

PROJECTS ΜΑΘΗΜΑΤΟΣ
ΣΥΓΧΡΟΝΕΣ ΔΙΑΔΙΚΤΥΑΚΕΣ ΕΦΑΡΜΟΓΕΣ

1^η Έκδοση (19/10)

Ιωάννης Αναγνωστόπουλος, Αναπλ. Καθ. Παν. Θεσσαλίας
Γεράσιμος Ραζής, Μεταδιδάκτορας Ερευνητής Παν. Θεσσαλίας
Μάριος Κόνιαρης, Μέλος Ε.ΔΙ.Π, Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών, ΕΜΠ

Τα projects ανατίθενται είτε ατομικά είτε σε ομάδες των δύο ατόμων και αποτελούν υλοποίηση συγκεκριμένων εφαρμογών. Απαιτείται καλή γνώση προγραμματισμού, τεκμηρίωσης κώδικα και εφαρμογής, καθώς και ευελιξία στη σύνθεση διαδικτυακών υπηρεσιών και εφαρμογών.

Τα παραδοτέα ανά project είναι:

- α) υλοποίηση εφαρμογής (κώδικας),
- β) τεκμηρίωση κώδικα στο github,
- γ) οδηγός χρήσης της εφαρμογής στο github και
- δ) παρουσίαση (σε ppt).

Οι συναντήσεις θα πραγματοποιούνται διαδικτυακά, κατόπιν συνεννόησης με τον εκάστοτε Project Leader.

Δήλωση ενδιαφέροντος (1^η φάση): μέχρι 04/11/2022 με e-mail στον εκάστοτε Project Leader και κοινοποίηση στον Υπεύθυνο (με θέμα: {ΤεχνοΟικονομικά – Project Σύγχρονες Διαδικτυακές Υπηρεσίες})

Ανάθεση εργασίας (επιβεβαίωση από εμάς): μέχρι 11/11/2022

ΘΕΜΑΤΑ

1. Δημιουργία εφαρμογής ανάλυσης φωτογραφιών κοινωνικών δικτύων μέσω της πλατφόρμας Google Vision AI (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: Γ. Ραζής {razis@dib.uth.gr}

Τα τελευταία χρόνια έχει εμφανιστεί η τάση στα Κοινωνικά Δίκτυα (ΚΔ) της ραγδαίας αύξησης του πλήθους των διαμοιραζόμενων μηνυμάτων πολυμεσικού περιεχομένου (εικόνες, βίντεο). Αποτέλεσμα είναι η εμφάνιση ΚΔ (π.χ. Instagram) στοχευμένα αποκλειστικά και μόνο σε τέτοιου είδους μηνύματα. Είναι αρκετά συνήθης η προσπάθεια των χρηστών να χαρακτηρίσουν με διάφορες τεχνικές το περιεχόμενο αυτών των μηνυμάτων (π.χ. μέσω hashtags ή άλλων λεκτικών περιγραφών). Ωστόσο, αυτό δεν είναι επαρκές καθώς δεν υιοθετείται μία καθολική τεχνική και οι περιγραφές είναι συχνά υποκειμενικές.

Σκοπός της εργασίας είναι η δημιουργία ενός συστήματος αυτόματης ανάλυσης και χαρακτηρισμού/επισήμανσης εικόνων ΚΔ μέσω της πλατφόρμας Google Vision AI. Το Google Vision API προσφέρει προ-εκπαιδευμένα μοντέλα μηχανικής μάθησης (machine learning) για άμεση χρήση σε εικόνες όπως ανάθεση ετικετών σε προκαθορισμένες κατηγορίες, εντοπισμός αντικειμένων και προσώπων, και ανάγνωση κειμένου. Συνεπώς, το σχετικό περιεχόμενο θα αντλείται από τα API των ΚΔ (ή μπορούν να δοθεί υπάρχον σχετικό περιεχόμενο) και θα εμπλουτίζεται μέσω του Google Vision API. Έπειτα, οι πληροφορίες θα αποθηκεύονται με σκοπό τη μεταγενέστερη ανάκτηση και την περαιτέρω αξιοποίησή τους.

