

Διπλωματικές εργασίες ακαδημαϊκού έτους 2016 - 17

Υπεύθυνοι καθηγητές, Αθανάσιος Τσακαλίδης, Χρήστος Μακρής

Παρουσίαση των Θεμάτων στην αίθουσα Π200

Παρασκευή 14/10/2016 16.00 - 19.00

1. Αλληλεπίδραση με conversational interface και ανάπτυξη ενός bot

Υπάρχουν πολλοί τρόποι αλληλεπίδρασης ανθρώπου υπολογιστή. Τα τελευταία χρόνια εξαιτίας της μεγάλης ανάπτυξης των chat εφαρμογών (viber, fb messenger, skype κ.α.) υπάρχει μία τάση από τους χρήστες να προτιμούν αυτού του είδους την αλληλεπίδραση μέσω ενός conversational interface. Αρκετές μεγάλες εταιρίες πληροφορικής (facebook, microsoft, skype, slack, google, amazon) έχουν φτιάξει πλατφόρμες για να υποστηρίξουν την ανάπτυξη bots που θα αλληλοεπιδρούν με τους χρήστες μέσω του εκάστοτε conversational interface.

Στόχος της διπλωματικής εργασίας είναι η δημιουργία ενός bot σαν τα ακόλουθα <https://bots.botframework.com/>.

Προαπαιτούμενα : Πολύ καλή γνώση αντικειμενοστραφούς προγραμματισμού, πολύ καλή γνώση c#, Προγραμματισμός στον παγκόσμιο Ιστό

Επιβλέπων: Αθανάσιος Τσακαλίδης

Συνεπιβλέπων: Μανόλης Βιέννας, Αγγελική Ράπτη

Επικοινωνία: biennas@ceid.upatras.gr, arapti@ceid.upatras.gr

Σχετικές Αναφορές:

- <https://dev.botframework.com/>
- <https://www.youtube.com/watch?v=spj9YOsKKXc>
- <https://messengerplatform.fb.com/>
- <https://bots.botframework.com/>
- <https://api.ai/>
- <http://venturebeat.com/2016/06/26/the-bot-revolution-how-conversational-interfaces-will-replace-apps/>
- <http://www.theverge.com/2016/10/4/13164882/google-assistant-actions-on-google-developer-sdk>
- <https://developer.amazon.com/echo>
- <https://www.youtube.com/watch?v=bglVydquFNE>
- <https://assistant.google.com/>

2. Οπτικοποίηση και αλληλεπίδραση με μεγάλο αριθμό δεδομένων

Αρκετές μεγάλες εταιρίες όπως το Spotify, το eBay και το Netflix έχουν μετατρέψει την υποδομή τους να κάνει χρήση της αρχιτεκτονικής microservices (μεταβαίνοντας από την παλαιότερη monolithic).

Στα πλαίσια της διπλωματικής αυτής θα αναπτυχθεί μία εφαρμογή με χρήση της αρχιτεκτονικής microservices και πιο συγκεκριμένα με χρήση του Azure Service Fabric.

Προαπαιτούμενα : Πολύ καλή γνώση αντικειμενοστραφούς προγραμματισμού, πολύ καλή γνώση c#, Προγραμματισμός στον παγκόσμιο Ιστό

Επιβλέπων: Αθανάσιος Τσακαλίδης

Συνεπιβλέπων: Μανόλης Βιέννας

Επικοινωνία: biennas@ceid.upatras.gr

Σχετικές Αναφορές:

- <https://en.wikipedia.org/wiki/Microservices>
- <http://martinfowler.com/microservices/>
- <https://www.infoq.com/microservices/>
- <https://azure.microsoft.com/en-us/services/service-fabric/>
- <https://azure.microsoft.com/en-us/documentation/services/service-fabric/>
- <https://www.youtube.com/watch?v=7LGPeBgNFuU>
- <https://www.infoq.com/presentations/migration-cloud-native>
- <https://cloud.docker.com/>
- <http://kubernetes.io/>
- <http://microservices.io/patterns/microservices.html>

3. Υλοποίηση Record Linkage Αλγορίθμων σε περιβάλλον spark

Ο σχεδιασμός και υλοποίηση συστημάτων λογισμικού για καταναεμημένες εφαρμογές είναι μια ιδιαίτερα σύνθετη και επίπονη διαδικασία. Βασικά χαρακτηριστικά είναι η αποδοτική επεξεργασία μεγάλου όγκου δεδομένων και η κλιμάκωση του συστήματος έτσι ώστε με την προσθήκη νέων πόρων να μπορεί να ανταπεξέλθει σε μεγαλύτερο φόρτο. Το περιβάλλον ανάπτυξης εφαρμογών Spark [1,2,3] αποτελεί την πιο γρήγορη προσέγγιση στο μοντέλο Map/Reduce σε κύρια (για πρώτη φορά) και δευτερεύουσα μνήμη και προσφέρει ιδιαίτερα καινοτόμες διαδικασίες και μεθόδους που επιτρέπουν την ανάπτυξη εφαρμογών που επιτυγχάνουν μεγάλη κλιμάκωση και ταχύτατη επεξεργασία δεδομένων μεγάλου όγκου.

Στα πλαίσια της διπλωματικής αυτής ζητούνται:

- Η μελέτη του περιβάλλοντος Spark σε γλώσσα προγραμματισμού Scala.
- Η υλοποίηση Record Linkage Αλγορίθμων [4,5,6,7,8] σε γλώσσα Scala.

Η εκτενής πειραματική αξιολόγηση της παραπάνω υλοποίησης

Επιβλέποντες: Α. Τσακαλίδης, Σ. Σιούτας, { sioutas@ceid.upatras.gr }

Ενδεικτική Βιβλιογραφία

- [1] <http://spark.apache.org/>
- [2] <http://simin.me/projects/spatialspark/>
- [3] <http://spark-summit.org/east>
- [4] https://en.wikipedia.org/wiki/Bloom_filter
- [5] <http://billmill.org/bloomfilter-tutorial/>
- [6] <http://prakhhar.me/articles/bloom-filters-for-dummies/>

[7] <http://infolab.stanford.edu/~ullman/mmds/book.pdf>

[8] <http://ceur-ws.org/Vol-1330/paper-04.pdf>

4. Ανάλυση μεγάλου όγκου Δεδομένων της HBase με χρήση του περιβάλλοντος ανάπτυξης εφαρμογών spark

Ο σχεδιασμός και υλοποίηση συστημάτων λογισμικού για κατανεμημένες εφαρμογές είναι μια ιδιαίτερα σύνθετη και επίπονη διαδικασία. Βασικά χαρακτηριστικά είναι η αποδοτική επεξεργασία μεγάλου όγκου δεδομένων και η κλιμάκωση του συστήματος έτσι ώστε με την προσθήκη νέων πόρων να μπορεί να ανταπεξέλθει σε μεγαλύτερο φόρτο. Το περιβάλλον ανάπτυξης εφαρμογών Spark [1,2,3,4] αποτελεί την πιο γρήγορη προσέγγιση στο μοντέλο Map/Reduce σε κύρια (για πρώτη φορά) και δευτερεύουσα μνήμη και προσφέρει ιδιαίτερα καινοτόμες διαδικασίες και μεθόδους που επιτρέπουν την ανάπτυξη εφαρμογών που επιτυγχάνουν μεγάλη κλιμάκωση και ταχύτατη επεξεργασία δεδομένων μεγάλου όγκου.

Στα πλαίσια της διπλωματικής αυτής ζητούνται:

- Η μελέτη του περιβάλλοντος Spark σε γλώσσα προγραμματισμού Java, Python ή Scala.
- Η μελέτη των τεχνικών διασύνδεσης της spark με την HBase [5].
- Η υλοποίηση αλγορίθμων ταξινόμησης (classification) και συσταδοποίησης (clustering) στο περιβάλλον spark με δεδομένα που είναι αποθηκευμένα σε πίνακες της HBase.
- Η εκτενής πειραματική αξιολόγηση της παραπάνω υλοποίησης.

Επιβλέποντες: Α. Τσακαλίδης, Σ. Σιούτας, { sioutas@ceid.upatras.gr }

Ενδεικτική Βιβλιογραφία

[1] <http://spark.apache.org/>

[2] <http://simin.me/projects/spatialspark/>

[3] <http://spark-summit.org/east>

[4] <http://spark.apache.org/docs/latest/programming-guide.html>

[5] <http://hortonworks.com/blog/spark-hbase-dataframe-based-hbase-connector/>

5. Επεξεργασία δεδομένων από δορυφορικές εικόνες για συσχέτιση με οικονομικούς δείκτες

Οι ερευνητές αναζητούν νέες μεθόδους για την εξαγωγή στατιστικών δεδομένων από πληθυσμούς χωρίς να χρειαστεί να καταφεύγουν στις επίπονες και ακριβές διαδικασίες των ερωτηματολογίων. Στο πρόβλημα έρχεται να συμβάλει η επεξεργασία δορυφορικών εικόνων.

