

Topics which are hot and will remain hot for next 5–10 years are listed below.

- 1.Next generation sequencing data analysis.
- 2.Multi-omics data integration and analysis.
- 3.Personalised medicine and disease modelling.
- 4.Chemoinformatics (modelling and prediction).

->Ontologies and ontology mapping is pretty big, as are topics related to epigenetics and epigenetic data analysis.

->**Bioinformatics Research pathway** focuses on three areas of research: Bioinformatics and computational biology; Genetics and genomics; Systems biology.

In Research areas bioinformatics and computational biology involve the use of techniques including applied mathematics, informatics, statistics, computer science, artificial intelligence, chemistry and biochemistry to solve biological problems usually on the molecular level.

Research in computational biology often overlaps with systems biology.

Major research efforts in the field include sequence alignment, gene finding, genome assembly, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, and the modeling of evolution...<http://sciencedaily.com>

The Bioinformatics pathway focuses on three areas of research:

Bioinformatics and computational biology; Genetics and genomics; and Systems biology...[*Bioinformatics Research Group at SRI International*](#)

1. 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.

2. 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.huberlin.de/de/forschung/gebiete/wbi/research/publications/2012/dmc_in_ngs.pdf

3. Εξόρυξη Ιατρικών/Βιολογικών Οντοτήτων από συλλογές Κειμένων Περιγραφή Στην παρούσα διπλωματική θα μελετηθούν και αξιολογηθούν πειραματικά αλγόριθμοι εξόρυξης ιατρικών και βιολογικών οντοτήτων και σχέσεων μεταξύ των οντοτήτων (πχ. πρωτεΐνες και πρωτεϊνικές αλληλεπιδράσεις ή γονιδίων, φαρμάκων και

αλληλεπιδράσεων τους) από συλλογές βιολογικών κειμένων που βρίσκονται στο διαδίκτυο. Ποιο συγκεκριμένα θα εστιάσει στην εφαρμογή τεχνικών εξόρυξης στο PUBMED, την μεγαλύτερη online βάση δεδομένων ιατρικών και βιολογικών επιστημονικών εργασιών. **Προαπαιτούμενα: Βιοπληροφορική, Ανάκτηση Πληροφορίας, Εξόρυξη Γνώσης, Αλγόριθμοι, Γλώσσες Προγραμματισμού** Επιβλέπων: Τσακαλίδης Συνεπιβλέπων: Θεοδωρίδης Επικοινωνία: theodori@ceid.upatras.gr Σχετικές Αναφορές: 1. Jorg Hakenberg, Robert Leaman, Nguyen Ha Vo, Siddhartha Jonnalagadda, Ryan Sullivan, Christopher Miller, Luis Tari, Chitta Baral, and Graciela Gonzalez. 2010. Efficient Extraction of Protein-Protein Interactions from Full-Text Articles. IEEE/ACM Trans. Comput. Biol. Bioinformatics 7, 3 (July 2010), 481-494. 2. Nate Sutton , Laura Wojtulewicz , Neel Mehta , Graciela Gonzalez, Automatic approaches for gene-drug interaction extraction from biomedical text: corpus and comparative evaluation, Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, June 08-08, 2012, Montreal, Canada

4. Αλγόριθμοι Ανάλυσης Δεδομένων Next Generation

Sequencing Περιγραφή Σκοπός της παρούσας διπλωματικής είναι η θεωρητική και πρακτική μελέτη αλγόριθμων και δομών δεδομένων για την αποδοτική ανάλυση μεγάλου πλήθους ακολουθιών DNA στην κύρια μνήμη υπολογιστικών συστημάτων. Θα γίνει εστίαση είτε στο πρόβλημα της εξαγωγής των μέγιστων κοινών υπο-συμβολοσειρών /προθεμάτων (Longest Common Prefix/Substring) ή στους αλγόριθμους ομαδοποίησης και κατηγοριοποίησης ακολουθιών (sequence clustering). Θα εξεταστούν τεχνολογίες επεξεργασίας μεγάλου όγκου όπως το Hadoop, Apache Spark, SeqPig, SparkSeq και άλλες. **Προαπαιτούμενα: Βιοπληροφορική, Ανάκτηση Πληροφορίας, Εξόρυξη Γνώσης, Αλγόριθμοι, Γλώσσες Προγραμματισμού** Επιβλέπων: Τσακαλίδης Συνεπιβλέπων: Θεοδωρίδης Επικοινωνία: theodori@ceid.upatras.gr Σχετικές Αναφορές: 1. Markus J. Bauer, Anthony J. Cox, Giovanna Rosone, and Marinella Sciortino. 2012. Lightweight LCP construction for next-generation sequencing datasets. In Proceedings of the 12th international conference on Algorithms in Bioinformatics 2. Kamil Salikhov, Gustavo Sacomoto, Gregory Kucherov: Using Cascading Bloom Filters to Improve the Memory Usage for de Bruijn Graphs. 364-376 WABI 2013

5. 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: Χ. Μακρής

Συνεπιβλέπων: Α. Καναβός

Επικοινωνία: kanavos@ceid.upatras.gr

Αναφορές:

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

6. 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/>

7. 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 γραφών. 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

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

To 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). DSubs: an

R package for flexible identification of differentially expressed subpathways

using RNA-seq experiments. *Bioinformatics*, btw544. 3.

<http://shiny.rstudio.com/>

9. 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

Κοινωνικά Δίκτυα

1. 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:

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

2. 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. Μακρής

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>

3. Τίτλος : Μελέτη τεχνικών και μεθόδων ανάλυσης συναισθήματος και εφαρμογή τους σε κοινωνικά δίκτυα

Περιγραφή

Στη βιβλιογραφία γνωρίζουμε ότι για ένα κείμενο, μπορούμε να αποφασίσουμε εάν έχει

θετική ή αρνητική χροιά, ή αν είναι ουδέτερο. Στη διπλωματική αυτή λοιπόν ερευνώνται

μέθοδοι ανάλυσης συναισθήματος και ακολούθως γίνεται εφαρμογή τους σε διαθέσιμα

κοινωνικά δίκτυα, όπως το twitter.

Προαπαιτούμενα: Ανάκτηση Πληροφορίας, Γνώση Προγραμματισμού (C++, Java, κ.α.),

Εξόρυξη Γνώσης

Επιβλέπων: Α. Τσακαλίδης

Συνεπιβλέπων: Α. Καναβός Επικοινωνία: kanavos@ceid.upatras.gr

Σχετικές Αναφορές:

1. Liu B. Sentiment Analysis and Opinion Mining
2. <http://www.fastcoexist.com/3029598/using-citizens-as-human-sensors-to-improve-public-services>

4. 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

5. 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

6. 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>