Αντικείμενα ενασχόλησης: προγραμματισμός με Java ή Python, χρήση του Twitter / Instagram API, Google Vision AI API

2. Ανάλυση αναρτημένου περιεχομένου λογαριασμών κοινωνικών δικτύων ιδίων χρηστών (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: Γ. Ραζής {razis@dib.uth.gr}

Η εργασία περιλαμβάνει τη συλλογή και ανάλυση των αναρτήσεων σε κοινωνικά δίκτυα (π.χ. Twitter, Facebook, Instagram κλπ.) προερχόμενες από λογαριασμούς, οι οποίοι ανήκουν στον ίδιο χρήστη. Πρωταρχικός σκοπός αυτής της εργασίας είναι η ανάλυση των Ονομάτων Χρηστών (Usernames) (π.χ. βαθμός ομοιότητας) τα οποία χρησιμοποιούν οι χρήστες στα κοινωνικά αυτά δίκτυα, η εκτενής ανάλυση του διαχεόμενου περιεχομένου τους (π.χ. συχνότητα δημιουργίας και αναδιανομής περιεχομένου, θεματικές περιοχές διαχεόμενου περιεχομένου, κ.λπ.), καθώς και η προτίμηση χρήσης συγκεκριμένων κοινωνικών δικτύων.

Αντικείμενα ενασχόλησης: προγραμματισμός με Python ή Java, χρήση του Twitter API, Facebook API και Instagram API

3. Δημιουργία εφαρμογής παρουσίασης ενοποιημένων προφίλ χρηστών από διαφορετικά κοινωνικά δίκτυα (π.χ. Twitter, Facebook, Instagram) και αλληλεπίδραση με αυτά (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: Γ. Ραζής {razis@dib.uth.gr}

Είναι γεγονός πως η παρουσία χρηστών στα κοινωνικά δίκτυα ολοένα και αυξάνεται. Εκτός αυτού, αυξάνεται και το πλήθος των κοινωνικών δικτύων. Ως αποτέλεσμα, οι χρήστες διαμοιράζονται περιεχόμενα σε πολλαπλά δίκτυα ταυτόχρονα, δυσκολεύοντας τους

«αναγνώστες» τους να τους παρακολουθήσουν, κάτι που οδηγεί σε απώλεια χρόνου και πληροφορίας.

Σκοπός της εργασίας είναι η δημιουργία μιας εφαρμογής η οποία παρουσιάζει ενοποιημένα προφίλ χρηστών από διαφορετικά κοινωνικά δίκτυα (π.χ. Twitter, Facebook, Instagram), καθώς και του περιεχομένου που διαμοιράζονται σε αυτά. Συνεπώς, όλη η σχετική πληροφορία θα αντλείται από τα API των κοινωνικών δικτύων και θα παρουσιάζεται ενοποιημένη. Επιπροσθέτως, θα υπάρχει η δυνατότητα αλληλεπίδρασης με αυτά (π.χ. retweet, like, share, κλπ) μέσω ενός του interface της εφαρμογής το οποίο όμως θα εκτελεί την ενέργεια στο επιλεγμένο κοινωνικό δίκτυο του χρήστη.

Αντικείμενα ενασχόλησης: προγραμματισμός με Python και/ή Java, χρήση του Twitter API, Facebook API και Instagram API, HTML, Javascript, AJAX

4. Δημιουργία ενός visualized SPARQL query builder για συγκεκριμένη οντολογία (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: Γ. Ραζής {razis@dib.uth.gr}

Είναι γεγονός πως για να μπορέσει ένας χρήστης να θέσει ερωτήματα (queries) σε έναν γράφο σημασιολογικών δεδομένων πρέπει να γνωρίζει τόσο τη δομή της οντολογίας όσο και τη σύνταξη της SPARQL (τη γλώσσα ερωτημάτων γράφων).