Θα γίνει χρήση δεδομένων από δορυφορικές εικόνες και θα γίνει συσχέτιση με βάση δεδομένων από οικονομικούς δείκτες. Στόχος είναι η δημιουργία εργαλείων για την απευθείας εξαγωγή των οικονομικών δεικτών από τις εικόνες.

Θα απαιτηθεί η εξαγωγή χαρακτηριστικών και η ταξινόμησή / κωδικοποίηση με χρήση deep learning Neural Networks.

Από την εργασία αναμένεται η εξαγωγή δημοσιεύσιμων αποτελεσμάτων.

Επιβλέποντες: Α. Τσακαλίδης, Δ. Κοσμόπουλος, { dkosmo@upatras.gr }

Απαιτούμενες Γνώσεις:

C/Matlab, αναγνώριση προτύπων, νευρωνικά δίκτυα, επεξεργασία εικόνας/σήματος

6. Επεξεργασία εικόνων/βίντεο για την βελτίωση της θέασης από άτομα με προβλήματα αχρωματοψίας

Άτομα που δε μπορούν να διακρίνουν χρώματα λόγω αχρωματοψίας χρειάζονται μια μέθοδο που θα μετατρέπει τις εικόνες/βίντεο σε κάποια γειτονική χρωματική περιοχή που θα μπορούν να αντιλαμβάνονται χωρίς ωστόσο να αλλάζει το σημασιολογικό περιεχόμενο. Στόχος της εργασίας είναι η δημιουργία εργαλείου που θα κάνει αυτή ακριβώς την εργασία.

Θα απαιτηθεί η κατάτμηση και η εξαγωγή χρώματος από κάθε τμήμα. Στη συνέχεια γειτονικά τμήματα θα αξιολογούνται ως προς την εγγύτητά τους και την χρωματική τους ομοιότητα για άτομα με αχρωματοψία. Τέλος με βάση μια μέθοδο βελτιστοποίησης θα γίνεται αντιστοίχιση των χρωμάτων της αρχικής εικόνας στην τελική που θα είναι ορατή στους έχοντες αχρωματοψία.

Από την εργασία αναμένεται η εξαγωγή δημοσιεύσιμων αποτελεσμάτων.

Επιβλέποντες: Α. Τσακαλίδης, Δ. Κοσμόπουλος, { dkosmo@upatras.gr }

Απαιτούμενες Γνώσεις:

C ή Matlab, επεξεργασία εικόνας/βίντεο, τεχνικές βελτιστοποίησης

7. Επεξεργασία βίντεο αποκρίσεων του ανθρώπινου εγκεφάλου για την αναγνώριση νόσου alzheimer

Η νόσος του Alzheimer είναι μια από τις πλέον συνήθεις εκφυλιστικές νόσους του νευρικού συστήματος, τα αίτια της οποίας είναι ελάχιστα γνωστά.

Σε μια προσπάθεια μελέτης της νόσου συγκεντρώσαμε δεδομένα από εγκεφαλικά ηλεκτρόδια, τα οποία μετρούν τα ηλεκτρικά δυναμικά του εγκεφάλου δημιουργώντας σχετικά βίντεο. Στόχος της μελέτης είναι η κατάταξη των ασθενών σε κατηγορίες ανάλογα με το στάδιο της νόσου.

Για το σκοπό αυτό θα χρησιμοποιηθούν conditional random fields για την μάθηση και ταξινόμηση των σχετικών βίντεο

Επιβλέποντες: Α. Τσακαλίδης, Δ. Κοσμόπουλος, { dkosmo@upatras.gr }

Απαιτούμενες Γνώσεις:

C/Matlab, επεξεργασία σήματος, νευρωνικά δίκτυα, αναγνώριση προτύπων

8. Ανάλυση σημάτων ηλεκτροεγκεφαλογραφήματος (EEG)

Οι συσκευές ηλεκτροεγκεφαλογραφήματος (EEG) μας δίνουν τη δυνατότητα της καταγραφής της ηλεκτρικής δραστηριότητας του εγκεφάλου. Για την καταγραφή της δραστηριότητας του εγκεφάλου τοποθετούνται ηλεκτρόδια στο δέρμα της κεφαλής σε διαφορετικές θέσεις. Η ανάλυση σημάτων EEG χρησιμοποιείται συνήθως για τη διάγνωση ασθενειών όπως επιληψία, διαταραχές ύπνου κ.λπ. Πολύ συχνά χρησιμοποιούνται όμως και για τη μελέτη της λειτουργίας του εγκεφάλου στην παρουσίαση ερεθισμάτων (οπτικά, ακουστικά, σωματοαισθητικά).

Στα πλαίσια της διπλωματικής αυτής ζητούνται:

- Η μελέτη του της λειτουργίας της συσκευής EEG Emotiv EPOC+ και του SDK
- Η συλλογή δεδομένων από ικανό δείγμα στην οπτική διέγερση διαφορετικών χρωμάτων
- Η συλλογή δεδομένων από ικανό δείγμα στην ακουστική διέγερση διαφορετικών ήχων

- Ανάλυση των δεδομένων και εξαγωγή εγγενών χαρακτηριστικών

Επιβλέποντες: Α. Τσακαλίδης, Α. Φωκά (foka@ceid.upatras.gr)

Ενδεικτική Βιβλιογραφία

- [1] <https://www.emotiv.com/epoc/>
 [2] <https://github.com/Emotiv/community-sdk>

9. Ανίχνευση γεγονότων στο Twitter

Η πληθώρα δημοσιεύσεων στο twitter μπορεί να χρησιμοποιηθεί για να εντοπίσει σημαντικά γεγονότα που συμβαίνουν σε κάποιο τομέα (πολιτική, αθλητικά, κ.λπ.) σε πραγματικό χρόνο, στο μέλλον αλλά και τον απόηχο γεγονότων στο άμεσο παρελθόν.

Στα πλαίσια της διπλωματικής αυτής ζητούνται:

- Η μελέτη του Twitter API
- Η μελέτη βιβλιοθηκών για την επεξεργασία δεδομένων από το Twitter: CMUTweetTagger, Tweet4J, Tweepy, twitter-text-python, Tweet NLP
- Η συλλογή δεδομένων από το twitter και επεξεργασία τους για την ανίχνευση γεγονότων
- Η αξιολόγηση των ανιχνευμένων γεγονότων

Επιβλέποντες: Α. Τσακαλίδης, Α. Φωκά (foka@ceid.upatras.gr)

Ενδεικτική Βιβλιογραφία

- [1] <https://dev.twitter.com/rest/public>
 [2] <https://github.com/ianozsvald/ark-tweet-nlp-python>
 [3] <http://twitter4j.org/en/>
 [4] <http://www.tweepy.org/>
 [2] <https://github.com/edburnett/twitter-text-python>
 [3] <http://www.cs.cmu.edu/~ark/TweetNLP/>

10. Πρόβλεψη από δεδομένα social media

Αναζήτηση παγκόσμιων τάσεων και συναισθημάτων από πολυμέσα που διαμοιράζονται σε πλατφόρμες κοινωνικής δικτύωσης. Κάθε φορά που μια εικόνα ή βίντεο αναρτάται ή προβάλλεται σε ένα μέσο κοινωνικής δικτύωσης έμμεσα δίνει μια «ψήφο» υπέρ ή κατά του θέματος που περιλαμβάνει. Η «ψήφος» εμπεριέχει δεδομένα όπως τον χρόνο και την τοποθεσία. Η συσχέτιση αυτών των δεδομένων από εκατομμύρια χρήστες του διαδικτύου μπορούν να αποκαλύψουν τάσεις σε τομείς όπως την πολιτική, την οικονομία και το μάρκετινγκ. Αρχικά θα εξαχθούν μεταδεδομένα από εικόνες στο Flickr καθώς και στοιχεία ομοιότητας εικόνων. Στη συνέχεια θα εφαρμοστούν μοντέλα πρόβλεψης όπως Autoregressive Moving Average (ARMA) και Bass Model.

Επιβλέποντες: Α. Τσακαλίδης, Α. Φωκά (foka@ceid.upatras.gr)

Ενδεικτική Βιβλιογραφία

- [1] Jin X, Gallagher A, Cao L, Luo J, Han J (2010) The wisdom of social multimedia: using Flickr for prediction and forecast. In Proceedings of ACM Multimedia
 [2] Roy SD, Mei T, Zeng W, Li S (2013) Towards cross-domain learning for social video popularity prediction. IEEE Trans Multimedia 15(1255-1267).

[3] Roy, S.D.; Tao Mei; Wenjun Zeng; Shipeng Li, "Empowering Cross-Domain Internet Media with Real-Time Topic Learning from Social Streams," in Multimedia and Expo (ICME), 2012 IEEE International Conference on , vol., no., pp.49-54, 9-13 July 2012

11. Υλοποίηση συναρτήσεων μοντελοποίησης ανάπτυξης φυτών σε γλώσσα προγραμματισμού MATLAB.

