαγραφών για τον διαμοιρασμό δεδομένων, η δυνατότητα σχολιασμού θέσεων, η αύξηση της αποθηκευτικής ικανότητας, η ικανότητα από κοινού παρατήρησης οπτικοποιημένων δεδομένων κλπ.
- Πρέπει να δικτυωθούμε και να θεωρήσουμε τον εαυτό μας σαν κόμβο για τη διαβίβαση δεδομένων, τον έλεγχο θεωριών και προσομοιώσεων
- Πρέπει πάντως να έχουμε υπόψη μας ότι υπάρχουν και περιορισμοί και ότι η δικτύωση στη γνώση είναι πιο δύσκολη από τη δικτύωση στοιχείων.
- Εντούτοις η πλημμυρίδα των δεδομένων που είναι πραγματικό γεγονός δεν αφήνει άλλα περιθώρια από τη συλλογική δουλειά που θα αθροίσει τις επιμέρους ικανότητες μεγάλου αριθμού ερευνητών.
- Τέλος η εργασία σε ένα ανοιχτό σύστημα που επιδέχεται τροποποιήσεις και βελτιώσεις σημαίνει ότι η δουλειά μας τελικά θα επιβιώσει ενώ σε ένα κλειστό σύστημα απλώς θα καταστραφεί.



Από το web 2.0 στην παγκόσμια βάση δεδομένων

Timo Hannay Nature Publishing Group



Το web 2.0 σήμερα

- Είναι πλέον κοινός τόπος ότι η έννοια του υπολογιστή είναι συνδεδεμένη με την έννοια του web 2.0 και ότι δεν είναι συνδεδεμένο στο διαδίκτυο δεν θεωρείται καν υπολογιστής.
- Η τεράστια επιτυχία του web 2.0 σε σχέση με άλλα δίκτυα έγκειται στο ότι ενσωμάτωσε μία σειρά καλών σχεδιαστικών πρακτικών, όπως η δυνατότητα δικτύωσης, η αρχιτεκτονική ενθάρρυνσης της συμμετοχής, ο ανοιχτός χαρακτήρας κλπ.
- Οι περιπτώσεις της Wikipedia και η λειτουργία του Google δείχνουν τα αποτελέσματα που μπορεί να έχει η συνεργασία και η συλλογική εργασία στο διαδίκτυο.
- Εντούτοις στον τομέα της επιστήμης υπήρξε στην αρχή μία επιφύλαξη και μία μικρότερη αξιοποίηση των δυνατοτήτων του διαδικτύου σε σχέση με άλλους τομείς και αυτό μπορεί να αποδοθεί σε τεχνικούς, ψυχολογικούς και κοινωνικούς λόγους.



Η σχέση της επιστήμης με το διαδίκτυο σήμερα

- Οι όποιες παγιωμένες συνήθειες και επιφυλάξεις που έκαναν τους επιστήμονες διστακτικούς απέναντι στο διαδίκτυο δεν μπορούν πλέον να διατηρηθούν άλλο ενόψει των νέων δεδομένων.
- Οι νέες συνθήκες που κάνουν αναπόφευκτη τη χρήση του διαδικτύου περιλαμβάνουν χαρακτηριστικά όπως, ανοιχτό και άμεσο διαμοιρασμό δεδομένων, μικρότερο ελάχιστο μέγεθος δημοσίευσης, μέτρηση παραγωγικότητας με άλλα κριτήρια από τις δημοσιεύσεις κλπ.
- Τα εργαστήρια του μέλλοντος θα είναι συντονισμένα στις νέες συνθήκες της αξιοποίησης της υπολογιστικής και του διαδικτύου.
- Ενδεικτικό είναι το παράδειγμα της χημείας όπου στο πρώτο συνεργατικό site του διαδικτύου τη Wikipedia μπορείς να βρεις λήμματα για χιλιάδες χημικές ενώσεις, ενώ σε πιο εξειδικευμένα site μπορείς να βρεις πληθώρα στοιχείων.
- Ακόμη και σε τομείς που η προστασία της πνευματικής ιδιοκτησίας είναι πρώτης σημασίας, όπως ο φαρμακευτικός τομέας, υπάρχουν περιπτώσεις ανοιχτών στο διαδίκτυο καινοτομικών πρωτοβουλιών που σχετίζονται με ερευνητικά θέματα.