Ο σκοπός της εργασίας είναι να δημιουργηθεί μία ιστοσελίδα η οποία θα επιτρέπει στον χρήστη τη δημιουργία ερωτημάτων σύμφωνα με τις επιλογές του (π.χ. μέσω drop down λιστών, κουμπιών προσθήκης περιορισμών, auto complete πεδίων, κλπ). Αυτός ο visualized SPARQL query builder θα απευθύνεται αποκλειστικά προς μία συγκεκριμένη οντολογία (μπορείτε να επιλέξετε εσείς μία ή να σας προταθεί).

Αντικείμενα ενασχόλησης: προγραμματισμός με Python ή Java, HTML, Javascript, AJAX, SPARQL

5. Δημιουργία Νομικού Chat Bot (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: M. Κόνιαρης {mkoniari@central.ntua.gr}

Τα συστήματα τύπου Chat Bots προσπαθούν να εξυπηρετήσουν τους χρήστες στην ανάκτηση της πληροφορίας που καλύπτει τις ανάγκες τους. Σκοπός της εργασίας είναι η υλοποίηση ενός συστήματος που θα επιτρέπει την αυτόματη απάντηση νομικών ερωτήσεων/ αναζητήσεων σε κοινωνικά δίκτυα/ συστήματα επικοινωνίας πχ facebook, skype. Η εργασία περιλαμβάνει χρήση της ελληνικής νομοθεσίας σε μορφή LegalDocML, την υλοποίηση διαλογικού συστήματος τύπου Chat Bot καθώς και την αξιολόγηση της αποτελεσματικότητάς του.

Αντικείμενα ενασχόλησης: προγραμματισμός με Python ή Java, xml

6. Μηχανισμοί Αυτόματης Κατηγοριοποίησης Νομικών Δεδομένων (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: M. Κόνιαρης {mkoniari@central.ntua.gr}

Σκοπός της εργασίας είναι η υλοποίηση ενός συστήματος που θα επιτρέπει την αυτόματη κατηγοριοποίηση νομικών κειμένων (Νομοθεσία / Νομολογία). Η εργασία περιλαμβάνει την εξέταση της δυνατότητας χρήσης του EuroVoc (πολύγλωσσος θησαυρός της Ευρωπαϊκής

Ένωσης), την υλοποίηση αλγόριθμων και κριτηρίων κατηγοριοποίησης κειμένων καθώς και την αξιολόγηση της αποτελεσματικότητά τους.

Αντικείμενα ενασχόλησης: προγραμματισμός με Python ή Java, xml

7. Οντολογία Κωδικοποίησης Ελληνικής Νομοθεσίας (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: M. Κόνιαρης {mkoniari@central.ntua.gr}

Σκοπός της εργασίας είναι προέκταση του οντολογικού σχήματος της Ευρωπαϊκής Νομοθεσίας ELI (<http://publications.europa.eu/mdr/eli/index.html>), για την αποτύπωση της Ελληνικής Νομοθεσίας. Για την υλοποίηση της εργασίας, οι σπουδαστές θα έχουν στην διάθεση τους το υπάρχον σχήμα της Ευρωπαϊκής νομοθεσίας. Η εργασία περιλαμβάνει και την δημιουργία προγράμματος εμπλουτισμού/ μετασχηματισμού της ελληνικής νομοθεσίας από LegalDocML στο νέο σχήμα

Αντικείμενα ενασχόλησης: προγραμματισμός με Python ή Java, xml, xslt

8. Ανάλυση και μελέτη εξέλιξης μετρητή παραπομπών (citations) επιστημονικών δημοσιεύσεων (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: Γ. Ραζής {razis@dib.uth.gr}