Περιγραφή

Η τεχνολογία, τα εργαλεία και οι μέθοδοι μοντελοποίησης φυτών παρουσιάζουν ραγδαία ανάπτυξη και η μελλοντική εξέλιξη και προοπτική τους θα εστιάζεται στην ανάπτυξη μοντέλων που θα βασίζονται σε χημικούς (ενδογενείς ή εξωγενείς) και περιβαλλοντικούς (φως, θερμοκρασία κλπ) παράγοντες. Σκοπός της παρούσας διπλωματικής εργασίας είναι η υλοποίηση συναρτήσεων μοντελοποίησης των πολύπλοκων μηχανισμών της ανάπτυξης των φυτών σε γλώσσα προγραμματισμού MATLAB.

Επιβλέποντες: Α. Τσακαλίδης, Λ. Λιόπα aliopa@teiwest.gr aliopa3@gmail.com

Σχετικές Αναφορές:

- Kupchik, Matthew J., and Richard F. Shaw. "Age, growth, and recruitment of larval and early juvenile Atlantic croaker (*Micropogonias undulatus*), determined from analysis of otolith microstructure." *Fishery Bulletin* 114, no. 1 (2016).
- Tjørve, Even, and Kathleen MC Tjørve. "A unified approach to the Richards-model family for use in growth analyses: why we need only two model forms." *Journal of theoretical biology* 267, no. 3 (2010): 417-425.
- Yin, Xinyou, J. A. N. Goudriaan, Egbert A. Lantinga, J. A. N. Vos, and Huub J. Spiertz. "A flexible sigmoid function of determinate growth." *Annals of botany* 91, no. 3 (2003): 361-371.
- Tsoularis, A., and J. Wallace. "Analysis of logistic growth models." *Mathematical biosciences* 179, no. 1 (2002): 21-55.
- Hunt, R., D. R. Causton, B. Shipley, and A. P. Askew. "A modern tool for classical plant growth analysis." *Annals of Botany* 90, no. 4 (2002): 485-488.
- Richards, F. J. "A flexible growth function for empirical use." *Journal of experimental Botany* 10, no. 2 (1959): 290-301.

12. Μελέτη συσχέτισης χρηστών με events του Twitter

Στη διπλωματική αυτή μελετώνται τα χαρακτηριστικά που επηρεάζουν τους χρήστες με σκοπό την αλληλεπίδρασή τους με events που διοργανώνονται και γίνονται γνωστά μέσω του Twitter σε πραγματικό χρόνο. Με χρήση τεχνικών machine learning και εξαγωγής των χαρακτηριστικών του χρήστη με βάση τη συμπεριφορά του στο κοινωνικό δίκτυο αλλά και το δίκτυο των φίλων που έχει, μπορούμε να προβλέψουμε το αν θα λάβει μέρος και θα συμμετέχει σε δραστηριότητες που θα του προταθούν.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Αναφορές:

1. Yuheng Hu, Shelly Farnham, Kartik Talamadupula. Predicting User Engagement on Twitter with Real-World Events
2. C.J. Hutto, Sarita Yardi, Eric Gilbert. A longitudinal study of follow predictors on twitter

13. Αναπαράσταση κειμένων με χρήση γράφων

Στόχος της εργασίας είναι η μελέτη διαφόρων γραφο-θεωρητικών μοντέλων αναπαράστασης κειμένου και η ακόλουθη εφαρμογή τους σε συλλογές κειμένων. Παράδειγμα τέτοιων εφαρμογών αποτελεί η ομαδοποίηση και ταξινόμηση κειμένων, η εξαγωγή κυρίων όρων και λέξεων-κλειδιών, κλπ.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Αναφορές:

1. Charu C. Aggarwal, Peixiang Zhao. Towards graphical models for text processing.
2. Wei Jin, Rohini K. Srihari. Graph-based text representation and knowledge discovery.

14. Εξόρυξη γνώσης με σκοπό τον εντοπισμό της αρχικής πηγής της είδησης και των ακόλουθων πηγών

Στόχος της εργασίας είναι η μελέτη διαφόρων τεχνικών που αφορούν την αναπαραγωγή μιας είδησης από πολλές ειδησεογραφικές πηγές τόσο στο Διαδίκτυο όσο και στα κοινωνικά δίκτυα. Ακολούθως να εντοπιστούν οι πηγές αυτές που δημοσιεύουν πρώτες και έπειτα οι πηγές που αναπαράγουν την είδηση αυτή με σκοπό την ταξινόμηση των πηγών ανά επιθυμητή κατηγορία ή το συσχετισμό πηγών που φιλοξενούν ίδιο περιεχόμενο μεταξύ τους. Αντικείμενο της μελέτης αποτελεί επίσης ο αντίκτυπος και το αν και σε ποιο βαθμό συμβάλλει η φήμη ή η επιρροή της αντίστοιχης πηγής. Αυτό μας επιτρέπει να κρίνουμε αυτόματα την ποιότητα και την ακρίβεια του περιεχομένου με σκοπό την πρόβλεψη των μελλοντικών τάσεων με μεγαλύτερη ακρίβεια.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Αναφορές:

1. William Dolan, Chris Quirk, and Chris Brockett. Unsupervised Construction of Large Paraphrase Corpora: Exploiting Massively Parallel News Sources

2. Roja Bandari, Sitaram Asur, Bernardo A. Huberman. The Pulse of News in Social Media: Forecasting Popularity
3. Julio Reis, Fabrício Benevenuto, Pedro Olmo, Raquel Prates, Haewoon Kwak, Jisun An. Breaking the News: First Impressions Matter on Online News

15. Αναγνώριση φημών σε κοινωνικά δίκτυα

Η αξιοπιστία των ειδήσεων και η αντίστοιχη αναγνώριση φημών αποτελεί ένα πολύ σημαντικό αντικείμενο μελέτης στα κοινωνικά δίκτυα. Παρόλο που έρευνες έχουν αποδείξει ότι τα μηνύματα και οι ειδήσεις που δημοσιοποιούνται μέσω κοινωνικών δικτύων δεν στερούνται αξιοπιστίας, εντούτοις εγκυμονεί πάντα η πιθανότητα παραπληροφόρησης.

Στόχος της εργασίας είναι η μελέτη διαφόρων τεχνικών αναγνώρισης τέτοιου είδους ειδήσεων τόσο με χρήση τεχνικών μηχανικής μάθησης όσο και με βάση τον τρόπο με τον οποίο διαχέονται τέτοιες πληροφορίες μέσα σ' ένα δίκτυο.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Αναφορές:

1. Carlos Castillo, Marcelo Mendoza, Barbara Poblete. Information Credibility on Twitter
2. Arkaitz Zubiaga, Maria Liakata, Rob Procter, Kalina Bontcheva, Peter Tolmie. Towards Detecting Rumours in Social Media

16. Big Data Analytics με σκοπό την ανάλυση συναισθηματικού περιεχομένου συγκεκριμένων προϊόντων

Στόχος της εργασίας είναι η δημιουργία ενός ακατέργαστου συνόλου δεδομένων από κάποιο κοινωνικό δίκτυο και η περαιτέρω συναισθηματική ανάλυσή του. Τα αρχικά κείμενα θα είναι σε αδόμητη μορφή JSON και ως εκ τούτου θα χρησιμοποιηθεί μια NoSQL βάση δεδομένων (π.χ. MongoDB) για την αποθήκευσή τους. Για την ανάλυση των κειμένων θα χρησιμοποιηθούν NLP πακέτα, όπως το Apache OpenNLP. Επίσης μπορούν να χρησιμοποιηθούν διάφορες εικόνες με χρήση του Google API.