Η συνεργασία και η συλλογική εργασία καταλύτρες για τη χρήση του διαδικτύου

- Η χρήση του διαδικτύου από την επιστημονική κοινότητα είναι πλέον μία αδήριτη ανάγκη λόγω της αλλαγής των συνθηκών διεξαγωγής της έρευνας.
 - Πέρασε πλέον ο καιρός του μοναχικού επιστήμονα που μόνος του αντιμετωπίζει όλα τα προβλήματα και η έρευνα πλέον από τη μικρή κλίμακα έχει περάσει στη βιομηχανική κλίμακα με καταμερισμό καθηκόντων, οργάνωση και στόχευση.
 - Η συλλογή δεδομένων αυξάνει εκθετικά σε τομείς όπως η βιολογία όπου από την εργασία της αλληλουχίας των γονιδίων περάσαμε σε αυτή των γονιδιωμάτων και φτάσαμε στα οικοσυστήματα.
 - Η τεράστια ποσότητα των δεδομένων που πρέπει να διαχειριστούν μετατρέπει το θέμα και σε ποιοτικό κόνοντας εφικτές ανακαλύψεις, λόγω ακριβώς της αλλαγής κλίμακας στην ποσότητα των δεδομένων.
 - Η μετατροπή της ποσότητας σε ποιότητα και η φράση “το πολύ είναι διαφορετικό” είναι φιλοσοφικές απόψεις που ταιριάζουν στην περίπτωση.
-



Τι να κάνουμε

- Οι νέες συνθήκες απαιτούν συγκεκριμένες δράσεις από ερευνητές και εκδότες για να γίνουν τα δεδομένα πιο χρήσιμα.
- Τέτοιες ενέργειες θα είναι η καθιέρωση πρότυπων μορφών για τα δεδομένα, παγκόσμια συμφωνημένους εντοπιστές και οντολογίες, ετικέτες σε κείμενα.
- Τα κατάλληλα δομημένα δεδομένα και κείμενα θα δώσουν μεγάλες δυνατότητες για αυτόματη εξόρυξη δεδομένων και πληροφοριών από βάσεις δεδομένων και από κείμενα.
- Η κατάλληλη μορφοποίηση των δεδομένων πρέπει να αρχίσει από την πηγή δηλαδή το εργαστήριο. Ήδη παρατηρούμε περιπτώσεις όπου αντιδραστήρια, πειράματα, και δεδομένα οργανώνονται με ένα προηγμένο πληροφοριακό σύστημα μεταβαίνοντας από το χαρτί στην οθόνη.
- Εκτός από την έκρηξη των δεδομένων παρατηρούμε και την αλληλοσυνδεσιμότητα καθώς η χρήση συνδέσμων, ετικετών και ID ταυτοτήτων δίνει τα δεδομένα που τελικά θα μας δώσει μία παγκόσμια βάση δεδομένων.



Κοιτάζοντας μπροστά (1/2)

CRAIG MUNDIE | Microsoft

- Η εφαρμογή στην πράξη του τέταρτου επιστημονικού προτύπου αυτού της επιστήμης των πυκνών δεδομένων σε συνδυασμό με τις δυνατότητες δικτύωσης μέσω του διαδικτύου θα δώσει τη δυνατότητα για ταχύτατη ανάπτυξη σε όλους τους επιστημονικούς τομείς.
- Η ευρύτατη διάδοση αυτόματων αισθητήρων σε όλη τη γη, που είναι ένα εικονικό “μακροσκόπιο” μαζί με τις νέες τεχνολογίες της υπολογιστικής και της διαχείρισης δεδομένων, θα μας δώσει τη δυνατότητα να καταλάβουμε το περιβάλλον, τη γη και να προβλέψουμε τις κλιματικές αλλαγές.
- Ραγδαίες θα είναι οι εξελίξεις στον τομέα της ιατρικής, όπου η δυνατότητα της γρήγορης ανάλυσης των γονιδίων, η απόκτηση, ανάλυση και αξιολόγηση ιατρικών δεδομένων θα μας οδηγήσει στην προσωπική ιατρική.
- Τα υπολογιστικά συστήματα θα γίνονται ολοένα πιο ικανά να συλλέξουν και να επεξεργαστούν πληροφορίες και θα απαλλάξουν τους επιστήμονες από ανιαρά έργα, ώστε να επικεντρωθούν στην ουσία της έρευνάς τους .