Σκοπός της εργασίας είναι η μελέτη εξέλιξης της προόδου των αριθμών παραπομπών επιστημονικών δημοσιεύσεων. Θα αξιοποιηθούν τα διδόμενα από υφισταμένες πηγές, τα οποία παρέχονται μέσω SPARQL endpoints. Οι πληροφορίες θα τοποθετούνται σε γράφους μέσω της βιβλιοθήκης networkX της γλώσσας Python, σύμφωνα με ορισμένα κριτήρια, και κατόπιν θα αναλύονται για την ανεύρεση συγκεκριμένων χαρακτηριστικών. Ενδεικτικά παρατίθενται τα εξής:

- εύρεση των top-k συγγραφέων, ιδρυμάτων, κλπ, σε συγκεκριμένο χρονικό πλαίσιο.
- ανεύρεση των top-k ερευνητικών περιοχών σε συγκεκριμένο χρονικό πλαίσιο (π.χ. τελευταία 3 χρόνια) και πρόβλεψη των επόμενων δημοφιλών περιοχών.
- προτάσεις σε ερευνητές για υποψήφιους συν-συγγραφείς δεδομένης της ομοιότητάς τους σε ερευνητικά ενδιαφέροντα.
- προσδιορισμός δυνητικών νέων top-k συγγραφέων, ιδρυμάτων, κλπ. Βάση ρυθμού αύξησης του πλήθους των παραπομπών που λαμβάνουν.
- άμεση παρουσίαση σημαντικών ερευνητικών εργασιών ανά έτος δημοσίευσης.

Αντικείμενα ενασχόλησης: προγραμματισμός με Python, χρήση του OpenCitations SPARQL endpoint (<http://opencitations.net/sparql>), SPARQL, networkX Python library

9. Δημιουργία Οντολογίας Νομικών Δεδομένων (1-2 άτομα)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leaders: M. Κόνιαρης {mkoniari@central.ntua.gr}

Γ. Ραζής {razis@dib.uth.gr}

Η δημοσίευση των Νομικών Δεδομένων (Legal Data) με την εφαρμογή τεχνολογιών του Σηματολογικού Ιστού αποτελεί πλέον κοινή πρακτική (ΗΠΑ, Ην Βασίλειο - <http://www.legislation.gov.uk/>, Ευρωπαϊκή Ένωση - <http://eur-lex.publicdata.eu/>). Η εφαρμογή των κανόνων των Συνδεδεμένων Δεδομένων (LOD) μέσω οντολογιών επιτρέπει επίσης τη διασύνδεση των υφιστάμενων νομικών συλλογών που σχετίζονται σημασιολογικά σχηματοποιώντας το "νέφος" των πηγών συνδεδεμένων νομικών δεδομένων.

Σκοπός της εργασίας είναι η δημιουργία ενός οντολογικού σχήματος της Ελληνικής Νομοθεσίας. Για την υλοποίηση της εργασίας, οι σπουδαστές θα έχουν στην διάθεση τους το υπάρχον σχήμα της Ελληνικής νομοθεσίας σε xml.

Τεχνολογίες: Java, Javascript, Sparql, RDF,xml

Σχετικό Υλικό:

Linked Data: Evolving the Web into a Global Data Space: <http://linkeddatatobook.com>

Simple Protocol and RDF Query Language (SPARQL),

<http://www.slideshare.net/olafhartig/an-introduction-to-sparql>

“Opening Public Data: a path towards innovative legal services,” Law via Internet Conf.,

2011, <http://www.hkliv.hk/conference/paper/2D2.pdf>

Σχετική προσέγγιση του LATC, Linked Open Data Around-The-Clock, <http://eur-lex.publicdata.eu/>

10. Δημιουργία εφαρμογής αποτύπωσης σε χάρτη μηνυμάτων κοινωνικών δικτύων κρατικών υπηρεσιών (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: Γ. Ραζής {razis@dib.uth.gr}