Τελικά θα δημιουργηθεί μια φιλική προς το χρήστη, εφαρμογή η οποία μπορεί να επεκταθεί σε μεγαλύτερη κλίμακα, προσθέτοντας περισσότερες δυνατότητες για να υποστηρίξει μια λεπτομερή ανασκόπηση ενός συγκεκριμένου προϊόντος.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Αναφορές:

1. Stavros Anastasios Iakovou, Andreas Kanavos, Athanasios K. Tsakalidis. Customer Behaviour Analysis for Recommendation of Supermarket Ware. AIAI 2016.
2. <https://arxiv.org/ftp/arxiv/papers/1601/1601.06971.pdf>
3. <http://ceur-ws.org/Vol-1096/paper1.pdf>

17. Big Data Analytics σε μετα-δεδομένα προϊόντων του Amazon

Σκοπός της εργασίας είναι η χρήση του Hadoop με σκοπό την εκτέλεση analytics σε μετα-δεδομένα προϊόντων του Amazon. Το σύνολο δεδομένων παρέχει πληροφορίες όπως το προϊόν, η τιμή, ο χρήστης που το αγόρασε, η βαθμολογία, η ώρα, κλπ. Ως analytics μπορούμε να εξάγουμε τα προϊόντα με την υψηλότερη και χαμηλότερη βαθμολογία, τους ενεργούς πελάτες, κλπ.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Αναφορές:

1. Stavros Anastasios Iakovou, Andreas Kanavos, Athanasios K. Tsakalidis. Customer Behaviour Analysis for Recommendation of Supermarket Ware. AIAI 2016.
2. <https://snap.stanford.edu/data/web-Amazon.html>
3. <https://webscope.sandbox.yahoo.com/catalog.php?datatype=a>

18. Αξιολόγηση Big Data τεχνολογιών

Στόχος της εργασίας είναι η αξιολόγηση διάφορων πλατφόρμων επεξεργασίας μεγάλου όγκου δεδομένων σε διάφορα θεματικά πεδία, όπως:

1. Αλγόριθμοι επεξεργασίας γράφων
2. Αλγόριθμοι εξόρυξης γνώσης
3. Αλγόριθμοι επεξεργασίας κειμένων

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. <http://giraph.apache.org/>
2. <http://storm.apache.org/>
3. <http://hadoop.apache.org/>
4. <https://spark.apache.org/>
5. <https://flink.apache.org/>

19. Εύρεση κοινοτήτων σε πολυδιάστατα κοινωνικά δίκτυα

Ως πολυδιάστατο δίκτυο θεωρείται ένα δίκτυο που αποτελείται από κόμβους και ακμές πολλών ειδών. Σκοπός της εργασίας είναι η εύρεση κοινοτήτων που θα επιτρέψει την εξαγωγή χρήσιμων συμπερασμάτων από πολυδιάστατα δίκτυα. Η αξιολόγηση της εργασίας θα γίνει τόσο με πραγματικά όσο και με συνθετικά δεδομένα.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. V. D.Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre. Fast unfolding of community hierarchies in large networks.
2. S. Fortunato. Community detection in graphs.
3. A. Clauset, M. E. J. Newman and C. Moore. Finding community structure in very large networks.
4. <https://snap.stanford.edu/data/higgs-twitter.html>

20. Εξόρυξη “σημαντικών” υπο-δικτύων από κοινωνικά δίκτυα

Στη διπλωματική αυτή θα ερευνηθούν μέθοδοι εξόρυξης γνώσης από κοινωνικά δίκτυα (Facebook, Twitter, Linkedin). Το πρόβλημα που θα εστιάσουμε είναι η εξόρυξη ενός αριθμού χρηστών, που σχηματίζουν μεταξύ τους υπο-δίκτυο, και που να έχουν κάποια συγκεκριμένα γνωρίσματα ώστε το τελικό υπο-δίκτυο να θεωρείται efficient για μεταφορά κάποιου μηνύματος σε όλο το εύρος του συνολικού δικτύου.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre. Fast unfolding of community hierarchies in large networks.
2. S. Fortunato. Community detection in graphs.
3. A. Clauset, M. E. J. Newman and C. Moore. Finding community structure in very large networks.

21. Χαρακτηρισμός ακμών σε κοινωνικά δίκτυα

Στα κοινωνικά δίκτυα, η φύση των συνδέσμων μεταξύ των κόμβων είναι τις περισσότερες φορές αδιευκρίνιστη. Για παράδειγμα, δεν είναι γνωστό αν δύο άτομα-κόμβοι που συνδέονται είναι φίλοι, συνεργάτες, ή ακόμα έχουν σχέση προϊσταμένου-υφισταμένου. Έτσι αναδύεται το πρόβλημα του αυτόματου χαρακτηρισμού των κόμβων. Σε αυτή την κατεύθυνση μπορούν να χρησιμοποιηθούν τεχνικές μηχανικής μάθησης.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. Sloan L., Morgan J., Housley W., Williams M., Edwards A., Burnap P. and Rana O. Knowing The Twitters - Deriving Sociologically Relevant Demographics From Twitter
2. Priedhorsky R., Culotta A. and Del Valle L. Inferring the Origin Locations of Tweets with Quantitative Confidence
- 3.

22. HealthCare Analytics με χρήση του Hadoop

Στην εποχή μας, οι περισσότεροι από τους οργανισμούς σε όλο τον κόσμο πρέπει να συνεργαστούν με τεράστια σύνολα δεδομένων με σκοπό την εξαγωγή analytics. Μέρα με τη μέρα, ο όγκος των δεδομένων που είναι αποθηκευμένα αυξάνεται με εκθετικούς ρυθμούς. Επιπλέον η αποθήκευση τεράστιων συνόλων δεδομένων και η ακόλουθη επεξεργασία τους με χρήση των παραδοσιακή RDBMS είναι εξαιρετικά χρονοβόρα. Το Hadoop ξεπερνάει τους περιορισμούς των υφιστάμενων RDBMS και παρέχει απλοποιημένα εργαλεία για την αποθήκευση δεδομένων καθώς και ταχύτερους χρόνους επεξεργασίας για την ανάλυση των δεδομένων.

Αυτή η εργασία έχει ως σκοπό την επεξεργασία δεδομένων υγειονομικής περίθαλψης με χρήση του Hadoop. Τα σύνολα δεδομένων μπορούν να αποτελούνται από εκατομμύρια ασθενείς, με στοιχεία όπως το όνομα, την ηλικία, τη διεύθυνση, το φύλο, πιθανές ασθένειες, φάρμακα, κλπ.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Βιοπληροφορική, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. <http://algo12.fri.uni-lj.si/reg/proc/presentations/ibUokiJBCgToSuThLVicuQOCO.pdf>
2. https://www.informatik.hu-berlin.de/de/forschung/gebiete/wbi/research/publications/2012/dmc_in_ngs.pdf

23. Υλοποίηση διαδικτυακού εργαλείου για την εύρεση «βιοδεικτών ασθενείας» (disease biomarkers) με τη μορφή διαφορικά εκφρασμένων υπογράφων μέσα σε βιολογικούς γράφους

Στην εποχή που διανύουμε, όπου ακμάζει η Συστημική Βιολογία (Systems Biology) και οι προεκτάσεις της, συναντάται έντονο ερευνητικό ενδιαφέρον στην εύρεση «υπογράφων» οι οποίοι διαταράσσονται (perturbed subgraphs) σε μια μελέτη γονιδιακής έκφρασης για μία ασθένεια. Οι συγκεκριμένοι υπογράφοι συνήθως θεωρούνται ως «υποψήφιοι βιοδείκτες ασθενειών» (potential disease biomarkers). Σκοπός της διπλωματικής είναι να υλοποιηθεί ένα διαδικτυακό εργαλείο, το οποίο θα δέχεται ως είσοδο γονιδιακές εκφράσεις (αρχείο .xls ή .txt) και θα εφαρμόζει τον αλγόριθμο DEsubs (Vrahatis et al., 2016), ο οποίος έχει υλοποιηθεί σε γλώσσα R. Η διπλωματική στοχεύει στην κατασκευή ενός φιλικού προς το χρήστη (user-friendly) διαδικτυακού εργαλείου (Sales et al., 2013), το οποίο θα παρέχει και πολλές επιλογές οπτικοποίησης των αποτελεσμάτων. Το θέμα δεν απαιτεί γνώσεις Βιοπληροφορικής ή έρευνα πάνω στον αλγόριθμο, αλλά περιορίζεται μόνο στην αποτελεσματική κατασκευή του διαδικτυακού εργαλείου.

Προαπαιτούμενα: Γνώση Προγραμματισμού (PHP, HTML, CSS)

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. Vrahatis, A. G., Balomenos, P., Tsakalidis, A. K., & Bezerianos, A. (2016). DEsubs: an R package for flexible identification of differentially expressed subpathways using RNA-seq experiments. *Bioinformatics*, btw544.
2. Sales, G., Calura, E., Martini, P., & Romualdi, C. (2013). Graphite Web: Web tool for gene set analysis exploiting pathway topology. *Nucleic acids research*, gkt386.

24. Σχεδιασμός και ανάπτυξη εργαλείου ομαδοποίησης σε Βιολογικούς γράφους με χρήση της πλατφόρμας Cytoscape

Το 'Cytoscape' είναι η πιο καθιερωμένη πλατφόρμα (ανοιχτής πρόσβασης) για σχεδιασμό και οπτικοποίηση πολύπλοκων βιολογικών γράφων. Το εργαλείο αυτό δέχεται επεκτάσεις (plugins) με τις οποίες μπορούν να ενσωματωθούν αλγόριθμοι και μεθοδολογίες που αφορούν την ανάλυση βιολογικών δικτύων. Σκοπός της διπλωματικής είναι να κατασκευαστεί ο αλγόριθμος ομαδοποίησης DMSP (Maraziotis et al., 2007), ο οποίος είναι υλοποιημένος σε γλώσσα Matlab, ως plugin στο Cytoscape. Ο αλγόριθμος DMSP έχει κατασκευαστεί για την εύρεση «σημαντικών υποδικτύων» μέσα σε ένα πρωτεϊνικό δίκτυο. Το θέμα δεν απαιτεί γνώσεις Βιοπληροφορικής ή έρευνα πάνω στον αλγόριθμο, αλλά περιορίζεται μόνο στην αποτελεσματική κατασκευή του plugin.