Κοιτάζοντας μπροστά (2/2)

CRAIG MUNDIE | Microsoft

- Υπηρεσίες μεγάλης κλίμακας θα μπορούν να φιλοξενηθούν στο νέφος κάνοντας εφικτές σύνθετες λειτουργίες όπως η οπτικοποίηση, ανάλυση και διάδραση με τα δεδομένα.
- Θα γίνει έτσι εφικτή η παγκόσμια ψηφιακή βιβλιοθήκη που θα περιλάβει όλο τον κύκλο της έρευνας από το ξεκίνημα μέχρι τη δημοσίευση.



Συμπεράσματα (1/2)

- Ήδη από τα μέσα της δεκαετίας του 90 ο πρωτοπόρος στον τομέα της πληροφορικής James Gray διείδε την τεράστια σημασία που θα έπαιζε στην ανάπτυξη της επιστήμης η επιτυχής διαχείριση του τεράστιου όγκου των δεδομένων που πλέον παράγονταν και που αργότερα χαρακτήρισε την κατάσταση αυτή ως το τέταρτο επιστημονικό πρότυπο.
- Παράλληλα και άλλοι αναγνώρισαν τον τεράστιο ρόλο της πληροφορικής στην ανάπτυξη της επιστήμης που εκτός των άλλων θα συντελούσαν στην διεπιστημονική συνεργασία και καθιέρωσαν τον όρο eScience για να δηλώσουν την σύνδεση της πληροφορικής με τις άλλες επιστήμες.
- Ο ίδιος ο James Gray καλούσε το 2007 τους οργανισμούς χρηματοδότησης να χρηματοδοτήσουν τομείς που θα επιτάχυναν την ανάπτυξη των απαραίτητων εργαλείων για την υλοποίηση των στόχων της eScience και οι απόψεις αυτές παραμένουν ακόμη και σήμερα επίκαιρες.
- Στα άρθρα που παρατέθηκαν αναδύεται η ανάγκη για στενή και ισότιμη συνεργασία των επιστημόνων του τομέα IT με τους άλλους επιστήμονες, με ιδιαίτερη έμφαση στους τομείς υγείας και περιβάλλοντος λόγω της τεράστιας σημασίας τους στην ζωή μας.



Συμπεράσματα (2/2)

- Στην παραπάνω κατεύθυνση θα είναι σημαντική η ανάπτυξη κατάλληλης κυβερνο-υποδομής (cyberinfrastructure) με κατάλληλα εργαλεία hardware, γρήγορων δικτύων, βάσεων δεδομένων, λογισμικού, ακόμη της παγκόσμιας ψηφιακής βιβλιοθήκης και με βασική φιλοσοφία την ανοιχτή επιστήμη.
- Στην προσπάθεια αυτή πρέπει να στρατευτούν η επιστημονική κοινότητα, οι οργανισμοί χρηματοδότησης, οι κυβερνήσεις και οι γονείς, διδάσκοντες και διδασκόμενοι βάζοντας σαν στόχους τη σύζευξη όλων των επιστημονικών τομέων με την πληροφορική, τη διεπιστημονική συνεργασία, την ενεργή συμμετοχή στις διεργασίες της eScience κλπ.
- Οι παραπάνω αρχές πρέπει να γίνουν όραμα για όλους μας και ο καθένας με τον τρόπο του να βοηθήσει να γίνουν πράξη για το συμφέρον της ανθρωπότητας.



Λίγα λόγια για τον James Gray

Ο James Gray γεννήθηκε το 1944 και χάθηκε στη θάλασσα το 2007. Πήρε διδακτορικό το 1969 στην πληροφορική από το Πανεπιστήμιο Berkeley και ήταν από τους πρωτοπόρους του τομέα. Εργάστηκε στον τομέα της πληροφορικής με μεγάλη επιτυχία και ως ερευνητής σε διάφορες θέσεις και από το 1995 στην Microsoft. Επικεντρώθηκε στην eScience και υπήρξε πρωτοπόρος σε πολλούς τεχνικούς τομείς. Ανέπτυξε την θεωρία για το τέταρτο πρότυπο της επιστήμης σε συνέχεια των τριών προηγούμενων του εμπειρισμού, της θεωρίας και της προσομοίωσης Τιμήθηκε με το βραβείο Turing (1998) και με άλλες διακρίσεις. Η επιστημονική κοινότητα χρωστάει πολλά στο James Gray και για την πρωτοπορία του σε πολλά τεχνικά επιτεύγματα, αλλά και για την οραματική σκέψη του.