Τα τελευταία χρόνια ολοένα και περισσότερες κρατικές υπηρεσίες αξιοποιούν τα Κοινωνικά Δίκτυα (ΚΔ) με σκοπό την άμεση ενημέρωση των πολιτών. Οι λογαριασμοί του Πυροσβεστικού Σώματος (@pyrosvestiki) και της Ελληνικής Αστυνομίας (@hellenicpolice) στο Twitter είναι χαρακτηριστικά παραδείγματα. Σε καθημερινή βάση δεκάδες μηνυμάτων κοινοποιούνται, ενημερώνοντας για σχετικές τους δράσεις και ενέργειες, αναφέροντας παράλληλα σε μικρό ή μεγάλο βαθμό γεωγραφικές πληροφορίες. Ωστόσο, η γεωγραφική πληροφορία δεν είναι διαθέσιμη σε κάποια συγκεκριμένη ή μηχαναγνώσιμη μορφή καθιστώντας την οπτικοποίηση των κοινωνικών μηνυμάτων σε χάρτη ως μία πρόκληση.

Σκοπός της εργασίας είναι η δημιουργία ενός συστήματος αυτόματης ανάλυσης μηνυμάτων των προαναφερθέντων λογαριασμών από το Twitter, κατηγοριοποίησης και αποτύπωσής τους σε χάρτη, με βάση την αναγνωρισμένη γεωγραφική πληροφορία. Συνεπώς, το περιεχόμενο (tweets) θα αντλείται από το Twitter API, μέσω διαφόρων τεχνικών (ευρετικοί κανόνες ή ανάλυσης φυσικής γλώσσας) οι υποψήφιοι όροι θα ανακαλύπτονται, και τέλος τα κατάλληλα μηνύματα θα αποτυπώνονται σε χάρτη (π.χ. Google Maps). Τεχνολογίες: Java, Javascript, Sparql, RDF,xml

Αντικείμενα ενασχόλησης: προγραμματισμός με Java ή Python, χρήση του Twitter API, Google Maps API, HTML, Javascript

11. Ανεύρεση Σημαντικών Οντοτήτων του ακαδημαϊκού βιβλιογραφικού τομέα (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: Γ. Ραζής {razis@dib.uth.gr}

Πρόσφατα, έχει παρατηρηθεί μια αυξημένη τάση σε όλους τους επιστημονικούς κλάδους για τη σύνταξη δημοσιεύσεων με πολλαπλούς συγγραφείς. Συνεπώς, η θέση των ονομάτων των συντελεστών διαδραματίζει σημαντικό ρόλο τόσο όσον αφορά την απεικόνιση των

πραγματικών συνεισφορών, όσο και στις αξιολογήσεις των επιτροπών αξιολόγησης. Συνήθως, ο πρώτος συγγραφέας συνεισφέρει περισσότερο και έτσι λαμβάνει το μεγαλύτερο μέρος της πίστωσης, ενώ η θέση των επόμενων συγγραφέων καθορίζεται από διάφορους παράγοντες, όπως οι συνεισφορές, η αλφαβητική σειρά ή η αντίστροφη αρχαιότητα. Η κατάταξη των συγγραφέων σε δημοσιεύσεις με δύο συγγραφείς είναι απλή, ωστόσο η ερμηνεία της θέσης κάθε συγγραφέα γίνεται ολοένα και πιο αυθαίρετη καθώς ο αριθμός τους αυξάνεται.