Προαπαιτούμενα: Γνώση Προγραμματισμού (Java)

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. Maraziotis, I. A., Dimitrakopoulou, K., & Bezerianos, A. (2007). Growing functional modules from a seed protein via integration of protein interaction and gene expression data. *Bmc Bioinformatics*, 8(1)
2. http://www.cytoscape.org/manual/Cytoscape2_6Manual.pdf

25. Μοντελοποίηση και Ανάλυση Βιολογικών γράφων με χρήση της Διαδικτυακής εφαρμογής Rshiny

Το rshiny (<http://shiny.rstudio.com/>) είναι ένα διαδικτυακό πλαίσιο εφαρμογής για τη γλώσσα R. Είναι απλό στη χρήση του και δεν απαιτεί γνώσεις HTML, CSS, ή JavaScript. Σκοπός της διπλωματικής είναι να χρησιμοποιηθούν κομμάτια από έτοιμα πακέτα σε γλώσσα R (Vrahatis et al., 2016a; Vrahatis et al., 2016b) και να κατασκευαστεί ένα rshiny εργαλείο, το οποίο θα μοντελοποιεί, θα οπτικοποιεί και θα κάνει πολλών ειδών αναλύσεις πάνω σε βιολογικά δίκτυα (σηματοδοτικά, μεταβολικά κτλ.). Το θέμα δεν απαιτεί γνώσεις Βιοπληροφορικής ή έρευνα πάνω στον αλγόριθμο, αλλά περιορίζεται μόνο στην αποτελεσματική κατασκευή του εργαλείου.

Προαπαιτούμενα: Γνώση Προγραμματισμού (κυρίως σε web εφαρμογές)

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. Vrahatis, A. G., Dimitrakopoulou, K., Balomenos, P., Tsakalidis, A. K., & Bezerianos, A. (2016a). CHRONOS: a time-varying method for microRNA-mediated subpathway enrichment analysis. *Bioinformatics*, 32(6), 884-892.
2. Vrahatis, A. G., Balomenos, P., Tsakalidis, A. K., & Bezerianos, A. (2016b). DEsubs: an R package for flexible identification of differentially expressed subpathways using RNA-seq experiments. *Bioinformatics*, btw544.
3. <http://shiny.rstudio.com/>

26. Εξαγωγή τοπολογικών και λειτουργικών χαρακτηριστικών σε Βιολογικά Δίκτυα

Ο χαρακτηρισμός των κόμβων (ή γκρουπ κόμβων) σε ένα βιολογικό δίκτυο με βάση τη τοπολογία του ή το λειτουργικό του ρόλο, δίνει σημαντική πληροφορία τόσο για τις «τοπικές» λειτουργίες όσο και για τη συνολική εικόνα του δικτύου. Σκοπός της διπλωματικής είναι (α) να γίνει ερευνητική μελέτη για το συγκεκριμένο θέμα και να καταγραφούν οι υπάρχουσες σχετικές μετρικές και (β) να προταθούν είτε νέες μετρικές ή βελτιώσεις των όσων υπάρχουν στη βιβλιογραφία.

Προαπαιτούμενα: Βασικές Γνώσεις στη θεωρία γράφων και στον Προγραμματισμό

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. Vrahatis, A. G., Dimitrakopoulou, K., Balomenos, P., Tsakalidis, A. K., & Bezerianos, A. (2016a). CHRONOS: a time-varying method for microRNA-mediated subpathway enrichment analysis. *Bioinformatics*, 32(6), 884-892.
2. Mason, O., & Verwoerd, M. (2007). Graph theory and networks in biology. *IET systems biology*, 1(2), 89-119.
3. Xu, J., & Li, Y. (2006). Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics*, 22(22), 2800-2805.

27. Υλοποίηση αλγορίθμων Βιοπληροφορικής με χρήση της πλατφόρμα R/Bioconductor

Η γλώσσα R είναι μια ταχέως αναπτυσσόμενη γλώσσα στατιστικών υπολογισμών και έχει ευρεία χρήση στο πεδίο της Βιοπληροφορικής. Η 'Bioconductor', είναι μια πλατφόρμα ανοιχτής πρόσβασης, η οποία δημοσιεύει λογισμικά πακέτα σε γλώσσα R, τα οποία σχετίζονται με τα πεδία της Βιοπληροφορικής, Υπολογιστικής Βιολογίας, Συστημικής Βιολογίας κ.α. Σκοπός της διπλωματικής είναι να υλοποιηθεί σε γλώσσα R και στα πρότυπα της Bioconductor, μια μεθοδολογία εύρεσης «σημαντικών» υπογράφων σε ένα βιολογικό δίκτυο (Vrahatis et al., 2013). Η μεθοδολογία είναι υλοποιημένη σε γλώσσα MatLab. Το

θέμα δεν απαιτεί γνώσεις Βιοπληροφορικής ή έρευνα πάνω στον αλγόριθμο, αλλά περιορίζεται μόνο στην αποτελεσματική υλοποίηση του στα πρότυπα της πλατφόρμας Bioconductor.

Προαπαιτούμενα: Βασικές Γνώσεις Προγραμματισμού

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Α. Βραχάτης

Επικοινωνία: kanavos@ceid.upatras.gr, agvrahatis@upatras.gr

Αναφορές:

1. Vrahatis, Aristidis G., et al. "Network-based modular markers of aging across different tissues." XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013. Springer International Publishing, 2014.
2. <https://www.bioconductor.org/>

28. Εξαγωγή γνώσης από βάσεις δεδομένων σχετικά με παρενέργειες φαρμάκων

Η φύση των κλινικών δοκιμών ενός υπό ανάπτυξη φάρμακου καθώς και η συχνά ελλιπής γνώση πάνω στον ακριβή τρόπο λειτουργίας αυτού, κάνει την εκτίμηση του συνόλου των παρενεργειών εξαιρετικά δύσκολη. Στόχος της διπλωματικής εργασίας είναι η μελέτη, ανάπτυξη και υλοποίηση μεθόδων για εξαγωγή γνώσης που αφορά σε παρενέργειες φαρμάκων από βάσεις φαρμακοεπαγρύπνησης/βιολογικών δεδομένων/ιατρική βιβλιογραφία/κοινωνικά δίκτυα.

Προαπαιτούμενα: Γνώση Προγραμματισμού (R/C++/Java)

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός, Π. Μπαλομένος

Επικοινωνία: kanavos@ceid.upatras.gr, balomenos@upatras.gr

Αναφορές:

1. Sarker, Abeed, et al. "Utilizing social media data for pharmacovigilance: A review." Journal of biomedical informatics 54 (2015): 202-212.

29. Μελέτη γλωσσικών μοντέλων (language models) και τεχνικών μηχανικής μάθησης για εξάλειψη πλεονασμού στην ανάκτηση πληροφορίας και πειραματική αξιολόγηση τους με χρήση του συνόλου δεδομένων Clueweb12.

Υπεύθυνος επ. καθ.: Μακρής Χρήστος

Περιγραφή

Η χρήση Γλωσσικών Μοντέλων στην Ανάκτηση Πληροφορίας αρχίζει να επεκτείνεται τα τελευταία χρόνια καθώς η απαίτηση για καλύτερη κατανόηση της συμπεριφοράς των χρηστών συστημάτων ανάκτησης πληροφορίας και για ποιοτικότερη ανακτώμενη πληροφορία αρχίζει να γίνεται ολοένα και πιο έντονη. Παράλληλα αρχίζει να γίνεται κοινός τόπος η χρήση τεχνικών μηχανικής μάθησης για ποιοτικότερη (χωρίς πλεονασμούς) παρουσίαση πληροφορίας στον χρήστη που την αναζητά. Αντικείμενο της συγκεκριμένη διπλωματικής εργασίας είναι η μελέτη και πειραματική επιβεβαίωση των πιο πρόσφατων σχετικών τεχνικών στο ClueWeb12 σύνολο δεδομένων, και η σύγκριση τους με κλασσικά μοντέλα. Θα επιχειρηθεί να γίνει καταγραφή των βασικών χαρακτηριστικών των σχετικών αλγορίθμων, και ίσως προσπάθεια επέκτασής τους.

Προαπαιτούμενα:

Δομές Δεδομένων, Ανάκτηση Πληροφορίας, Αλγόριθμοι, Τεχνολογίας Υλοποίησης , Αλγορίθμων, C++, Java, C#

Αναφορές

- R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval: the concepts and technology behind search, second edition, ACM Press, 2011.
- Zhai, C.X., Statistical Language Models for Information Retrieval, A critical Review, Foundations and Trends in Information Retrieval, 2(3), 2008, 137-213.

