#### ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

#### ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Project

**Η μηχανή αναζήτησης ASTERAS**

**1η Προγραμματιστική Εργασία**

**Ηλίας Αλεξανδρόπουλος**

### 2022201900007

**Βίνσε Ράντσες**

### 2022201900196

#### Τρίπολη, Ιανουάριος 2023

**Περιεχόμενα**

Περιεχόμενα

[Προεπεξεργασία περιεχοµένου ερευνητών 3](#_Toc123663226)

[Κατασκευή ευρετηρίου και εισαγωγή/διαγραφή ερευνητών 3](#_Toc123663227)

[Αναζήτηση και συνάφεια ερευνητών 3](#_Toc123663228)

[Παραδείγµατα επίδειξης 4](#_Toc123663229)

[Bonus 9](#_Toc123663230)

[User Manual 10](#_Toc123663231)

[Αναφορές 11](#_Toc123663232)

# 

# Προεπεξεργασία περιεχοµένου ερευνητών

Για την διαδικασία αυτή χρησιμοποιήσαμε έναν έτοιμο parser bib αρχείων(BibTeX-Parser). Ο οποίος δεν κάνει κάποιο case folding ή stemming ή punctuation removal.

Αυτά τα κάνουμε μέσω του lucene όταν ο χρήστης δώσει κάποιο query.

# Κατασκευή ευρετηρίου και εισαγωγή/διαγραφή ερευνητών

Ο χρήστης έχει την δυνατότητα να εισάγει ή να διαγράψει κάποιον ερευνητή μέσω του γραφικού περιβάλλοντος του webpage. Κάθε φορά που προσθέτει έναν καινούργιο ερευνητή κατασκευάζουμε το ευρετήριο ξανά. Δηλαδή προσθέτουμε το αρχείο του ερευνητή στο φάκελο Data και ξανα φτιάχνουμε indexes για όλα τα αρχεία του φακέλου αυτού. Τέλος, στην περίπτωση εισαγωγής ερευνητή που υπάρχει ήδη διαγράφουμε το υπάρχων αρχείο και βάζουμε το καινούργιο.

# Αναζήτηση και συνάφεια ερευνητών

Η αναζήτηση γίνεται μέσω του γραφικού περιβάλλοντος του webpage και συγκεκριμένα ενώς search-bar όπου ο χρήστης έχει την δυνατότητα να πληκτρολογήση το query του. Η εφαρμογή στην συνέχεια θα του εμφανίσει του ερευνητές που θεωρεί πιο σχετικούς για το συγκεκριμένο query. Επιπλέον ο χρήστης έχει την δυνατότητα να εισάγει έναν αριθμό (k) όπου αναπαριστά τα top-k αποτελέσματα (default k = 5).

Ο υπολογισμός της συνάφειας περιγράφεται παρακάτω:

Α) Υπολογισμός συνάφειας ζεύγους ερευνητών: ο χρήστης δίνει σαν query τα δύο ονόματα των ερευνητών διαχωρισμένα με κενό μέσω του search-bar που αναφέραμε παραπάνω και η εφαρμογή του εμφανίζει την συνάφεια μεταξύ τους.

B) Υπολογισμός συνάφειας ενός ερευνητή: ο χρήστης δίνει σαν query το όνομα του ερευνητή μέσω του search-bar που αναφέραμε παραπάνω και η εφαρμογή του εμφανίζει την συνάφεια μεταξύ του συγκεκριμένου ερευνητή και όλων των ερευνητών που έχει η εφαρμογή.

# Παραδείγµατα επίδειξης

Εισαγωγή ερευνητών

Graphical user interface, text, application

Description automatically generated

A screenshot of a computer

Description automatically generated

Και αφού πατήσουμε Submit θα μας εμφανιστεί μία λίστα με τους ερευνητές.

Graphical user interface, application, Teams

Description automatically generated

Όπως φαίνεται ο χρήστης μπορέι να διαγράψει κάποιον ερευνητή μέσω του textfield ή μέσω του εικονιδίου «κάδου» δεξία από το όνομα κάθε ερευνητή.

Τώρα εάν θέλουμε να τρέξουμε κάποιο query:

Text

Description automatically generated with medium confidence

Επιλέγουμε την επιλογή Query στο select menu και γράφουμε το query στο searchbar. Επίσης γράφουμε και τον επιθυμητό αριθμό top αποτελεσμάτων στο textfield «K:». Τέλος αφού πατήσουμε το κουμπί Search παίρνουμε το παρακάτω αποτέλεσμα.

Graphical user interface, text, application, email

Description automatically generated

Όπως φαίνεται στο screenshot βρήκαμε 2 αποτελέσματα τα οποία είναι και ταξινομημένα από το μεγαλύτερο score στο μικρότερο. Επιπλέον ο χρήστης έχει την δυνατότητα να τρέξει το ίδιο query αλλά με διαφορετικό K μέσω του αντίστοιχου textfield.

Τέλος για την συνάφεια ερευνητών έχουμε δύο περιπτώσεις.

1. Υπολογισμός συνάφειας ζεύγους ερευνητών

Graphical user interface, text

Description automatically generated

Πρέπει αρχικά να επιλέξουμε την επιλογή “Cosine Similarity” και να γράψουμε στο searchbar τα δύο ονόματα των ερευνητών. Προφανώς στο παράδειγμα μας περιμένουμε αποτέλεσμα 1 διότι δώσαμε το ίδιο όνομα ερευνητή.

Τώρα αφού πατήσουμε το κουμπί search μας εμφανίζεται το αποτέλεσμα.

Graphical user interface, text, application

Description automatically generated

1. Υπολογισμός συνάφειας ενός ερευνητή

Graphical user interface

Description automatically generated

Πρέπει αρχικά να επιλέξουμε την επιλογή “Cosine Similarity” και να γράψουμε στο searchbar το όνομα του ερευνητή. Τέλος εάν θέλουμε μπορούμε να αλλάξουμε και το K.

Τώρα αφού πατήσουμε το κουμπί search μας εμφανίζεται το αποτέλεσμα.

Graphical user interface, text, application

Description automatically generated

# Bonus

* Η υλοποίηση του προγράµµατος ως web εφαρµογη
* χρήστης μπορεί να ανεβάζει απευθείας κάποιο .zip αρχείο που περιέχει πολλά .bib αρχεία μέσα για την εισαγωγή ερευνητών.

# User Manual

Αφου γίνει η αποσυμπιεση του AlexandropoulosRantses.zip ανοίγουμε ένα τερματικό και κάνουμε cd στο directory στο οποίο κάναμε extract το zip.

Και τρέχουμε την παρακάτω εντολή:

java -jar .\jetty-runner-9.4.43.v20210629.jar --path /Asteras .\dist\Asteras.war

Στην συνέχεια ελαχιστοποιούμε το τερματικό χωρίς να πατησουμε πάνω σε αυτό και ανοίγουμε έναν browser.

Μέσω του browser και της σελίδας localhost:8080/Asteras μπορόυμε να περιηγηθούμε στην εφαρμογή.

# Αναφορές

* BibTeX-Parser για την προεπεξεργασία των .bib αρχείων  
  <https://github.com/jakubowiczish/BibTeX-Parser>
* Χρησιμοποιήθηκε για τον υπολογισμό της ομοιότητας συνημιτόνου

<https://stackoverflow.com/questions/54427663/using-cosine-similarity-for-two-text-files>