Στη βιβλιογραφία αναλύονται αρκετά σχετικά κριτήρια και μετρικές, καθώς υπάρχει σημαντική αύξηση του αριθμού των συγγραφέων σε έγγραφα που υποβάλλονται σε βιοϊατρικά περιοδικά καθώς και στην πρακτική της «δωρεάς» συγγραφής. Ωστόσο, εξακολουθεί να λείπει μια μεθοδολογία για τον προσδιορισμό της πίστωσης όσον αφορά τις βαθμολογίες που σχετίζονται με συγγραφείς, άρθρα, χώρους (δηλ. περιοδικά και συνέδρια), ιδρύματα, καθώς και θέματα σε όλο τον επιστημονικό και εκδοτικό τομέα. Σκοπός της εργασίας είναι η αξιολόγηση και περαιτέρω βελτίωση της μεθοδολογίας «Ανεύρεση Σημαντικών Οντοτήτων», η οποία αξιοποιώντας τις παραπομπές κατατάσσει και βαθμολογεί όλες τις σημαντικές οντότητες του επιστημονικού και εκδοτικού τομέα, όπως συγγραφείς, άρθρα, χώρους (δηλ. περιοδικά και συνέδρια), ιδρύματα, καθώς και θεματικές περιοχές. Αυτή είναι η πρώτη μεθοδολογία αξιοποίησης παραπομπών που αποδίδει ταυτόχρονα βαθμολογίες σε όλες τις προαναφερθείσες οντότητες. Χρησιμοποιώντας αυτές τις βαθμολογίες, μπορούμε να πάρουμε μία αντικειμενική μετρική του αντίκτυπου μιας ακαδημαϊκής οντότητας (π.χ. τον αντίκτυπο ενός συγγραφέα και ενός ερευνητικού εγγράφου).

Αντικείμενα ενασχόλησης: προγραμματισμός με Java ή Python, ανάλυση δεδομένων

12. Συλλογή και ανάλυση δεδομένων σχετιζόμενων με τηλεοπτικές σειρές (1 άτομο)

Υπεύθυνος: I. Αναγνωστόπουλος {janag@uth.gr}

Project leader: Γ. Ραζής {razis@dib.uth.gr}

Τα τελευταία χρόνια έχει παρατηρηθεί μια ραγδαία αύξηση του πλήθους των [τηλεοπτικών σειρών](#), άμεση απόρροια της αλλαγής της τάσης των τηλεθεατών ως προς την διασκέδαση τους και τις χρονικές τους διαθέσιμότητες. Μία συνήθης πρακτική των εξαιρετικά δημοφιλών σειρών, είναι η τεχνητή επέκταση της διάρκειάς τους μέσω της εισαγωγής επεισοδίων ή ακόμα και ολόκληρων κύκλων χωρίς κάποια άμεση προσφορά στη βασική πλοκή (γνωστά ως "πληρώσεις"/"fillers"). Σε αυτές τις περιπτώσεις η βαθμολογία είναι ικανή να αποτυπώσει αυτό το γεγονός, ούσα αρκετά χαμηλότερη από τις βαθμολογίες των υπόλοιπων επεισοδίων/κύκλων.

Σκοπός της εργασίας είναι η συλλογή δεδομένων που σχετίζονται με τηλεοπτικές σειρές μέσω συγκεκριμένων διεπαφών (API), την ανάλυσή τους με σκοπό την αξιολόγηση του ανωτέρω σημείου, τη μελέτη για ύπαρξη συγκεκριμένων μοτίβων ως προς τις βαθμολογίες και τις συμπεριφορές των τηλεθεατών, καθώς και την ανάλυση των δεδομένων από την οπτική των κατηγοριών των σειρών. Ενδεικτικά ερωτήματα που θα απαντηθούν με το πέρας της εργασίας:

- Πόσους κύκλους διαφορά έχουν οι δημοφιλείς σειρές σε σχέση με τις λιγότερο/μη δημοφιλείς;
- Ποιος κύκλος των δημοφιλών σειρών έχει την υψηλότερη/χαμηλότερη βαθμολογία και ποιος των υπολοίπων; Πόση απόκλιση υπάρχει μεταξύ τους;
- Επηρεάζει (και πόσο) ο αριθμός κύκλων και επεισοδίων τη συνολική βαθμολογία μία σειράς;

- Υπάρχει συσχέτιση μεταξύ της συνολικής βαθμολογίας μιας σειράς και του πλήθους ατόμων που ψήφισαν;
-

Αντικείμενα ενασχόλησης: προγραμματισμός με Java ή Python, χρήση του [IMDB API](#), Βάσεις Δεδομένων (οποιοδήποτε είδους), ανάλυση δεδομένων