Κατανεμημένη Υλοποίηση Υψηλής Οριζόντιας Κλιμάκωσιμότητας σε Υποδομή Νεφοϋπολογιστικής Νευρωνικών Δικτύων Βαθιάς Αρχιτεκτονικής Deep Neural Networks (DNN) για Επισημείωση Αδόμητου Κειμένου με Οντότητες Wikipedia και οπτικοποίηση.

Υπεύθυνος επ. καθ.: Μακρής Χρήστος

Συνεπιβλέπων: Σίμος Μιχαήλ - Άγγελος (υποψήφιος Διδάκτωρ)

Οι αλγόριθμοι βαθιάς αρχιτεκτονικής μηχανικής μάθησης (Deep Learning) αποτελούν υποσύνολο αλγορίθμων μηχανικής μάθησης που επιχειρούν τη μοντελοποίηση μοτίβων και αφαιρέσεων υψηλού επιπέδου στα δεδομένα χρησιμοποιώντας βαθιά πολυεπίπεδα γραφήματα αποτελούμενα από γραμμικούς και μη γραμμικούς μετασχηματισμούς. Η επισημείωση αδόμητου κειμένου με υποκείμενη εννοιολογική πληροφορία από κάποια οντολογία, είναι σημαντικό βήμα προεπεξεργασίας σειράς εργασιών στα πεδία της ανάκτησης πληροφορίας, τεχνητής νοημοσύνης, μηχανικής μάθησης, διαχείρισης δεδομένων, κλπ. Η δημιουργία των οντοτήτων (άρθρα) και η επεξεργασία της γνώσης στη Wikipedia συγκλίνει σε κοινά αποδεκτές λεκτικές περιγραφές εννοιών που μπορεί να θεωρηθεί ότι συνθέτουν μια οντολογία. Η διαδικασία επισημείωσης κειμένου με οντότητες Wikipedia περιλαμβάνει την αναγνώριση κυρίαρχων εννοιών ενός τμήματος κειμένου, και η αντιστοίχισή τους με άρθρο αντίστοιχο του εννοιολογικού τους περιεχόμενο στο εκάστοτε πλαίσιο συμφραζομένων.

Η εργασία περιλαμβάνει τη μελέτη της σχετικής βιβλιογραφίας και εν συνεχεία σχεδιασμό και ανάπτυξη καινοτόμου αλγορίθμου μηχανικής μάθησης με επαυξημένες δυνατότητες οριζόντιας κλιμάκωσης σε υποδομές νεφοϋπολογιστικής. Η υλοποίηση της εργασίας θα πραγματοποιηθεί με χρήση του framework Apache Spark, και τα πειράματα θα εστιάζουν σε μεγάλου όγκου δεδομένα αδόμητου κειμένου. Θα πραγματοποιηθεί οπτικοποίηση των τελικών και ενδιάμεσων αποτελεσμάτων επιμέρους σταδίων επεξεργασίας.

Προαπαιτούμενα:

Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού (C++, Java), Εξόρυξη Γνώσης

Αναφορές:

Learning entity representation for entity disambiguation (2013) Zhengyan He , Shujie Liu , Mu Li , Ming Zhou , Longkai Zhang , Houfeng Wang

<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.495.8680>

30. Κατανεμημένη Υλοποίηση Υψηλής Οριζόντιας Κλιμάκωσιμότητας σε Υποδομή Νεφοϋπολογιστικής Deep Belief Networks (DBN) για Επισημείωση Αδόμητου Κειμένου με Οντότητες Wikipedia και οπτικοποίηση.

Υπεύθυνος επ. καθ.: Μακρής Χρήστος

Συνεπιβλέπων: Σίμος Μιχαήλ - Άγγελος (υποψήφιος Διδάκτωρ)

Οι αλγόριθμοι Deep Belief Networks (DBN) αποτελούν υποσύνολο αλγορίθμων βαθιάς αρχιτεκτονικής μηχανικής μάθησης που επιχειρούν τη μοντελοποίηση μοτίβων και αφαιρέσεων υψηλού επιπέδου στα δεδομένα χρησιμοποιώντας ένα πιθανοτικά επαγόμενο βαθύ πολυεπίπεδο γράφημα. Η επισημείωση αδόμητου κειμένου με υποκείμενη εννοιολογική πληροφορία από κάποια οντολογία, είναι σημαντικό βήμα προεπεξεργασίας σειράς εργασιών στα πεδία της ανάκτησης πληροφορίας, τεχνητής νοημοσύνης, μηχανικής μάθησης, διαχείρισης δεδομένων, κλπ. Η δημιουργία των οντοτήτων (άρθρα) και η επεξεργασία της γνώσης στη Wikipedia συγκλίνει σε κοινά αποδεκτές λεκτικές περιγραφές εννοιών που μπορεί να θεωρηθεί ότι συνθέτουν μια οντολογία. Η διαδικασία επισημείωσης κειμένου με οντότητες Wikipedia περιλαμβάνει την αναγνώριση κυρίαρχων εννοιών ενός τμήματος κειμένου, και η αντιστοίχισή τους με άρθρο αντίστοιχο του εννοιολογικού τους περιεχόμενο στο εκάστοτε πλαίσιο συμφραζομένων.

Η εργασία περιλαμβάνει τη μελέτη της σχετικής βιβλιογραφίας και εν συνεχεία σχεδιασμό και ανάπτυξη καινοτόμου αλγορίθμου μηχανικής μάθησης με επαυξημένες δυνατότητες οριζόντιας κλιμάκωσης σε υποδομές νεφοϋπολογιστικής. Η υλοποίηση της εργασίας θα πραγματοποιηθεί με χρήση του framework Apache Spark, και τα πειράματα θα εστιάζουν σε μεγάλου όγκου δεδομένα αδόμητου κειμένου. Θα πραγματοποιηθεί οπτικοποίηση των τελικών και ενδιάμεσων αποτελεσμάτων επιμέρους σταδίων επεξεργασίας.

Προαπαιτούμενα:

Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού (C++, Java), Εξόρυξη Γνώσης

Αναφορές

Applying Deep Belief Networks to Word Sense Disambiguation Peratham Wiriayathammabhum, Boonserm Kijirikul, Hiroya Takamura, Manabu Okumura

<http://arxiv.org/pdf/1207.0396v1.pdf>

31. Κατανεμημένη Υλοποίηση Υψηλής Οριζόντιας Κλιμάκωσιμότητας σε Υποδομή Νεφοϋπολογιστικής του αλγορίθμου Girvan and Newman Community detection και οπτικοποίηση

Υπεύθυνος επ. καθ.: Μακρής Χρήστος

Συνεπιβλέπων: Πισπιρίγκος Γεώργιος (υποψήφιος Διδάκτωρ)

Ένα από τα πιο σημαντικά προβλήματα των γραφημάτων μοντελοποίησης πραγματικών συστημάτων έγκειται στην αναγνώριση δομών κοινοτήτων, ήτοι την οργάνωση των κορυφών σε συστάδες, με πολλαπλές ακμές διασύνδεσης κορυφών του ιδίου συμπλέγματος και συγκριτικά λίγες ακμές διασύνδεσης κορυφών των διαφορετικών συμπλεγμάτων. Ο αλγόριθμος Girvan and Newman κατατάσσεται στην κατηγορία των "διαιρετικών" αλγορίθμων. Υποκείμενη αρχή λειτουργίας του έγκειται στην αφαίρεση των ακμών διασύνδεσης διαφορετικών κοινοτήτων. Η εργασία περιλαμβάνει τη μελέτη της σχετικής βιβλιογραφίας και εν συνεχεία σχεδιασμό και ανάπτυξη καινοτόμου αλγορίθμου με επαυξημένες δυνατότητες οριζόντιας κλιμάκωσης σε υποδομές νεφοϋπολογιστικής. Η υλοποίηση της εργασίας θα πραγματοποιηθεί με χρήση του framework Apache Spark, και τα πειράματα θα εστιάζουν σε μεγάλο όγκο τεχνητά δεδομένα κοινωνικών δικτύων. Θα πραγματοποιηθεί οπτικοποίηση των τελικών και ενδιάμεσων αποτελεσμάτων επιμέρους σταδίων επεξεργασίας.

Προαπαιτούμενα:

Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού (C++, Java), Εξόρυξη Γνώσης

Αναφορές

Girvan M. and Newman M. E. J., Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99, 7821–7826 (2002)

<http://www.pnas.org/content/99/12/7821.full.pdf>

32. Μελέτη αποδοτικών αλγορίθμων για δυναμική διαχείριση ανεστραμμένων αρχείων.

Υπεύθυνος επ. καθ.: Μακρής Χρήστος

Περιγραφή

Αντικείμενο της συγκεκριμένης διπλωματικής εργασίας είναι η μελέτη αποδοτικών δομών δεδομένων για την διαχείριση ανεστραμμένων αρχείων σε ένα δυναμικό περιβάλλον όπου μπορεί να γίνει ένθεση και διαγραφή αρχείων. Θα μελετηθούν θεωρητικά οι βασικότερες τεχνικές και θα γίνει πειραματική επιβεβαίωση της συμπεριφοράς τους. Το θέμα είναι αρκετά σημαντικό και με ερευνητική διάσταση καθώς οι τεχνικές που προτείνονται είναι λίγες σε πλήθος και βασίζονται κυρίως σε προσεκτικά δρομολογημένα επαναχτισίματα της δομής δεικτοδότησης.

Προαπαιτούμενα:

Δομές Δεδομένων, Ανάκτηση Πληροφορίας, Αλγόριθμοι, Τεχνολογία Υλοποίησης, Αλγορίθμων, C++, Java, C#

Αναφορές

- Stefan Buttcher and Charles L. A. Clarke. A Hybrid Approach to Index Maintenance in Dynamic Text Retrieval Systems. In Proceedings of the 28th European Conference on Information Retrieval, pages 229{240, London, UK, April 2006.
- Stefan Buttcher and Charles L. A. Clarke. Indexing Time vs. Query Time Trade-offs in Dynamic Information Retrieval Systems. University of Waterloo Technical Report CS-2005-31, October 2005.

33. Αποδοτικά B-δένδρα/Λεξικά για μεταβλητού μήκους κλειδιά και τεχνικές βελτιστοποίησης τους.

Περιγραφή:

Οι περισσότερες παραλλαγές των B- δένδρων λειτουργούν για ένα σύνολο κλειδιών σταθερού μήκους. Στη παρούσα διπλωματική μελετώνται (θεωρητικά και πειραματικά) παραλλαγές των B-δένδρων για μεταβλητού μήκους κλειδιά. Οι δομές αυτές εκπίπτουν σε λεξικά που αντιστοιχούν τα αρχικά κλειδιά σε ατομικά αντικείμενα επιλέγοντας με κατάλληλο τρόπο τα στοιχεία διαχωρισμού έτσι ώστε να είναι δυνατή η αναζήτηση στο B-δένδρο. Θα μελετηθούν τεχνικές που έχουν ήδη προταθεί και θα γίνει προσπάθεια βελτίωσης τους και στη συνέχεια υλοποίησης τους. Η βασική δομή δεδομένων πάνω στην οποία θα κινηθούμε είναι η δομή που περιγράφεται στο paper: M. Bender, H. Hu, B. Kuszmaul, Performance guarantees for B-trees with different-sized atomic keys, PODS 2010, 305-316 (κοιτάξτε ενδεικτική βιβλιογραφία).

Σχετικά Μαθήματα: Δομές Δεδομένων, Αλγόριθμοι, Οντοκεντρικός Προγραμματισμός

Επιβλέπων: Μακρής **Συνεπιβλέπων:** Θεοδωρίδης

Επικοινωνία: theodori@ceid.upatras.gr, makri@ceid.upatras.gr

Ενδεικτική. Βιβλιογραφία: <http://dl.acm.org/citation.cfm?doid=1807085.1807125>

8. Τίτλος: Μελέτη αποδοτικών αλγορίθμων για διαχείριση δεδομένων συμβολοσειρών στη δευτερεύουσα μνήμη διαχρονικά, και πειραματική επιβεβαίωσή τους.

Υπεύθυνος επ. καθ.: Μακρής Χρήστος

Περιγραφή

Αντικείμενο της συγκεκριμένης διπλωματικής εργασίας είναι η ανάπτυξη καινοτόμων αλγορίθμων διαχείρισης συμβολοσειρών, στη κυρία και στη δευτερεύουσα μνήμη, για την επίλυση κυρίως προβλημάτων στο χώρο της Βιοπληροφορικής. Ιδιαίτερα θα εστιάσουμε σε διαχρονικές υλοποιήσεις.

Ο έλεγχος των αλγορίθμων που θα αναπτυχθούν για την διαχείριση βιολογικών ακολουθιών θα γίνει με τη χρήση δεδομένων από βιολογικές βάσεις δεδομένων όπως η Genbank (<http://en.wikipedia.org/wiki/GeneBank>) που περιέχει δεδομένα βιολογικών ακολουθιών σε όρους DNA και η Swissprot (<http://en.wikipedia.org/wiki/Swiss-Prot>) που περιέχει δεδομένα σε όρους πρωτεϊνών. Επίσης, αν χρειαστεί θα χρησιμοποιούν και δεδομένα από διαθέσιμες βάσεις δεδομένων μικροσυστοιχιών όπως [ArrayExpress](http://www.ebi.ac.uk/microarray-as/ae/) (<http://www.ebi.ac.uk/microarray-as/ae/>), και [Stanford Microarray Database \(SMD\)](http://smd.stanford.edu/) (<http://smd.stanford.edu/>).

Προαπαιτούμενα:

Δομές Δεδομένων, Ανάκτηση Πληροφορίας, Εισαγωγή στη Βιοπληροφορική, Αλγόριθμοι, Τεχνολογίας Υλοποίησης, Αλγορίθμων, C++, Java, C#

Αναφορές

- Iliopoulos, C.S., Makris, C., Panagis, Y., Perdikuri, K., Theodoridis, E., Tsakalidis, A.K.: The Weighted Suffix Tree: An Efficient Data Structure for Handling Molecular Weighted Sequences and its Applications. Fundam. Inform. 71(2-3): 259-277 (2006)
- Askitis, N., Zobel, J.: B-tries for disk-based string management. VLDB J. 18(1): 157-179 (2009).
- Algorithms on Strings, Trees and Sequences, D. Gusfield, Cambridge University Press, 1999

34. Χρήση τεχνικών μηχανικής μάθησης για εντοπισμό πνευματικών διαταραχών

Υπεύθυνος επ. καθ.: Μακρής Χρήστος

Περιγραφή

Αντικείμενο της διπλωματικής είναι η χρήση τεχνικών μηχανικής μάθησης για την ανίχνευση σημαδίων ψυχικής αδυναμίας και μεταπτώσεων της προσωπικότητας από τη γλωσσική επεξεργασία των γραπτών μηνυμάτων ενός ατόμου. Θα είναι σε θέση να ενεργεί ενεργά σε e-mails, facebook, twitter messages.

Προαπαιτούμενα:

Ανάκτηση Πληροφορίας, Εξορυξη Δεδομένων και Γνώσης

Αναφορές

- [1] Munmun De Choudhury, Michael Gamon, Scott Counts, Eric Horvitz, Predicting Depression via Social Media, ICSW 2013
- [2] Eleanna Kafeza, Andreas Kanavos, Christos Makris, Pantelis Vikatos: T-PICE: Twitter Personality Based Influential Communities Extraction System. BigData Congress 2014: 212-219
- [3] Robert E Krueger, Avshalom Caspi, and Terrie E. Moffitt Personality Traits Are Differentially Linked to Mental Disorders: A Multitrait-Multidiagnosis Study of an Adolescent Birth Cohort, Journal of Abnormal Psychology Copyright 1996 by the American Psychological Association, Inc. 1996, Vol. 105, No. 3, 299-312

35. Μελέτη αποδοτικών αλγορίθμων για διαχείριση συμβολοσειρών με χρήση τεχνικών συμπίεσης ανεστραμμένων αρχείων

Αντικείμενο της διπλωματικής εργασίας είναι η μελέτη αποδοτικών δομών δεδομένων για διαχείριση ανεστραμμένων αρχείων σε ένα δυναμικό περιβάλλον. Μελετώνται βασικές τεχνικές και θα γίνει πειραματική αξιολόγησή τους. Με τις τεχνικές αυτές χρησιμοποιούνται επαναληπτικά υπο-συμβολοσειρές που κατασκευάζουν τις τελικές συμβολοσειρές.

Προαπαιτούμενα: Ανεστραμμένα Αρχεία, Γνώση Προγραμματισμού, Δομές Δεδομένων, Εξορυξη Γνώσης, n-grams

Επιβλέπων: Χ. Μακρής

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Αναφορές:

1. Kim, M. S., Whang, K. Y., Lee, J. G. and Lee M. J. n-Gram/2L: A Space and Time Efficient Two-Level n-Gram Inverted Index Structure
2. Iliopoulos, C. S., Makris, Ch., Panagis. Y., Perdikuri K., Theodoridis, E. and Tsakalidis, A. The Weighted Suffix Tree: An Efficient Data Structure for Handling Molecular Weighted Sequences and its Applications

36. Εξόρυξη πρωτεϊνικών αλληλεπιδράσεων με χρήση οντολογιών στο Παγκόσμιο Ιστό

Στη διπλωματική αυτή θα μελετηθούν αλγόριθμοι εξόρυξης βιολογικών δεδομένων (πρωτεϊνικών αλληλεπιδράσεων) από συλλογές δεδομένων του Παγκόσμιου Ιστού. Ειδικότερα χρησιμοποιούνται κάποιες οντολογίες (Gene Ontology και Molecular Interaction Ontology) στο BioCreative.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Βιοπληροφορική, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Χ. Μακρής

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Αναφορές:

1. <http://www.readcube.com/articles/10.1093/database/bas017?locale=en>
2. <http://www.biocreative.org/>

37. Διερεύνηση της διάχυσης της πληροφορίας στα κοινωνικά δίκτυα

Σε προηγούμενη εργασία εξετάστηκε πόσο επικοινωνιακή είναι μια κοινότητα σε ένα κοινωνικό δίκτυο. Σε αντίστοιχη εργασία έχει επιβεβαιωθεί η βιβλιογραφία που ισχυρίζεται ότι στο Twitter οι αλυσίδες μετάδοσης είναι μικρές. Αυτή η εργασία θα εξετάσει την μετάδοση της πληροφορίας στα κοινωνικά δίκτυα. Σε κάθε κοινότητα θα εξετάσει με βάση τα re-tweets τον αριθμό και την δομή των μικρών δέντρων της μετάδοσης της πληροφορίας. Θα εξετάσει επίσης και την ίδια πληροφορία για μετάδοση ανάμεσα σε κοινότητες. Με βάση τα αποτελέσματα θα εξαχθούν συμπεράσματα για την καταλληλότητα των αλγορίθμων δημιουργίας κοινοτήτων όσον αφορά το κριτήριο της μετάδοσης της πληροφορίας.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Χ. Μακρής

Συνεπιβλέπων: Ελεάννα Καφέζα, Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Σχετικές Αναφορές:

1. Goel S., Watts D. & Goldstein D. The Structure of Online Diffusion Networks

38. Προσωπικότητα (personality) κόμβων και πως επιδράει στη διάχυση των πληροφοριών στα κοινωνικά δίκτυα

Η διπλωματική έχει σχέση με κοινωνικά δίκτυα και θέλουμε να χρησιμοποιήσουμε κάποια από τα μοντέλα για διάχυση των πληροφοριών (diffusion) που να λαμβάνει υπόψη το personality του κόμβου. Υπάρχουν δυο βασικά μοντέλα, το The Linear Threshold και το Independent Cascade. Αυτά τα μοντέλα είναι πιθανοτικά και βασίζονται σε τυχαίες επιλογές.

Στη διπλωματική αυτή θέλουμε να εξετάσουμε μήπως η πιθανότητα να μεταδώσει κάποιος το μήνυμα εξαρτάται και από το personality. Επιπλέον θέλουμε να μελετήσουμε το αν η μετάδοση σ' ένα κοινωνικό δίκτυο ακολουθεί κάποιο από αυτά τα δύο μοντέλα.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Χ. Μακρή

Συνεπιβλέπων: Ελεάννα Καφέζα, Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Σχετικές Αναφορές:

1. Kempe D., Kleinberg J. & Tardos E. Maximizing the Spread of Influence through a Social Network

39. Εξατομίκευση περιεχομένου με αξιοποίηση γνώμης - κριτικής και των κοινωνικών δικτύων (aspect based opinion mining)

Περιγραφή

Στόχος της διπλωματικής είναι η ενσωμάτωση σ' ένα μοντέλο των κριτικών των χρηστών αλλά και του κοινωνικού δικτύου τους ώστε να καταλήγει στις καλύτερες προτάσεις.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού, Εξόρυξη Γνώσης

Επιβλέπων: Χ. Μακρή

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Σχετικές Αναφορές:

1. Community detection in graphs, Santo Fortunato
2. <http://dl.acm.org/citation.cfm?id=2010006>

40. Εξόρυξη δεδομένων σε χρονοσειρές

Σε αυτή τη διπλωματική θα μελετηθούν τεχνικές και αλγόριθμοι για εξόρυξη δεδομένων σε χρονοσειρές. Οι αλγόριθμοι θα εφαρμοστούν στο twitter όπου θα αναπτυχθούν μετρικές που σχετίζονται με την συμπεριφορά των χρηστών.

Αναφορές

Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Bill Chiu. 2003. A symbolic representation of time series, with implications for streaming algorithms. In Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery (DMKD '03).

Yasuko Matsubara, Yasushi Sakurai, and Christos Faloutsos. 2014. AutoPlait: automatic mining of co-evolving time sequences. In Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD '14).

41. Ένας νέος τρόπος για δημιουργία κοινοτήτων στο Twitter με χρήση πολλαπλών κριτηρίων

Σε αυτή τη διπλωματική οι χρήστες του κοινωνικού δικτύου θα μοντελοποιηθούν σαν σημεία στον πολυδιάστατο χώρο. Στη συνέχεια θα δημιουργηθούν κοινότητες στο Twitter με χρήση διαφορετικών αλγορίθμων και διαφορετικών κριτηρίων. Θα διερευνηθεί ο τρόπος για την εύρεση και την ταξινόμηση των κοινοτήτων. Θα διερευνηθεί η χρήση rectangular enclose τεχνικών για την εύρεση των καλύτερων κοινοτήτων

Αναφορές

Eleanna Kafeza, Andreas Kanavos, Christos Makris, Pantelis Vikatos: T-PICE: Twitter Personality Based Influential Communities Extraction System. BigData Congress 2014: 212-219

Lagogiannis G., Makris C., Panagis J., Sioutas S., Tsihlias K., Rectangle Enclosure Reporting in Linear Space Revisited, Journal of Automata and Combinatorics, 8(2003), pp. 633-647.

42. Σειρά διπλωματικών σε κοινωνικά δίκτυα (συν. Π. Βικάτος)

- 1) Θεωρητική και πειραματική μελέτη πρόβλεψης συνδέσεων σε multilayer graphs.
Προαπαιτούμενα : Εξόρυξη δεδομένων , θεωρία γραφημάτων .
https://www.repository.cam.ac.uk/bitstream/handle/1810/257431/Hristova_et_al-2016-EPJ_Data_Science-VoR.pdf?sequence=1
Γλώσσα προγραμματισμού : python
- 2) Ανίχνευση bots και spammers σε κοινωνικά δίκτυα
Προαπαιτούμενα : Εξόρυξη δεδομένων , θεωρία γραφημάτων .
<http://homepages.dcc.ufmg.br/~fabricio/download/ceas10.pdf>

<http://firstmonday.org/ojs/index.php/fm/article/view/4217/3700>

Γλώσσα προγραμματισμού : python

- 3) Πρόβλεψη της επιτυχίας crowdsourced campaigns από κοινωνικά δίκτυα.

Προαπαιτούμενα : Εξόρυξη δεδομένων , θεωρία γραφημάτων .

<http://dl.acm.org/citation.cfm?id=2852005>

Γλώσσα προγραμματισμού : python

- 4) Θεωρητική και πειραματική μελέτη του brand loyalty σε κοινωνικά δίκτυα

Προαπαιτούμενα : Εξόρυξη δεδομένων

Γλώσσα προγραμματισμού : python

- 5) Θεωρητική και πειραματική μελέτη μετρικών της φήμης εταιρειών στα κοινωνικά δίκτυα.

Προαπαιτούμενα : Εξόρυξη δεδομένων , θεωρία γραφημάτων .

Γλώσσα προγραμματισμού : python

- 6) Αναταξινόμηση προτάσεων βιβλίων με τεχνικές ομαδοποίησης.

Προαπαιτούμενα : Εξόρυξη δεδομένων , θεωρία γραφημάτων .

Γλώσσα προγραμματισμού : python

- 7) Θεωρητική και πειραματική μελέτη διάχυση της εμπιστοσύνης σε κοινωνικά δίκτυα.

Προαπαιτούμενα : Εξόρυξη δεδομένων , θεωρία γραφημάτων .

Γλώσσα προγραμματισμού : python

- 8) Θεωρητική και πειραματική της επέκτασης κοινού μέσω χαρακτηριστικών κοινωνικών δικτύων.

Προαπαιτούμενα : Εξόρυξη δεδομένων , θεωρία γραφημάτων .

<http://www.kdd.org/kdd2016/papers/files/adf0483-liuA.pdf>

Γλώσσα προγραμματισμού : python

- 9) Θεωρητική και πειραματική μελέτη της συμπεριφοράς σε κοινωνικά δίκτυα ατόμων με ψυχολογικές διαταραχές.

Προαπαιτούμενα : Εξόρυξη δεδομένων , θεωρία γραφημάτων .

Γλώσσα προγραμματισμού : python

- 10) Θεωρητική και πειραματική Cyberbullying σε κοινωνικά δίκτυα

<http://arxiv.org/pdf/1508.06257v1.pdf>

Προαπαιτούμενα : Εξόρυξη δεδομένων , θεωρία γραφημάτων .

Γλώσσα προγραμματισμού : python